



Parameterization of atmospheric convection in numerical climate models - Practices and epistemological challenges

Ludovic Touzé-Peiffer

► To cite this version:

Ludovic Touzé-Peiffer. Parameterization of atmospheric convection in numerical climate models - Practices and epistemological challenges. Meteorology. Sorbonne Université, 2021. English. NNT : 2021SORUS539 . tel-04215936

HAL Id: tel-04215936

<https://theses.hal.science/tel-04215936>

Submitted on 23 Sep 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

École doctorale 129 : Sciences de l'Environnement

THÈSE

pour obtenir le grade de docteur délivré par

Sorbonne Université

Spécialité doctorale “Sciences de l'Environnement”

*qui sera
présentée et soutenue publiquement par*

Ludovic TOUZÉ-PEIFFER

le 22 septembre 2021

Paramétrisation de la convection atmosphérique dans les modèles numériques de climat – Pratiques et enjeux épistémologiques

Directeur de thèse : **Hervé LE TREUT**

Co-Directeurs de thèse : **Anouk BARBEROUSSE, Nicolas ROCHETIN**

Jury

| | | |
|------------------------------|--|---------------|
| Vincent LAM | Professeur, Université de Bern | Rapporteur |
| David SALAS Y MELIA | Ingénieur en Chef des Ponts, des Eaux et des Forêts, CNRM | Rapporteur |
| Alessandra GIANNINI | Professeur, ENS | Examinatrice |
| Raymond PIERREHUMBERT | Professeur, Université d'Oxford | Examineur |
| François RAVETTA | Professeur, Sorbonne Université | Examineur |
| Hervé LE TREUT | Professeur, Sorbonne Université | Directeur |
| Anouk BARBEROUSSE | Professeur, Sorbonne Université | Co-Directrice |
| Nicolas ROCHETIN | Maître de conférence, ENS | Co-Directeur |

Paramétrisation de la convection atmosphérique dans les modèles numériques de climat – Pratiques et enjeux épistémologiques

Résumé : Historiquement, les modèles de circulation générale (MCG) ont joué un rôle essentiel dans notre capacité à comprendre et à prédire les changements climatiques, et donc, à fournir des informations scientifiques pertinentes aux décideurs politiques. Pourtant, depuis quelques années, des critiques ont été formulées à l’encontre des MCG et du paradigme sur lequel ils reposent. En particulier, certains scientifiques ont remis en cause une composante fondamentale de tout MCG, les paramétrisations de la convection, censées représenter l’effet moyen de la convection atmosphérique à l’échelle de la grille du modèle. L’objectif de ma thèse est de mener une analyse hybride des paramétrisations de la convection, à l’interface entre les sciences de l’atmosphère et la philosophie des sciences, afin d’étudier les défis épistémologiques et pratiques que leur utilisation soulève dans les modèles de climat.

Je commence par une réflexion générale sur ce que signifie comprendre la convection atmosphérique et sur le rôle joué par l’identification de certaines structures cohérentes – que j’appelle “objets” – dans cette compréhension. Les paramétrisations de la convection reposent elles-mêmes sur certains objets et elles nous donnent ainsi un cadre pour traduire notre compréhension des phénomènes atmosphériques dans les modèles climatiques. Je décris ce cadre et j’explore ses limites. J’examine ensuite les paramètres contenus dans les paramétrisations. La valeur de la plupart de ces paramètres est peu contrainte par les observations, et doit donc être ajustée. Ce “*tuning*” des modèles climatiques comporte des similarités avec ce que l’on appelle “calibration” dans d’autres domaines scientifiques. Pourtant, je montre que le *tuning* présente des spécificités qui en font une activité unique, dont le cadre épistémologique reste à définir. Cette réflexion est illustrée par un cas d’étude, où je règle les paramètres du MCG atmosphérique LMDZ suite à l’ajout d’une nouvelle paramétrisation. A travers cet exemple concret, je donne certains critères pour définir comment évaluer les progrès apportés par une nouvelle paramétrisation dans un MCG, un exercice qui présente des difficultés conceptuelles et méthodologiques importantes. Enfin, je me penche sur l’organisation collective de la recherche climatique actuelle autour des MCG, en examinant l’une de ses pierres angulaires, le projet d’intercomparaison des modèles couplés (CMIP). Grâce à CMIP, tous les 5 ou 6 ans, la plupart des MCG utilisés dans le monde sont comparés dans des conditions similaires. Un des effets de CMIP est d’avoir permis davantage de dialogue entre les laboratoires de modélisation du climat. En retour, j’explique cependant que CMIP a créé un fossé entre les développeurs de modèle et leurs utilisateurs, et a concentré l’effort de recherche sur les analyses de modèles au détriment du développement de paramétrisations.

Parameterization of atmospheric convection in numerical climate models – Practices and epistemological challenges

Summary: Historically, general circulation models (GCMs) have played a critical role in our ability to understand and predict climate change, and thus, to provide policy-relevant scientific information to decision makers. However, in recent years, criticisms have been raised against GCMs and the paradigm on which they are based. In particular, some scientists have questioned a fundamental component of any GCM, convection parameterizations, which represent the average effect of atmospheric convection at the model grid scale. The objective of my thesis is to conduct a hybrid analysis of convection parameterizations, at the interface between atmospheric sciences and philosophy of science, in order to investigate the epistemological and practical challenges their use raises in climate models.

I begin with a general reflection on what it means to understand atmospheric convection and on the role played by the identification of certain coherent structures – which I call “objects” – in this understanding. Convection parameterizations are themselves based on certain objects and thus give us a framework to translate our understanding of atmospheric phenomena into climate models. I describe this framework and explore its limitations. I then look at the parameters contained in parameterizations. The value of most of these parameters is poorly constrained by observations, so they must be tuned. The “tuning” of climate models bears similarities with what is called “calibration” in other scientific fields. However, I show that tuning has specificities that make it a unique activity, whose epistemological framework remains to be defined. This reflection is illustrated with a case study, where I tune the parameters of the atmospheric GCM LMDZ following the addition of a new parameterization. Through this concrete example, I give criteria to define how to evaluate the progress brought by a new parameterization in a GCM, an exercise that poses important conceptual and methodological difficulties. Finally, I analyze the collective organization of today’s climate research around GCMs by examining one of its cornerstones, the Coupled Model Intercomparison Project (CMIP). Thanks to CMIP, every 5-6 years, most of the GCMs used in the world are compared under similar conditions. One of the effects of CMIP is to have enabled more dialogue between climate modeling laboratories. In turn, however, I argue that CMIP has created a gap between model developers and model users, and has focused research effort on model analyses at the expense of parameterization development.

Remerciements / Acknowledgements

Je tiens en premier lieu à remercier mon directeur de thèse Hervé Le Treut, ainsi que mes co-directeurs Anouk Barberousse et Nicolas Rochetin, pour tout ce qu'ils m'ont apporté pendant ces trois années de thèse. Merci Hervé pour la confiance et la liberté que tu n'as cessé de m'accorder. Merci Anouk pour tes commentaires enthousiastes et bienveillants, mais aussi pour ton exigence qui m'a aidé maintes et maintes fois à creuser ma réflexion. Merci Nicolas pour ton sérieux dans ton travail, et en particulier dans le suivi de ma thèse, autant que pour ton non-sérieux dans la vie en général.

Un grand merci également aux deux rapporteurs de cette thèse, Vincent Lam et David Salas y Melia, qui ont accepté de lire et d'évaluer mon travail. Merci aussi à Alessandra Giannini, Raymond Pierrehumbert et François Ravetta, qui m'ont fait l'honneur d'accepter de faire partie de mon jury de thèse.

Merci aussi à tous les membres de mon comité de thèse, Sandrine Bony, Julie Jebeile, Riwal Plougonven et Romain Roehrig, qui ont suivi avec intérêt l'avancée de mes travaux et m'ont prodigué des conseils précieux pour les mener à bien.

Je tiens à remercier particulièrement Sandrine Bony de m'avoir permis de rejoindre l'équipage de la campagne EUREC⁴A. Merci de m'avoir donné cette chance de participer à ce projet enthousiasmant et plein de sens, et de rencontrer ainsi des personnes formidables pendant ces quelques semaines à la Barbade.

Merci aussi à Riwal Plougonven, avec qui j'ai eu la chance de donner des cours pendant mes deux premières années de thèse. Merci de m'avoir fait confiance alors que je ne connaissais rien ou presque aux éoliennes et d'avoir toujours été patient dans tes explications. Merci de m'avoir transmis au-delà des connaissances, une certaine pédagogie et un désir que les élèves comprennent d'abord "avec les mains", avant de faire des calculs compliqués.

Un grand merci également à Pierre Morel, qui a créé le Laboratoire de Météorologie Dynamique (LMD) il y a un peu plus de 50 ans, et que j'ai eu la chance de rencontrer à plusieurs reprises pour une série d'interviews juste avant de débiter ma thèse. Merci de m'avoir fait découvrir votre expérience vivante de l'histoire de la météorologie, des premiers satellites, des ballons Eole et du Programme Mondial de Recherche sur le Climat.

Un grand merci à tous ceux qui m'ont fait découvrir les dessous du modèle LMDZ, notamment Jean-Yves Grandpeix, Marie-Pierre Lefebvre, Frédéric Hourdin, Catherine Rio. Votre suivi et vos conseils tout au long de ma thèse m'ont aidé à appréhender ces milliers de lignes de code et à apporter une petite pierre à la construction de ce beau projet.

J'aimerais ensuite remercier chaleureusement Anna Lea Albright et Benjamin Fildier : notre amitié qui s'est construite peu à peu et s'est surtout renforcée lorsque nous étions dans la "coloc de l'ambiance" à la Barbade m'a porté au-delà de ce que je pouvais imaginer pendant cette thèse. J'espère sincèrement que nous continuerons à cultiver cette belle amitié. Merci aussi à tous les deux pour cette soirée perdue à faire une reprise de *Blowin' in the wind* adaptée à EUREC⁴A. Merci enfin à Ben d'avoir lancé le groupe de lecture qui nous a donné une bonne excuse pour nous retrouver une fois par semaine.

Merci aussi à Caroline, qui était également dans notre bungalow à la Barbade, pour sa bonne humeur, son accessibilité, son enthousiasme et sa passion pour les sciences. Merci d’avoir de l’énergie autant pour un footing matinal à la Barbade que pour sortir ta guitare au jardin du Luxembourg, en passant bien sûr par des ateliers de vulgarisation scientifique pour les enfants, ou des conférences pointues sur l’organisation de la convection profonde.

Merci à tous ceux que je n’ai pas encore cité dans la team EUREC⁴A, notamment Raphaela Vogel, qui m’a transmis sa passion pour les poches froides, et Jessica Vial qui nous a fait découvrir ce fameux cycle diurne des petits cumulus. Merci aussi à nos collègues et amis allemands Geet George, Theresa Lang, Hauke Schulz, Sabrina Schnitt, Theresa Mieslinger, pour ne citer qu’eux. Un grand merci également à Bjorn Stevens de m’avoir fait confiance pour porter l’*outreach team* avec Anna Lea et Ben pendant la campagne EUREC⁴A. Cette confiance nous a poussés à donner le meilleur de nous-mêmes pour toutes les responsabilités qui nous avaient été confiés pendant ces quelques semaines à la Barbade. Merci aussi à Robert Pincus de nous avoir suggéré l’idée d’un *data paper* sur les profils radiatifs pendant EUREC⁴A et de nous avoir encadrés tout au long de ce projet. Un grand merci enfin à nos amis barbadiens Branden Spooner, Shanice Whitehall et Rebecca Chewitt-Lucas de nous avoir aidés à organiser en moins de deux semaines un symposium international au *Carribean Institute for Meteorology and Hydrology*. Plus généralement, merci à tous les participants à cette campagne de terrain qui ont contribué, chacun à leur manière, à cette belle aventure et à l’émulation scientifique qui en a découlé.

Merci à tous les autres collègues et amis du LMD, et d’abord mes co-bureaux successifs, Jan Vatant d’Olonne, Florentin Lemonnier, Anna Lea Albright, Antoine Bierjon et Antony Delavois, mais aussi tous ceux que j’ai pu cotoyer au coin café, en réunion climat ou ailleurs, Karine Marquois, Sara Shamekh, Ehouarn Millour, Margaux Vals, Camille Risi, Aymeric Spiga, Jean-Louis Dufresne, Geneviève Sèze, Martin Turbet, Sandrine Guerlet, Deborah Bardet, Yuan Zhang, Frédérique Cheruy, Philippe Drobinski, Andre Szantai, Sebastien Lebonnois, Vladimir Zakharov, Max Popp, Alexandre Boissinot, Venance Journée, Laurent Li, Jean-Baptiste Madeleine, Etienne Vignon, Binta Diallo, Marion Saint-Lu, Addisu Semie, Yanfeng Zhao, Hugo Bellenger, Evelyne François-Ellocie, Thomas Pierron, Joseph Naar, Adriana Sima et tous les autres !

J’aimerais aussi remercier plus personnellement mes amis – et notamment Nico, Théo, Julie et Jean-Nicolas pour ne citer qu’eux – ainsi que mes proches, mes parents, mon frère Emmanuel et ma sœur Anne-Solveig. Merci à chacun de m’avoir soutenu tout au long de ma thèse. Un grand merci aussi à ma grand-mère, qui m’a hébergé plusieurs fois pendant ma thèse, dans un cadre reposant et propice à l’écriture.

Enfin, merci à Manon et au petit bébé qui devrait nous rejoindre à la fin de l’année, pour toute la joie qu’ils me donnent au quotidien.

Contents

| | | |
|----------|---|------------|
| 1 | Introduction | 9 |
| 1.1 | Motivations | 9 |
| 1.2 | Atmospheric convection and its role for climate | 11 |
| 1.3 | Scientific tools to study atmospheric convection | 16 |
| 1.4 | The parameterization debate | 20 |
| 1.5 | The perspective of philosophers of science | 24 |
| 1.6 | Organization of the manuscript | 28 |
| 2 | Nature and role of objects | 31 |
| 2.1 | Introduction | 32 |
| 2.2 | Definition and characteristics of objects | 33 |
| 2.3 | Objects in observations | 39 |
| 2.4 | Objects in computer simulations | 43 |
| 2.5 | The example of cold pools | 52 |
| 2.6 | Conclusion | 63 |
| 3 | The semi-empirical world of convection parameterizations | 65 |
| 3.1 | Introduction | 66 |
| 3.2 | Why do we need to parameterize atmospheric convection? | 67 |
| 3.3 | In pursuit of a theoretical foundation for mass-flux convection parameterizations | 71 |
| 3.4 | Object-based interpretation of LMDZ convection parameterizations | 82 |
| 3.5 | Conclusion | 90 |
| 4 | Epistemology of climate model tuning | 91 |
| 4.1 | Introduction | 92 |
| 4.2 | Principle and methods of climate model tuning | 93 |
| 4.3 | Climate model tuning vs calibration | 100 |
| 4.4 | Some terminology – Compensating errors, structural error and overtuning | 104 |
| 4.5 | Automatic tuning methods: a shift in tuning paradigm? | 108 |
| 4.6 | Conclusion | 111 |
| 5 | Parameterization development and tuning in practice | 113 |
| 5.1 | Introduction | 113 |
| 5.2 | Definition of progress for parameterizations | 115 |
| 5.3 | Presentation of the cloud and precipitation overlap parameterization | 117 |
| 5.4 | Results | 124 |
| 5.5 | Conclusion | 144 |

| | | |
|----------|---|------------|
| 6 | The Coupled Model Intercomparison Project | 147 |
| 6.1 | Preamble | 147 |
| 6.2 | Abstract | 148 |
| 6.3 | Introduction | 148 |
| 6.4 | History of climate model intercomparison projects | 149 |
| 6.5 | Uses of CMIP results and related controversies | 154 |
| 6.6 | Structural effects of CMIP on climate research | 160 |
| 6.7 | Conclusion | 162 |
| 7 | Conclusion | 165 |
| 7.1 | Summary | 165 |
| 7.2 | Perspectives | 167 |
| A | Publication list | 173 |
| B | Detecting cold pools from soundings during EUREC⁴A | 175 |
| B.1 | Preamble | 175 |
| B.2 | Abstract | 175 |
| B.3 | Introduction | 176 |
| B.4 | Radiosonde and dropsonde data | 177 |
| B.5 | Presentation of the method and test in a high-resolution simulation | 178 |
| B.6 | Application to EUREC ⁴ A data | 182 |
| B.7 | Conclusions | 193 |
| B.8 | Data availability | 194 |
| C | Atmospheric radiative profiles during EUREC⁴A | 195 |
| C.1 | Preamble | 195 |
| C.2 | Abstract | 196 |
| C.3 | Introduction | 196 |
| C.4 | Data and methods | 197 |
| C.5 | Preliminary results and discussion | 201 |
| C.6 | Uncertainty assessment | 209 |
| C.7 | Conclusions | 211 |
| C.8 | Code and data availability | 211 |
| D | Résumé long en français | 213 |

Chapter 1

Introduction

Contents

| | | |
|------------|---|-----------|
| 1.1 | Motivations | 9 |
| 1.2 | Atmospheric convection and its role for climate | 11 |
| 1.2.1 | Physics of convection | 11 |
| 1.2.2 | Importance of convection in the climate system | 13 |
| 1.3 | Scientific tools to study atmospheric convection | 16 |
| 1.3.1 | Observations | 16 |
| 1.3.2 | Numerical simulations | 16 |
| 1.4 | The parameterization debate | 20 |
| 1.4.1 | The need for epistemology in climate modeling | 20 |
| 1.4.2 | Convection parameterizations being questioned | 22 |
| 1.5 | The perspective of philosophers of science | 24 |
| 1.5.1 | Introduction to the philosophy of climate science | 24 |
| 1.5.2 | The (few) epistemological studies on parameterizations | 26 |
| 1.6 | Organization of the manuscript | 28 |

1.1 Motivations

We are conducting the most important and dangerous scientific experiment ever: changing the climate of our planet. The concentration of carbon dioxide (CO_2) in the atmosphere has been increasing steadily over the past century and is now at a level never reached in the last million years. This is likely not only to increase the Earth's surface temperature, but also change the intensity and distribution of precipitation, raise sea levels, cause glaciers to shrink, thaw permafrost, etc. In turn, these changes will affect all ecological systems and all forms of life, including our human societies. As we suffer their consequences, we will learn the results of this large-scale scientific experiment in which we are, whether we like it or not, all involved.

For obvious reasons, rather than waiting patiently for disasters to come, we prefer to study and understand the response of climate to an increase of CO_2 and eventually, to predict what lies ahead. As we do not have twins of the Earth on which we could carry out experiments

without undergoing the consequences, we use instead digital twins, i.e. numerical models. Since the advent of the computer age in the 60s, different types of numerical models have been used to simulate our climate system or some of its components. Historically, these models have played a key role in our progressive understanding of the climate system and its response to an increase in atmospheric CO₂ concentration. While useful, they have many uncertainties and should not be mistaken for the truth. Distinguishing legitimate uses from misuses of climate models is a challenging task and there has been growing debates among philosophers of science about the conceptual, methodological and epistemological issues associated with climate models, with questions such as: what is our confidence in these models based on? In which sense are they validated by observational data? How to characterize their uncertainties? Are climate models “fit-for-purpose”, i.e. good enough for the use we make of them?

When I arrived at the Laboratoire de Météorologie Dynamique (LMD) in September 2018 to begin a Ph.D. thesis at the interface between climate science and philosophy, I had these questions in mind. I wanted to firmly establish the epistemological foundations of models used in climate science in order to clarify their scope and purpose. Back then, I was concerned that my interest in philosophical questions about climate models and their role would not be shared, or even understood, by the scientists of my lab. What is the point of adding a philosophical layer on top of climate science? Isn’t there enough real scientific work to be done in the field? Why should a student like me waste his time with philosophical reflections that do not really advance science?

I quickly realized that these fears were unfounded. On the contrary, the scientists of my lab had asked themselves these questions before me. My first job was thus to listen what they had to say, in order to understand all these reflections which were rarely published in the scientific literature, but sparked debate at every coffee break, team meeting or conference. While there is a consensus among climate modelers on the reality of anthropogenic climate change, there is significant disagreement on how this change should be predicted and studied. During my Ph.D., I was confronted with various conceptions of models and their role, and with many discussions about the meaning of “doing good science” with models. These discussions were sometimes tense and had immediate consequences on the methods used by the scientists around me in their daily research.

A recurring topic concerned the standard way of representing atmospheric convection in climate models through “convection parameterizations”. The latter rely on various assumptions to account for the average effect of phenomena at the origin of clouds and precipitation in climate models without resolving them at the model grid scale. On the one hand, there are the opponents of convection parameterizations, who consider that the climate modeling community has wasted enough time with parameterizations and propose alternatives to replace them. On the other hand, there are their defenders, who are convinced that parameterizations are still the right way to go and that efforts should be focused on improving them. The debate about parameterizations is partly technical and concerns in particular the quality of the results that parameterizations and their alternatives produce in climate models. However, it is also partly epistemological: the foundations of parameterizations as well as their use in climate models are indeed often questioned and described as not entirely justified.

Epistemological questions about convection parameterizations are the central topic of this Ph.D. thesis. My approach will be at the interface between science and philosophy; I had the

chance, during my Ph.D. thesis, to participate in several scientific works that gave me a close look at various facets of the study of atmospheric convection. These different works have nourished my reflection and are used throughout my thesis as examples, case studies or pretexts for epistemological discussions on parameterizations and their foundations. I have thus built my thesis on three pillars: first, a critical study of the literature in philosophy of science and climate science on issues related to parameterizations. Second, the many informal discussions I had with the scientists I worked with during my thesis. Third, the various scientific works I conducted that allowed me to study in practice the methodological difficulties posed by the study and modeling of atmospheric convection.

In the rest of this introductory chapter, I provide the reader with some key elements to understand the content and relevance of the work presented in this manuscript. I start with a general introduction on atmospheric convection and its role for climate (Section 1.2), and present the different tools used to study atmospheric convection, in particular observations and numerical simulations (Section 1.3). I then justify the need for epistemology in climate modeling and introduce the parameterization debate (Section 1.4), before giving a brief overview of the relevant literature in philosophy of climate science (Section 1.5). Finally, I present the general organization of the rest of the manuscript (Section 1.6).

1.2 Atmospheric convection and its role for climate

1.2.1 Physics of convection

Convection refers to the transport of heat by the macroscopic movement of a fluid. It is one of the three modes of heat exchange between two physical systems, along with conduction and radiation. The manifestations of convection in our daily lives are countless: every time we see smoke rising from a chimney, or observe the movement of liquid water in a boiling pot, we are witnessing a convective phenomenon. What is less known, though, is that convection is also at the origin of most clouds, from small fair weather cumuli to massive cumulonimbus that darken the sky.

Atmospheric convection is the dominant form of vertical energy transport in the atmosphere. It can be driven solely by buoyancy (buoyancy-driven or free convection) or by a combination of buoyancy and mechanical force (forced convection). Buoyancy, according to Archimedes' principle, comes from a density difference between an air parcel and its environment. When a fluid is lighter than the fluid around it, it rises; when it is heavier, it sinks. At constant pressure, due to ideal gas law, the density of an air parcel is inversely proportional to its temperature. Heating close to the ground – for instance, due to the absorption of solar radiation by the ground, or to the presence of a warm sea surface temperature – might therefore trigger convection over the lowest layers. Forced convection happens when an obstacle – for instance, a mountain – forces air masses to rise. In this case, the horizontal kinetic energy of the flow is converted into vertical kinetic energy. A cold front – that is, the leading edge of a relatively cold air mass – might also behave in some situations as a material obstacle and help the lifting of surrounding air masses.

When an air parcel rises to lower pressure levels, its volume expands. Since the conductivity of the air is low, heat exchange between the rising air parcel and its environment is usually neglected, and this expansion is described as adiabatic. Nevertheless, the decrease in pressure



Figure 1.1: Different types of cumulus clouds: a) *Cumulus humilis*, b) *Cumulus mediocris*, c) *Cumulus congestus*, d) *Cumulonimbus*. Pictures taken from the official website of the World Meteorological Organization's (WMO) International Cloud Atlas (<https://cloudatlas.wmo.int/en/home.html>).

causes a decrease in temperature and the air becomes colder and colder as it goes to higher altitudes. When the air is dry, the rate of temperature decrease of a rising air parcel is generally stronger than that of the environment: after a few hundreds of meters, the rising parcel becomes colder, thus denser than its environment. It explains why dry convection stays generally confined over the lowest layers of the atmosphere. On the contrary, when the air is moist, i.e. when it contains water vapor, it may become saturated at a certain altitude. Indeed, according to the Clausius-Clapeyron relation, the amount of water vapor that a given parcel of air can contain depends on its temperature: the warmer the air, the more water vapor it can have. As an air parcel rises in the sky, it gets progressively colder and its temperature may thus reach its saturation temperature – the temperature for which the first liquid droplets appear. While water vapor is invisible, these small water droplets floating in the sky reflect the sunlight and are thus visible: they form what is called a cloud.

The condensation of water vapor releases some heat. This warms the saturated air parcel relative to its environment and makes it rise further. The apparition of liquid water thus fuels convection and helps it to grow. The altitude at which a rising air parcel stops depends mainly on its initial temperature and water vapor mixing ratio, but also on a process called *entrainment* – namely how much the parcel mixes with the surrounding air as it rises. In fact, a rising air parcel entrains some environmental air. If this entrained air is dry, it dilutes the water vapor the parcel contains, thus reduces the available latent heat to feed convection. Isolated air parcels in a dry environment generally give rise to fair weather cumulus clouds such as *cumulus humilis* or *cumulus mediocris* (see Fig. 1.1 a and b), which stays confined in the first two kilometers of the atmosphere. When the air is moister, air parcels may reach the

mid-levels of the atmosphere, up to 6 km and give birth to congestus (Fig. 1.1 c), or through the whole troposphere, up to 12 km or more, and form cumulonimbus (Fig. 1.1 d). At such altitudes, temperatures become negative and some of the parcel liquid droplets turn into ice crystals, releasing additional heat that sustains the ascent of the parcel.

Clouds are filled with tons of water droplets or ice crystals. Nevertheless, they float in the sky due to the very small size of the liquid droplets or ice crystals that compose them. Typically, the radius of water droplet or ice crystals ranges from a few microns (10^{-6} m) to a few tens of microns. The fall speed of any object is related to its mass and surface area, as can be seen by comparing the falling speed of the same mass of feathers and lead. The mass of a cloud droplet scales with its volume, thus the cube of its radius, whereas its surface scales with the square of its radius. The smaller the radius of a cloud droplet, the larger its surface/volume ratio, and thus the slower its fall. Typically, the velocity of cloud droplets which are about 10 microns in radius is only 1 cm/s, which means that it takes them more than a day to go through the atmosphere before reaching the surface!

In practice, droplets do not fall at all, but on the contrary rise. Such small fall velocities are indeed largely offset by the rising air motions – so-called, the updrafts – which support the clouds. Nevertheless, when the density of cloud droplets becomes large enough, various processes of collision and coalescence make them merge and form larger droplets. Typically, a rain drop of 2 mm contains the water equivalent to several millions of cloud droplets. The largest droplets may then fall at a sufficient speed to reach the ground without being fully evaporated. Depending on the conditions in which they were formed, these drops affect us under the form of rain or solid crystals such as snow or hail.

1.2.2 Importance of convection in the climate system

Clouds and climate

Why do we care – or should we care – about clouds? Not only because they make us stay at home when it rains, but also because by absorbing and reflecting electromagnetic radiation, clouds have a decisive effect on the temperatures of our planet.

Cloud radiative effects To explain the role of clouds on the climate, we must go back to what determines the temperatures at the surface of the Earth: the radiative budget, i.e. the balance between the energy coming from the Sun and the one re-emitted by our planet to space. Incoming solar radiation is mostly in the ultraviolet, visible and near-infrared part of the energy spectrum, so-called “shortwave radiation” (SW). As illustrated in Fig. 1.2, clouds reflect part of this SW radiation back to space. They thereby deprive the Earth of a fraction of solar energy, which lowers its temperature. This cooling due to the reflection of SW radiation is the first effect of clouds on the climate. To quantify this effect, one has to calculate the amount of heat that would have been absorbed by the Earth in the absence of clouds. It is not an easy task because the surface of our planet also reflects part of solar radiation. The amount of energy returned, called the “albedo” depends largely on the properties of the surface receiving sunlight. The oceans, darker than the ice, have for instance a lower albedo than the ice sheets. To determine the total radiative effect of clouds, it is necessary to take into account both the characteristics of clouds and their location: for example, a cloud located over the oceans will tend to deprive the Earth of more energy than a cloud located over the poles.

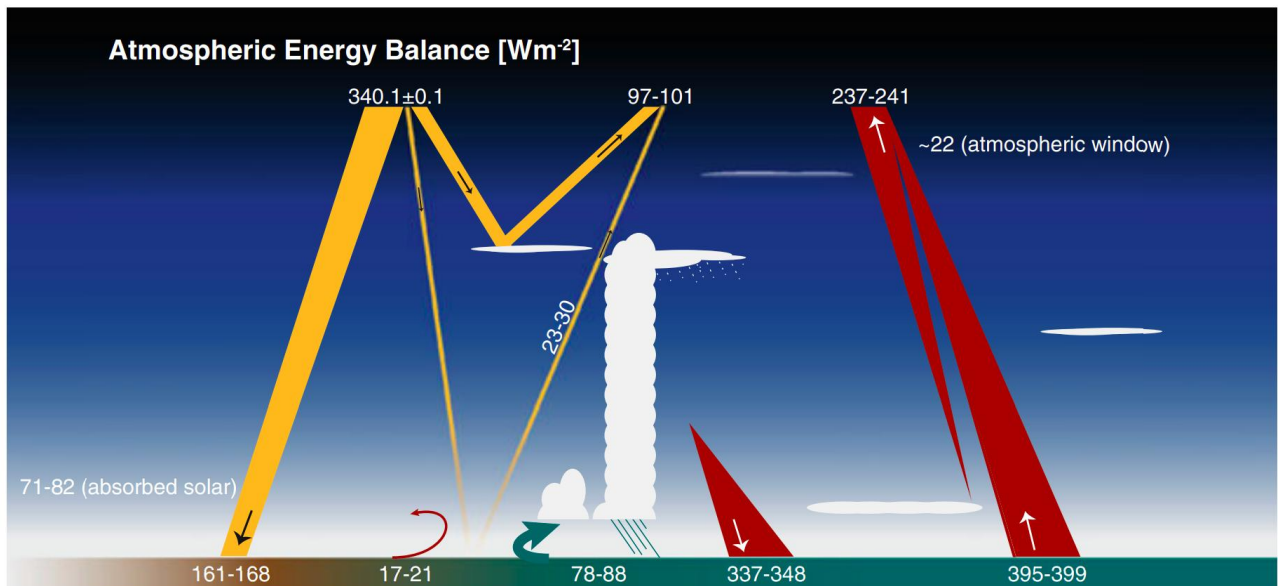


Figure 1.2: From Stevens and Schwartz (2012). Earth’s mean top-of-atmosphere (TOA) and surface energy budget. In yellow: incoming shortwave radiation and how it is reflected and absorbed by clouds, the atmosphere and the Earth’s surface. In red: longwave radiation emitted by the Earth’s surface and atmosphere. The green and red arrows near the surface correspond respectively to the surface sensible and latent heat fluxes. The values given for the different fluxes refer to the likely range subjectively determined by Stevens and Schwartz (2012) based on a review of the literature and global simulations.

In total, about 30% of the solar energy is reflected back to space. The remaining energy is absorbed by the Earth’s surface or atmosphere. Any heated body dissipates part of its thermal energy by emitting radiation whose wavelength depends on its temperature. The Earth is no exception: in its temperature range, it is mostly infrared radiation that is emitted – we will talk of “longwave radiation” (LW). Depending on their thickness and composition, clouds absorb a certain percentage of this radiation. They then re-emit it in all directions and part of it returns to the Earth (see Fig. 1.2). Clouds thus prevent a certain fraction of the infrared radiation emitted by the Earth from escaping into space. This is the second effect of clouds on the climate. Contrary to the first one, this second effect leads to a warming of the surface temperatures of our planet.

There are therefore two competing effects of clouds on the climate: on the one hand, they cool the Earth’s surface by reflecting solar radiation, and on the other, they warm it by trapping infrared radiation. The net impact of these cloud radiative effects (CRE) depends on many properties, including the location on Earth where they are formed, their height, their altitude, their microphysical properties, etc. In the current climate, satellite data show that the global and annual mean SW CRE is approximately -50 W/m^2 and the mean LW CRE, about 30 W/m^2 (Boucher et al., 2013). This results in a net global mean CRE of about -20 W/m^2 , thus a strong net cooling effect of clouds on the current climate.

Cloud feedback An important question is whether global warming will change this net CRE. Indeed, the distribution and type of clouds are expected to change with global warming, which could lead to an increase or decrease in net CRE. For instance, if the net cooling effect of

clouds is enhanced, this will compensate part of the global warming induced by an increase of CO₂ in the atmosphere – what is called a negative cloud feedback. On the contrary, if the cooling effect of clouds is weakened, the temperature of the Earth will rise even more, resulting in a positive cloud feedback. Assessing the magnitude of cloud feedback is the focus on many ongoing scientific studies. According to the last report of the Intergovernmental Panel on Climate Change (IPCC), the net cloud feedback is expected to be slightly positive, around 0.6 W m⁻² °C⁻¹ (Stocker et al., 2013). However, there are still many unknowns about the response of clouds to rising global temperature, thus on the amplitude of cloud feedback (S. Sherwood et al., 2020). Much of the uncertainty about future climate change is hence related to clouds.

Precipitation

Precipitation is another important effect of convection on the climate. The average residence time of water in the atmosphere is about 10 days (Ent & Tuinenburg, 2017), and at any time, part of the water present in the atmosphere returns to the surface in the form of precipitation. The location and intensity of precipitation determines how fresh water is distributed over land, and in particular the distribution of rivers, lakes and watersheds. Like many other species, we cannot live without fresh water, and humanity has therefore developed mainly where there are water resources. Precipitation also has a direct impact on soil moisture and on the plant species that grow in it, with once again consequences on our human societies. Too little precipitation can lead to droughts that make the soil unusable for farming. On the contrary, extreme precipitation is associated with natural disasters such as heavy storms or floods. Predicting the consequences of global warming on the distribution and intensity of precipitation around the Earth is therefore crucial for assessing the impact of climate change on humans.

Various modeling studies have shown that the Earth’s water cycle will intensify under global warming, leading to an increase of globally-averaged precipitation of about 1 to 3 % °C⁻¹ (O’Gorman & Schneider, 2009; Richter & Xie, 2008; Trenberth, 2011). Despite considerable regional variability, there is an overall tendency for a “wet-get-wetter” and “dry-get-drier” response over oceans. It can be understood because atmospheric circulations tend to import moisture from dry regions to moist regions. As these circulations are not expected to change significantly in a warmer climate, but the atmosphere is expected to be moister on average (I. M. Held & Soden, 2006; Mitchell, Wilson, & Cunningham, 1987), the wet regions will tend to import more moisture from dry regions. This result that have been confirmed by various numerical simulations (Chou, Neelin, Chen, & Tu, 2009; Muller, O’Gorman, & Back, 2011) and analyses of 20th century precipitation trends (Allan & Soden, 2007; Zhang et al., 2007). Local changes in precipitation are however more uncertain as they might depend on regional circulation shifts induced by global warming (Joshi, Gregory, Webb, Sexton, & Johns, 2008; Xie et al., 2010). There are also major uncertainties about the precipitation response over land due to the difficulties in predicting accurately moisture–precipitation feedbacks (Hohenegger, Brockhaus, Bretherton, & Schär, 2009).

The importance of clouds and their uncertain response to climate change motivates many scientific studies about atmospheric convection. We give an overview of the tools that are used in these studies in the next section.

1.3 Scientific tools to study atmospheric convection

1.3.1 Observations

To study atmospheric convection, the most natural method is probably to observe what happens in nature, what we will call “the natural phenomena”. Three main sources of observational data can be used to study atmospheric convection (Guillemot, 2009).

Observing networks

Observing networks consist mainly of weather stations over lands and buoys over oceans. Each station provides continuous data at one precise location. Twice a day, radiosondes are launched in some of these stations in order to measure profiles of pressure, wind, temperature and humidity in the atmosphere. The data collected are used to initialize weather forecast models, but can also help to better understand atmospheric convection.

Field campaigns

A field campaign is a period of intensive measurements over a targeted region, usually focused on one or a few specific phenomena. Over the past 50 years, many field campaigns have been organized to understand atmospheric convection. As part of my Ph.D., I had the opportunity to participate in the field campaign EUREC⁴A (Elucidating the role of clouds-circulation coupling in climate) which took place in January and February 2020 in the trade-wind environment east of Barbados (Bony et al., 2017, see also appendix B and C). A field campaign is interesting in that it allows to cross the measurements of different instruments offering complementary views of a given phenomenon. Panel a in Fig. 1.3 shows for instance the many instruments that have been used during EUREC⁴A, including four aircraft, four research vessels, a scanning precipitation radar, several drones, etc. More than two thousand atmospheric profiles have also been measured during the field campaign using radiosondes and dropsondes.

Satellite observations

While the observations of field campaigns give precise information, localized in space and time, satellites ensure a constant monitoring of the atmosphere all around the globe. Polar satellites evolve at typically 700 to 800 km above the ground and circle the entire surface of the Earth several times a day. Geostationary satellites evolve at about 36000 km from the Earth and, as their name suggest, remain above a given location on our planet, which allows for instance to follow the evolution of cloud fields in time. Equipped with various passive and active sensors, these two types of satellites give precious information about many atmospheric properties relevant for the study of convection such as microphysical and macrophysical properties of clouds, winds, atmospheric water vapor content, etc.

1.3.2 Numerical simulations

As data from observations are often incomplete, atmospheric scientists use numerical models to complement them. These models are based on fundamental equations of fluid dynamics, discretized over a given grid at a given resolution. By definition, they will not be able to adequately simulate phenomena that occur at a scale smaller than the model grid, nor phenomena that cannot be directly described by the equation of fluid dynamics – radiative processes for instance. To take their average effect at the scale of the grid into account, scientists include

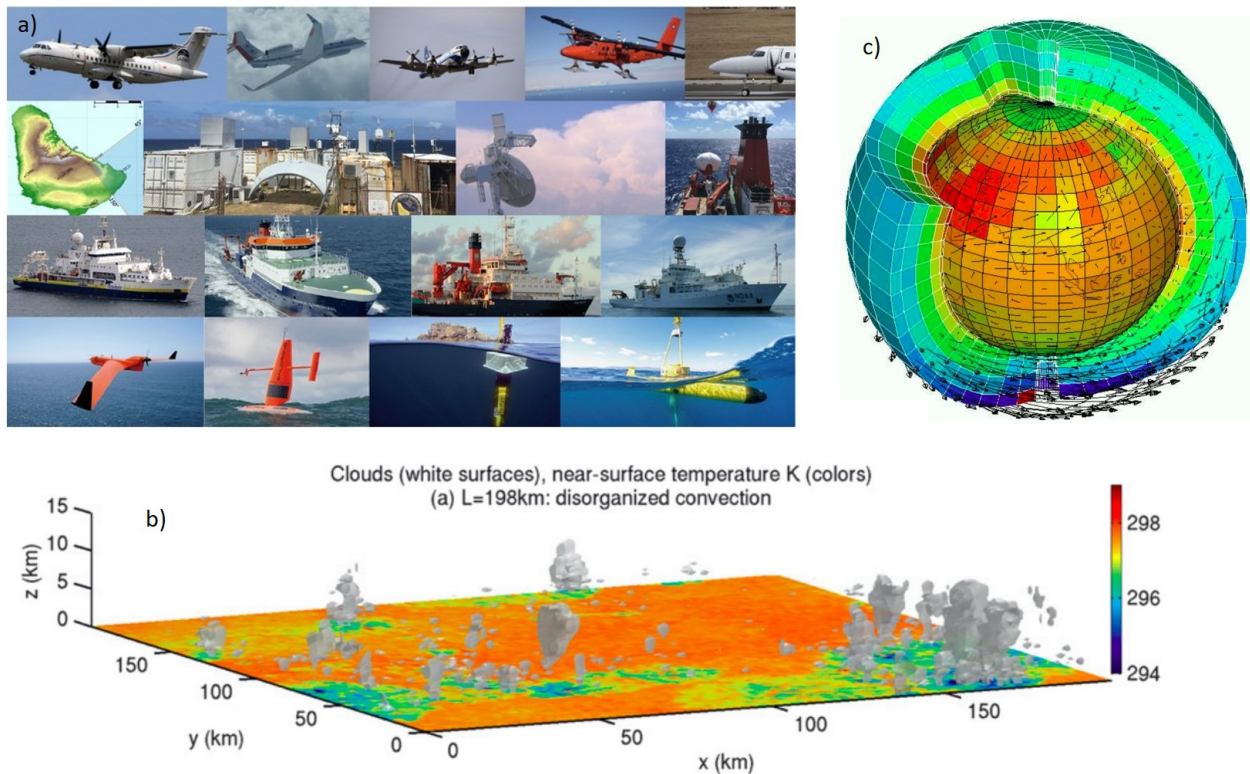


Figure 1.3: Scientific tools to study atmospheric convection a) Various platforms used to collect observations during the EUREC⁴A field campaign (EUREC⁴A official website, 2021) b) Instantaneous snapshots of clouds and near surface temperature in a CRM simulation (Muller & Held, 2012) c) Representation of horizontal and vertical grids of a GCM. Colors represent grid-box temperatures, and arrows indicate wind direction (credit: L. Fairhead, ISPL).

parameterizations. Depending on the resolution of the model, more or less phenomena need to be parameterized.

Large-eddy simulations

The typical resolution of large-eddy simulations (LES) ranges from a few tens of meters to a few hundreds of meters. At such resolution, it is necessary to parameterize 1) the turbulence, in order to represent the many small eddies that mix the air, but which are too small to be resolved at the scale of the model grid, 2) the microphysics, i.e. the different processes associated with cloud particles and precipitation (commonly referred to as “hydrometeors”), and 3) radiative transfer in the atmosphere. Conversely, at the LES grid scale, convective updrafts are relatively well described and do not need to be parameterized. LES are able to represent *explicitly* – that is, without using parameterization – the life cycle of shallow cumulus such as cumulus humilis, mediocris, stratocumulus, or deeper convective clouds such as congestus or cumulonimbus. The relatively fine resolution of LES nevertheless comes at a price: due to their computational cost, LES simulations can generally only be performed on domains of limited size, from about 5 km to 100 km, and for relatively short periods of time, a few days at most.

Cloud-resolving models

Various names have been given to models with a kilometric resolution: “cumulus ensemble models” (Sui, Lau, Tao, & Simpson, 1994; K.-M. Xu, Arakawa, & Krueger, 1992), “cloud system resolving models” (V. T. Phillips, Donner, & Garner, 2007) or, more simply, “cloud-resolving models (CRMs)” (Gao, Cui, Zhou, & Li, 2005; McCaul Jr, Goodman, LaCasse, & Cecil, 2009). We choose to use this last name here. CRMs and LES are based on similar equations. The main difference is that the resolution is coarser in CRMs than in LES, which allows them to be run over larger domains and for longer duration. Recently, some global simulations using CRMs have even been conducted for several weeks (Satoh et al., 2019; Stevens, Acquistapace, et al., 2020). Despite their coarse resolution compared to LES, CRMs are able to resolve many important features of convection (Prein et al., 2015; Weisman, Skamarock, & Klemp, 1997). However, shallow clouds are not simulated with the same accuracy as in LES. This led some modeling groups – but not all – to adopt subgrid-scale parameterizations of shallow convective clouds, in addition to the parameterizations of turbulence, microphysics and radiation also present in LES (Bogenschutz & Krueger, 2013).

General circulation models

Historically, atmospheric general circulation models (GCMs) have been used both for weather forecasting and for climate projections. In the latter case, atmospheric GCMs need to be coupled to an oceanic GCM and various other components of the climate system, such as land and sea-ice. The acronym GCM can then be used to mean Global Climate Model, or Global Coupled Model. The term Earth System Models (ESMs) is generally preferred when a Global Climate Model takes into account interactions with the biosphere, and in particular the carbon cycle. In the following, when we speak of GCMs, we implicit mean atmospheric GCMs, whether or not they are integrated into a Global Climate Model or an Earth System Model.

GCMs have a resolution much coarser than LES or CRMs, typically from 30-300km. At such resolution, an important difference with LES and CRMs is that it is possible to simplify the equations of fluid dynamics using the hydrostatic approximation, which assumes a balance between the gravity force and the pressure force in the vertical direction, and neglects the vertical acceleration due to buoyancy fluctuations. The resulting equations, so-called “primitive equations”, are discretized over the grid of the model. They form the “dynamics” of the GCM, as opposed to the “physics” which contains all the parameterizations. At each time step, typically every 10 minutes, the state variables of the model – that is the temperature, the specific humidity, the pressure, and the wind vectors – are calculated by the dynamics at all points of the 3D grid of the model.

Due to the coarse grid used in GCMs, many phenomena cannot be adequately represented at the model grid scale and have to be parameterized, i.e. their average effect on the state variables of the model must be calculated. In particular, contrary to LES and CRMs, GCM do not explicitly resolve clouds and convection and use parameterizations to take their effect into account. Over the past 50 years, different convection parameterizations for use in GCMs have been proposed (e.g. Arakawa & Schubert, 1974; DelGenio & Yao, 1993; K. A. Emanuel & Živković-Rothman, 1999). Due to the importance of atmospheric convection for the climate system, these convection parameterizations are a key element of any GCM. Consequently, most of the differences between GCM results come from differences in their convection parameterizations (Rio, Del Genio, & Hourdin, 2019).

Assessing convection parameterizations using SCM case studies To evaluate convection parameterizations, a paradigm has progressively emerged using so-called Single-Column Models (SCMs) (Browning, 1994; D. Randall, Khairoutdinov, Arakawa, & Grabowski, 2003). As its name suggests, a SCM can be seen as a single column of a GCM, taken in isolation from the neighboring columns. A SCM contains the same parameterizations as the full GCM, but gets rid of its large-scale dynamics. Its main advantage is that it allows to test parameterizations in a simplified framework. This comparison can be done at two levels. First, it is possible to test a SCM directly with observational data from a field campaign. This testing involves three different steps: 1) Initializing the SCM with observations. 2) Forcing the SCM with observations, that is to say applying at each time step a correction to make it respect some large-scale forcings derived from observations. 3) Comparing the results produced by the SCM in response to this observed forcing with other observations collected in the field, observations of cloudiness for instance.

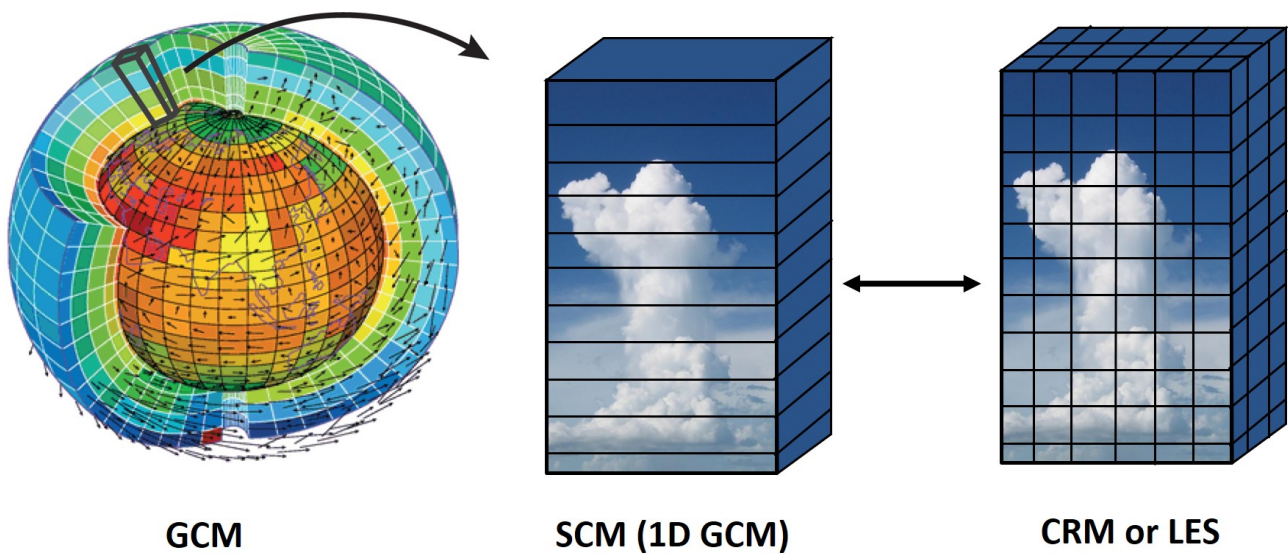


Figure 1.4: Diagram explaining how CRM or LES can be used to test a GCM through the use of its SCM version. SCM and CRM/LES results can be directly compared because both models are initialized and forced with the same boundary conditions.

Yet, in practice, it is difficult to extract the relevant information from observations to assess SCM results (step 3). Some important diagnostics for parameterizations, such as cloudiness at different levels or mass flux at cloud base, cannot be easily inferred from observations. It motivates the use of CRMs or LES, which can be initialized and forced (step 1 and 2) in exactly the same way as SCMs (see Fig. 1.4). The comparison (step 3) between models and observations is also easier for CRMs/LES than for SCMs. Indeed, as CRMs/LES explicitly resolve clouds, they are one step closer to the observing world. If we use the same initialization and forcing (step 1 and 2) for a SCM and a CRM/LES, the comparison between the two models is then straightforward: the CRM/LES results give comprehensive information about four-dimensional fields of temperature, water vapor, water liquid, ice and momentum which can then be compared directly to SCM results.

Progressively, observational data from various field campaigns have been used to develop different case studies comparing SCMs with CRMs/LES and observations. Early SCM stu-

dies included the FIRE case [First ISCCP (International Satellite Cloud Climatology Project) Regional Experiment] to study the stratocumulus-topped boundary layer (Bechtold et al., 1996; Zhu et al., 2005), the ASTEX case (Atlantic Stratocumulus Transition Experiment, C. S. Bretherton et al., 1999), to study cloudiness transition between stratocumulus and subtropical trade cumulus clouds, the ARM case (Atmospheric Radiation Measurement, Guichard et al., 2004; Lenderink et al., 2004), to study the diurnal cycle of convection over land, or the more recent RICO case (Rain In shallow Cumulus over the Ocean), based on a field campaign devoted to the study of the formation of rain in shallow cumuli over the Atlantic ocean (Raubert et al., 2007). With these many case studies, the comparison between SCMs, CRMs and observations has progressively become the dominant paradigm to evaluate and improve convection parameterizations.

1.4 The parameterization debate – or why do climate modelers care about epistemology?

What are the links between climate modeling and epistemology? According to the *Encyclopedia Britannica*, epistemology is “the philosophical study of the nature, origin, and limits of human knowledge” (Martinich & Stroll, 2021). The scope of epistemology is thus extremely broad: it attempts to study not only scientific knowledge, but any kind of knowledge. However, in the French tradition, the term epistemology is generally more specific and focuses exclusively on scientific knowledge (e.g. Berthelot, 2018; Le Moigne, 1990; Mucchielli, 2000; Nadeau, 1999). It could then be defined as the branch of philosophy of science that critically examines the foundations on which different scientific disciplines are based as well as the methods and principles they use, in order to determine the value and scope of the results they produce. This definition seems the most appropriate to describe the work done in this thesis and we use the term “epistemology” in this sense in the rest of the manuscript.

With this definition, scientists engage in epistemology when they reflect on their own scientific practices, on the methods they use, or discuss the validity of their results. As we will see, this is common in climate science. Many climate modelers have conducted in-depth epistemological reflections on their own work. Epistemology is thus not a meta-discipline, which would pose its own philosophical problems on top of climate modeling. On the contrary, many epistemological problems are at the heart of the daily work of climate modelers, especially those working on parameterizations. Epistemological reflections are needed to define what is meant by “good science” when working with GCMs and orient climate modeling research accordingly.

1.4.1 The need for epistemology in climate modeling

Many epistemological studies in climate science have been conducted by climate scientists themselves. There are two main reasons for this:

- first, **the appearance of climate science in the public sphere**. Faced with the risks of future climate change, many voices have risen in the scientific community to alert decision makers and the general public. The alert was given most notably through the IPCC reports, which provide policy makers with regular scientific assessments of climate change and its potential impacts. This visibility in the public sphere has exposed climate scientists to criticism. In the United States in particular, many influential personalities, sometimes even recognized scientists, claimed to be *climate-skeptics* and questioned the

results of the IPCC reports (Dunlap, 2013; Hoffman, 2011). At the core of some of these criticisms, there is the suspicion that future climate projections are based on numerical models, giving a biased picture of reality (Norton & Suppe, 2001). To answer these criticisms, scientists had to conduct an epistemological reflection to justify where their confidence in numerical climate models came from.

- secondly, and more fundamentally, **the increasing use of numerical models in science** challenges existing epistemological frameworks and raises new questions (Jebeile, 2013) such as: which kind of understanding do numerical models provide? how do we know whether a numerical model is “good enough” for the use we want to make of it? Reciprocally, how can we prove a model is wrong? How to assess the conditions of validity of a numerical model? Such questions strike climate modelers in their own scientific practice and lead them to epistemological reflections in order to justify the choices they make and the methods they use.

As a consequence, there are many instances in the scientific literature of atmospheric physicists questioning the methods used in their own field of research. For instance, Stevens and Lenschow (2001) attempt “to begin a dialogue on the philosophy of simulation”, with a focus on LES. In this paper, important epistemological questions are studied such as: how can we validate a LES using observations? What would be a critical test for LES? What role can LES play to complement observations and experiments? In particular, Stevens and Lenschow (2001) argue that the comparisons between LES and observations are most of the time uncritical – the observations used have not been designed explicitly to test LES and as a result, they may provide at best qualitative statements about the fidelity of LES. They claim that explicit criteria should be used to distinguish between the success and failure of a LES when it is confronted to an observational test.

Many other scientific papers are focused on GCMs, and try to address the challenges of interpreting and understanding such complex models. Again, many epistemological questions are raised: where does our confidence in GCMs come from? What criteria can be used to guide their development? What kind of understanding is enabled by such numerical simulations? At what point does a simulation become too complex to be useful? I. Held (2005) defines for instance the understanding of a complex system as the ability to relate its behavior to that of simpler systems. It leads him to claim that a hierarchy of models of increasing complexity is needed to make sense of the most complex GCMs. This claim is shared by Bony et al. (2013), which point out that the understanding of the content of GCMs is crucial because, unlike models used for numerical weather prediction, climate models cannot be accurately assessed using observations. According to the authors, our confidence in model projections critically depends on our physical understanding of the principles on which they are built. Jeevanjee, Hassan-zadeh, Hill, and Sheshadri (2017) provide an interesting complement, as they study among other things how hierarchies actually help scientists to formulate hypotheses, test them, and assess their robustness.

Another epistemological issue that climate modelers have been working on is the tuning of GCMs. The parameterizations used in GCMs contain many free parameters that are “tuned” in order to ensure that GCM results are sufficiently *realistic*. GCM tuning was sometimes criticized and seen as an attempt to hide certain model deficiencies (e.g. Räisänen, 2007; D. A. Randall & Wielicki, 1997). But in recent years, a rational discourse has emerged from the climate modeling community to justify the tuning and the methods on which it is based.

Particularly noteworthy are the papers of Hourdin et al. (2017a) and G. A. Schmidt et al. (2017) which discuss and justify the uses and limits of climate model tuning.

1.4.2 Convection parameterizations being questioned

Significant improvements have been made since the first parameterizations were developed in the 1970s (see Rio et al., 2019, for an overview of recent progress in parameterization development). Parameterizations have enabled climate models (either atmospheric GCMs, coupled GCMs or ESMs) to produce realistic climate projections, which have been extensively used in IPCC reports to support scientific statements about anthropogenic climate change. However, considering that the pace of progress was too slow, climate modelers who had themselves contributed to the development and improvement of parameterizations have become critical towards the use of parameterizations in climate models (Arakawa, 2004; D. Randall et al., 2003). In response to what these modelers called the “cloud parameterization deadlock”, alternative paths were proposed.

Super-parameterizations A first avenue consisted in the use of CRMs as a “super-parameterization” within each column of a GCM (Grabowski, 2001; Grabowski & Smolarkiewicz, 1999). In this approach, a CRM is used over a fraction (2D or 3D) of a GCM column and gives statistics about subgrid-scale convective processes that are then extrapolated to the whole column (Khairoutdinov, Randall, & DeMott, 2005). Like a conventional parameterization, a CRM receives as input the state variables from the dynamics of the GCM and returns as output the tendencies due to subgrid-scale processes, which are calculated by horizontally averaging the CRM fields over the whole column. Implicitly, the super-parameterization approach makes the hypothesis that there is a clear separation of scales between the small scale and the large scale, so that the small scale is “enslaved” to the large scale. This assumption also plays an important role in traditional convection parameterizations, but has been questioned by several observational studies (e.g. Heggem, Lende, & Løvseth, 1998; Ishida, 1989; Vinnichenko, 1970, see also Chapter 3)

Global cloud-resolving models A second approach is based on CRMs covering the entire atmosphere of the Earth called global cloud-resolving models (GCRMs). A main advantage of GCRMs compared to traditional parameterizations and super-parameterizations is that they better take into account the multi-scale nature of convection. In fact, GCRMs do not rely on a notion of scale separation between the subgrid scale and the large scale and are able to resolve the so-called “mesoscale”, which is neither parameterized nor resolved in traditional GCMs. GCRMs have many applications and can in particular serve as a virtual laboratory to better understand multi-scale interactions in the atmosphere (Stevens, Acquistapace, et al., 2020). However, a major effort is needed to couple GCRMs with ocean models of similar resolution and to run these coupled models over long period: decadal or multi-decadal climate simulations using GCRMs seems therefore out of reach in the near future.

Machine learning Finally, the last approach is based on the recent perspectives offered by the progress in *machine learning* (Balaji, 2021). Most modeling problems can be formulated as an attempt to link certain outputs to certain inputs. In parameterizations, in particular, the inputs are the state variables of the model at each grid point and the outputs are the tendencies given by the parameterizations that define the average effect of subgrid processes on the model grid. Basically, machine learning techniques aim to learn what is the best output

for each input, i.e. to find the best function mapping the set of possible inputs to the set of possible outputs. For that, the machine must first learn to connect inputs to outputs from a training data set, which consists of an ensemble of known outputs given by known inputs. Neural networks, random forests or various other machine learning algorithms can be used in this learning phase. Second, the rules learned by the machine can be applied to other inputs, for which the outputs are not known. For instance, Gentine, Pritchard, Rasp, Reinaudi, and Yacalis (2018) propose to replace convection parameterizations by artificial neural networks that could learn from a GCRMs or a super-parameterized GCM the tendencies associated with various large-scale situations. The main advantage, as they argue, is that the computational cost would be limited essentially to the training phase: once the correct tendencies are learned by the machine learning algorithm, using these tendencies in a GCM would be computationally efficient. It opens the door to decadal or multi-decadal projections at a reasonable computational cost (see also Brenowitz, Beucler, Pritchard, & Bretherton, 2020; O’Gorman & Dwyer, 2018).

Other scientists, however, are reluctant to abandon convection parameterizations and emphasize the significant progress in the development of parameterizations made in recent years (Gettelman et al., 2019; Rio et al., 2019). Recognizing the opportunities offered by machine learning and high-resolution models, some of them propose to take advantage of these opportunities to improve parameterizations and better tune their parameters. For instance, Couvreur et al. (2021) and Hourdin, Williamson, et al. (2020) use techniques from machine learning to explore GCM results for different sets of parameters, and find the parameter values in the whole parameter space that give the best results (see also Schneider, Lan, Stuart, & Teixeira, 2017, for a similar approach).

These different approaches are certainly opposed on a technical basis, but also on an epistemological basis. They reflect different conceptions of the role of models used in climate science. In parameterizations of atmospheric convection, understanding comes before modeling. Scientists working on parameterizations summarize their understanding of convective phenomena in the parameterizations they develop. Parameterizations give them a framework to think about different convective phenomena. Their confidence in GCM results comes as much from the realism of these results as from their understanding of the physical principles used to build the parameterizations. On the contrary, in super-parameterizations and GCRMs, understanding comes after modeling. Unlike parameterizations, CRMs do not contain the modelers’ vision of what nature looks like. The characteristics of simulated phenomena and their interactions appear spontaneously in CRMs, whereas they are prescribed in the parameterizations. This leads modelers to discover the causes and effects of convective phenomena in the outputs of CRM simulations. Some see this as an advantage, as it avoids contaminating CRMs with our potential erroneous conceptions (Stevens, Acquistapace, et al., 2020). However, others consider that rejecting parameterizations deprives scientists of the gradual understanding built up over the last 50 years on convective phenomena, which parameterizations summarize and incorporate into GCMs (Rio et al., 2019).

Similar arguments oppose machine learning to parameterizations. But the use of machine learning methods to replace conventional parameterizations presents also new epistemological challenges compared to super-parameterizations or GCRMs. It is indeed possible to analyze the outputs of a CRM – whether it is used as a super-parameterization or as a GCRM – in order to understand the phenomena that emerge from the simulation. This analysis can help

identify and correct certain biases in the model. On the contrary, it can help understand why certain phenomena seem to be well represented, hence give confidence in the model results. Conversely, machine learning models are difficult to interpret for the user, and sometimes described as “black boxes” (N. Jones, 2017). This criticism is rejected by machine learning advocates, who defend that machine learning is in part interpretable and offers a certain kind of understanding (Dueben & Bauer, 2018; McGovern et al., 2019). The debate then moves to an epistemological ground, with questions such as: what is understanding? What role does it play in climate modeling? Where does our confidence in climate models come from? Which kind of understanding do climate models provide? A critical look at these questions is given in the study of Jebeile, Lam, and R  z (2020), which compares the type of understanding provided by machine learning and the one given by other statistical methods commonly use in climate science such as statistical downscaling. Climate modeling is today at a turning point and we bet that such epistemological reflections are highly relevant to compare the various avenues open to us and draw the future of climate models.

1.5 The perspective of philosophers of science

Over the last 30 years, more and more philosophers have become interested in the specific problems posed by the study of climate and its modeling (Lenhard & Winsberg, 2010; Petersen, 2012; Winsberg, 2018b). It led to the emergence of the philosophy of climate science as a distinct sub-discipline of philosophy of science. This sections aims at introducing the reader to the growing literature in the domain, with a focus on the epistemology of climate modeling, and detailing some of the works that will be relevant in the rest of the manuscript.

1.5.1 Introduction to the philosophy of climate science

So far, the philosophy of climate science has been focused mainly on GCMs. To our knowledge, other numerical models used in climate science (e.g. LES or CRMs) have hardly received any attention from philosophers of science and the theoretical foundations of climate science have been addressed by only a few studies (e.g. Katzav & Parker, 2018). We give here a brief overview of some of the main issues that have been discussed so far by philosophers of science.

Uses of GCMs What is the purpose of computer simulations produced by GCMs? Which uses can we distinguish? Several answers to these questions have been given by philosophers of science. First, GCMs can serve as a **surrogate for experimentation**, as it is not possible to perform experiments on the real climate system (Petersen, 2012). In particular, we are forced to use simulations to study future climate change because we do not have a twin Earth on which we could test the effects of an increase in CO₂ concentration. Second, GCM results can be taken as a **substitute for observational data** (W. S. Parker, 2014). Many phenomena could theoretically be observed in nature, but are difficult to study in practice due to lack of relevant observational data. Computer simulations can help to study phenomena for which observational data are either incomplete or inaccurate. Third, GCMs can be used **heuristically**, that is to formulate and test hypotheses (Heymann & Hundeb  l, 2017), or **serve as educational resources**.

More generally, two main uses of GCM simulations are often distinguished: understanding the climate system and predicting it. While there are bridges between the two – in particular, a good understanding of climate is necessary to trust GCM simulations – they address distinct

research priorities and require different resources (Dalmedico, 2001; Heymann & Hundebøl, 2017). The models used to make climate projections are designed to represent all elements that may be important to the climate system. Over the last decades, they have become increasingly complex and have taken into account more and more phenomena. In particular, ESMs, the most comprehensive models used today, include interactions with the biosphere in addition to interactions between the atmosphere, land, ocean and sea ice. It is easy to get lost in the details of these many interactions and such comprehensive models are therefore difficult to interpret. If one wants to understand one aspect of the climate system, for example cloud feedbacks, comprehensive models are of limited use. On the contrary, simplified GCMs, which assume for example that oceans cover the entire surface of the Earth – so-called aqua-planet GCMs – have limited value for making realistic climate projections. Nevertheless, they provide a valuable tool to understand key aspects of the climate system.

Using climate models to understand or predict climate reflects two different epistemological conceptions of the role of models in climate science. In a world of limited human and computational resources, research strategies need to be defined and these two uses of models are sometimes in competition. This illustrates how epistemological reflections on the kind of science that is permitted, legitimate, or useful with models can influence research priorities in laboratories and how research budgets are allocated.

Validation of GCMs using observations GCMs are routinely confronted to observational data. The ability of GCMs to simulate accurately certain observations – the rate of warming in the twentieth century, for instance – is in fact often used as a criterion to evaluate their ability to predict future climate change (Chen & Frauenfeld, 2014; Koutroulis, Grillakis, Tsanis, & Papadimitriou, 2016). However, the expressions “validation of GCMs” or “verifications of GCMs” by observations have been criticized by philosophers of science due to the methodological difficulties such an exercise represents: to validate (or verify) a GCM, it would be necessary to eliminate all the reasons that could lead to bad results. It is impossible in practice, first because of the limited number of observations available and second, because GCM results are far from matching observations perfectly. Instead of the black and white picture given by the term validation, some philosophers proposed more nuanced terms such as “confirmation of models by observations” (Lloyd, 2009; Oreskes, Shrader-Frechette, & Belitz, 1994), or suggested to use observations only to assess the “adequacy for purpose” of GCMs, i.e. whether they are adequate or not for a certain use (W. S. Parker, 2009, 2020).

These terminological debates might seem anecdotal. Yet, they have consequences on the strategy used to evaluate GCMs: in particular, if the objective is to show the adequacy for purpose of a GCM, scientists should focus on the most relevant variables for the intended use of the GCM. For example, to predict the evolution of Arctic sea ice, they should compare past simulations of Arctic sea ice to observations in order to assess whether the GCM used is adequate for this particular purpose. On the contrary, terms such as validation or verification suggest a deeper examination of various GCM results in order to detect any potential incoherence with observations.

Interpreting multi-model ensembles As part of the Coupled Model Intercomparison Project (CMIP), most of the coupled GCMs used around the world are submitted to the same simulation protocols every 5 or 6 years to compare their results under similar conditions. In particular, CMIP gives a framework to compare future climate projections from different GCMs,

which serve as the basis for the IPCC reports. However, many philosophers have questioned the interpretation one can give to CMIP multi-model ensembles. In particular, the spread of multi-model ensembles has given rise to much debate: should we see it as an estimate of the uncertainty, as suggested by many scientific studies (Pirtle, Meyer, & Hemilton, 2010)? Or only as a lower bound of the uncertainty (W. S. Parker, 2013)? Katzav (2014) is even more cautious and argue that the spread of CMIP multi-model ensembles should only be seen as a range of “real possibilities”.

Philosophers of science also examined the robustness of CMIP results and investigated whether the agreement between different GCMs for a certain result legitimately gives confidence in that result. While Lloyd (2010) argues that the agreement between GCMs is indeed an argument on which we can base our confidence, W. S. Parker (2018) reaches an opposite conclusion and argues that “when today’s climate models agree that an interesting hypothesis about future climate change is true, it cannot be inferred [...] that the hypothesis is likely to be true, nor that confidence in the hypothesis should be significantly increased, nor that a claim to have evidence for the hypothesis is now more secure”. One of the reason for this claim is that, as suggested by Knutti (2018); Pirtle et al. (2010) and others, GCMs are not independent of each other but share ideas and sometimes even pieces of code. Therefore, they are likely to have common biases and the agreement between GCMs may simply reveal these common biases.

1.5.2 The (few) epistemological studies on parameterizations

Parameterizations, and in particular convection parameterizations, are at the heart of GCM uncertainties. Yet, few philosophers have looked specifically at the epistemological issues they raise. Until recently, parameterizations have been mainly a focus for sociological research, through fieldwork and interviews in climate modeling laboratories (e.g. Guillemot, 2017; Sundberg, 2007). This sociological research highlighted different ways of being a climate modeler – what Shackley (2001) called “epistemic lifestyles”. For instance, Shackley (2001) proposed a distinction between *climate seers*, who use GCMs mainly to understand the climate system and explore specific scientific questions, and *climate model constructors*, who attempt to capture the full complexity of the climate system, thus to build a versatile model that can be used for a variety of applications. Climate seers, according to Shackley (2001) do not strive for realism, but prefer idealized parameterizations, whose effects on the model or responses to perturbations are easy to interpret. For model constructors, on the other hand, parameterizations must be as realistic as possible, i.e. they should represent the real phenomena as accurately as possible. By interviewing various researchers from the LMD, Guillemot (2017) offers a glimpse of another epistemic lifestyle, unique to the lab where I did my Ph.D. thesis and characterized, according to the author, by the importance of the “physical understanding” of processes.

These sociological studies also raised epistemological questions about parameterizations. For example, Sundberg (2007) uses sociological fieldwork in a Swedish meteorological laboratory to distinguish different functions of parameterizations, and to study the links between parameterizations, observations and theory. But in such studies, the heart of the matter remains the distinction between different modeling practices, and epistemological reflections are relegated to the sidelines. Consequently, while philosophers have examined various epistemological issues related to GCMs and sociologists have investigated various ways of developing a climate model, the epistemology of parameterizations has so far received little attention. Notable exceptions are the following two issues, intrinsically related to parameterizations, and which have given

rise to (a few) epistemological studies.

Confirmation holism In philosophy of science, confirmation holism is the claim that an individual theory or hypothesis cannot be confirmed or refuted by an empirical test, as such a test would necessarily involve other theories or hypotheses. It is only possible to test an ensemble of theories or hypotheses *as a whole*. Lenhard and Winsberg (2010) defend a strong form of confirmation holism for GCMs: when GCM simulations are compared to observations, the different components of GCMs – namely, the parameterizations – cannot be tested individually. GCMs can only be tested as a whole. According to the authors, it is not possible to isolate the causes of success and failure of GCM simulations, and thus to use this *analytical understanding* to improve them.

The heart of their criticism concerns the parameterizations. Indeed, they explain that parameterizations are too interdependent to be tested in isolation. It is not possible to assess the merits and shortcomings of a given parameterization in a GCM independently of the other parameterizations. This criticism, which has been nuanced by other philosophers (e.g. Frigg, Thompson, & Werndl, 2015), points to the problems of interpretability of parameterizations in complex GCMs. It opens the door to epistemological reflections on the particular kind of understanding given by GCMs. However, as far as parameterizations are concerned, the reflexion proposed by Lenhard and Winsberg (2010) stays at the surface. They do not take into account the fact that a parameterization is neither a theory nor a hypothesis: testing a parameterization has thus a very different meaning than testing an hypothesis or a theory. Before speaking of confirmation holism, it would be necessary to define the particular type of scientific object parameterizations are, a discussion they completely omit.

Tuning and falsifiability Another issue that has been the subject of much epistemological research concerns the tuning of GCMs. Several philosophers reacted to a paper written by two atmospheric physicists, David A. Randall and Bruce A. Wielicki who accused tuning to “artificially prevent a model from producing a bad result” (D. A. Randall & Wielicki, 1997, p. 404). Randall and Wielicki argued that the tuning should be limited to parameters that are both important to the model results but poorly understood, such as parameters related to cloud microphysics. As our understanding progresses, fewer and fewer parameters should therefore be tuned. According to them, understanding the natural phenomena sufficiently to avoid the need for parameter tuning is a primary goal for climate modelers.

The philosopher Petersen compares the position of Randall and Wielicki with Popper’s philosophy, which states that the formulation of a new theory should not be based on ad hoc corrections of an old theory, but only on corrections that are theoretically justified (Petersen, 2012). According to this view, climate modelers should try to justify theoretically the corrections they bring to their climate models; otherwise, the confrontation of a climate model to observations would not be able to test a model adequately. Indeed, the agreement between model results and observational data can come from compensating errors introduced by the tuning process, and it is therefore not possible to use this agreement to validate the model. As Edwards (2001) noted, this is a particularly strong concern when the same observations are used to tune a GCM and to evaluate its performance. On the other hand, if there is a disagreement between the GCM results and observations, the observed disagreement could be due either to errors in the GCM or to incorrect parameter values. The causes of this disagreement would be difficult to diagnose. GCMs are therefore not falsifiable in a Popperian sense (Petersen, 2012).

However, Lloyd (2009) argues that this lack of falsifiability is not an issue as GCMs are not used to test a theory, but to apply theories to give a description (or projections) of the climate system. From this perspective, the use of tuning seems legitimate as it is likely to increase the quality of GCM results.

Again, we believe that without a thorough study of what parametrizations are, these epistemological debates are difficult to settle. If parametrizations were firmly established in theory, there would be no debate and tuning would be akin to calibration: parameters could be set once and for all before the GCMs are run. Conversely, if the parameterizations were entirely ad hoc, they would lack generality and could not be used in a GCM: each situation would require a particular tuning, impossible to generalize. Since parameterizations are partly theory-based and partly ad hoc, they live in a *semi-empirical world*, a term coined by Edwards (2001). In this semi-empirical world, partly determined by theory and partly determined by observations, epistemological debates arise about what is acceptable or not. To shed light on these debates, an examination of the foundations of parameterizations seems necessary.

1.6 Organization of the manuscript

The previous discussions illustrate the interest of both climate scientists and philosophers of science in the epistemological issues raised by GCMs. Parameterizations occupy a central place in GCMs and are at the heart of these debates. Sometimes seen as the “Achilles heel” of GCMs (Stone & Risbey, 1990), they are often described as major sources of uncertainties and opposed to the dynamics of GCMs, solidly anchored in well-established equations. Many epistemological studies consider parameterizations as one of the main causes of difficulties in interpreting and validating GCM results. Yet, in such studies, the description of parameterizations remains superficial: parameterizations are put in the same “physical package”, living in a rather vague “semi-empirical world”. In particular, no distinction is made between the different parameterizations used in GCMs, and their respective justification. However, if we look closely, the different types of parameterizations have little to do with each other: in particular, parameterizations of microphysics, radiation, turbulence and convection are very different in nature and involve different types of justification. Each type of parameterization would thus require a detailed epistemological examination to determine its basis, scope and limitations. Such studies could shed new light on the debates about the interpretation of GCM results.

In my Ph.D. thesis, I choose to limit myself to convection parameterizations. This choice is motivated by the present context, where alternative paradigms to convection parameterizations are being proposed by the climate modeling community – I have cited in particular the emergence of super-parameterizations, GCRMs and machine learning as rivals to convection parameterizations. A discussion on the foundations of parameterizations and the kind of understanding they give us could help assess what would be lost or gained if parameterizations were replaced by other approaches. I will try to answer questions such as: what are we trying to represent when we parameterize atmospheric convection? In what sense are the parameterizations partly ad hoc and partly determined by theory? What interpretations can be given to the parameterization formalism? How do these interpretations help to better understand the role of tuning in GCMs?

To answer these various questions, my approach will be at the interface of climate science and epistemology. I had the chance, during my Ph.D. thesis, to participate in several scientific

works that gave me a close look at various facets of the study of atmospheric convection. In particular, I took part in the EUREC⁴A field campaign and was involved in several studies that resulted from it (Albright, Fildier, & Touzé-Peiffer, 2020; Stevens et al., 2021; Touzé-Peiffer, Vogel, & Rochetin, 2021). This experience allowed me to understand the role and limitations of observations in the study of atmospheric convection. I also developed a parameterization for the atmospheric GCM LMDZ and conducted significant work of model re-tuning following the addition of this new parameterization. Thanks to this work, I acquired a practical knowledge of the epistemological challenges posed by model tuning and the new perspectives offered by automatic tuning methods. Finally, I was confronted with the world of LES and CRM simulations over large domains by contributing to a study defining a method to detect localized areas of cold air near the surface formed by evaporation of precipitation, called “cold pools”, in such simulations (Rochetin, Hohenegger, Touzé-Peiffer, & Villefranque, 2021). These different works are used throughout my thesis as examples, case studies or pretexts for epistemological reflections on parameterizations and their foundations. In the following, I give an overview of the organization of the manuscript.

Behind the complexity of atmospheric flows, we can see some coherent structures, which I call “objects”. The purpose of **Chapter 2** is to study these objects and to specify the role they play in our understanding of atmospheric convection. In particular, I explain that objects represent the conceptual framework on which the convection parameterizations are built. But their role does not stop there: objects also give words to analyze observations and high-resolution simulations. In this, I define them as the prism through which we see and study atmospheric convection. The different facets of the role of objects in our understanding of atmospheric convection are illustrated with the example of cold pools.

In **Chapter 3**, I explore the semi-empirical world of convection parameterizations, between theory and observations. I first study the context in which parameterizations emerged and explain that parameterizations were originally proposed to answer an operational need to “make the models work”, i.e. to ensure that GCMs are sufficiently realistic to be useful. Despite attempts to ground parameterizations solely on observations or to derive them exclusively from first principles, I then explain that most parameterizations used in GCMs offer an object-based representation of convection, relying on both some theoretical principles and more ad hoc assumptions based on observations. The advantages and limits of such framework are discussed.

I then focus on the tuning of GCMs in **Chapter 4**, and study the characteristics of this specific scientific activity. I distinguish on different levels parameters that can be used for tuning, and analyze why tuning is necessary and which tuning strategies can be used. I then explain in which sense climate model tuning is similar and different from calibration procedures used in other scientific fields. I also study the consequences of the non-accuracy of parameterizations on tuning and the risks to introduce “compensating errors” or to “overtune” a model. Finally, I mention recent progress in automatic tuning methods and the paradigm shift they represent for climate model tuning.

I explore how the different issues regarding parameterizations and tuning arise in practice in **Chapter 5**, where I implement a cloud and precipitation overlap parameterization in the atmospheric model LMDZ. Following this addition, I perform a re-tuning of the model using new methods that automate part of the tuning process. Through this case study, I question what it means to improve the parameterizations of a GCM and explain why in practice this

simple question leads to important methodological difficulties. To overcome these difficulties, I give some criteria to evaluate the progress brought by a new parameterization from a conceptual point of view and in terms of model results.

In **Chapter 6**, I extend the discussion to comparisons of different GCMs in CMIP, which serve as basis for the IPCC reports. I show that CMIP was able to highlight the flaws of parameterizations, but did not solve them. One of the historical motivations of CMIP was to provide a framework for improving GCM parameterizations. I explain why this motivation was not fulfilled and detail the various uses of CMIP in climate research today. I show that CMIP has steered climate studies in a certain direction, focused on model applicability rather than model improvement, and I mention the structural effects it has had on climate research.

Eventually, **Chapter 7** summarizes all the work presented in this manuscript and discusses possible avenues for future research.

Chapter 2

Nature and role of objects in the study of atmospheric convection

L'essentiel est invisible pour les yeux

Le petit prince, Antoine de Saint-Exupery

Contents

| | | |
|------------|--|-----------|
| 2.1 | Introduction | 32 |
| 2.2 | Definition and characteristics of objects | 33 |
| 2.2.1 | Splitting the atmosphere into objects | 33 |
| 2.2.2 | On the choice of objects | 34 |
| 2.2.3 | Mechanistic understanding vs functional understanding | 36 |
| 2.2.4 | Sugar, gravel, fish and flowers as examples of objects | 38 |
| 2.3 | Objects in observations | 39 |
| 2.3.1 | Defining and characterizing objects in observations | 40 |
| 2.3.2 | Limitations of observational measurements | 41 |
| 2.4 | Objects in computer simulations | 43 |
| 2.4.1 | Models and computer simulations in climate science | 43 |
| 2.4.2 | Two types of objects | 44 |
| 2.4.3 | Uses and limitations of computer simulations | 48 |
| 2.5 | The example of cold pools | 52 |
| 2.5.1 | Definition of a cold pool | 52 |
| 2.5.2 | The construction of the object “cold pools” | 52 |
| 2.5.3 | The role of cold pools as an object | 55 |
| 2.6 | Conclusion | 63 |

2.1 Introduction

In our daily life, we use words to describe our environment. To describe an entity which has a trunk, branches and leaves, we will use for instance the word “tree”. This single word aggregates a myriad of information coming from our perception: the vision of branches, the sound of wind in the leaves, the texture of a tree trunk, etc. In the word tree, consciously or unconsciously, our brain has attached many characteristics. Spontaneously, if we see something that has a trunk, branches and leaves, we will guess that it should also have roots. Through words, we can not only name the things around us, but also get to know them better.

In this sense, words are the mediator between our understanding and the world. Our language is constituted by the most useful words, concepts or notions which describe our environment. As scientists study phenomena that we do not encounter in our daily lives, they sometimes have to find new words to name them. Indeed, in science as in the real life, words are the lens through which we see the world: they order our environment and make it intelligible. From the particle physicist distinguishing different families of particle, to the biologist finding names for the constituents of a eukaryotic cell, the words created from scratch by scientists are numerous.

Atmospheric physicists are no exception and invented a whole vocabulary to talk about their field of study. One famous example is the *International Cloud Atlas*, first published in 1896 by the International Meteorological Organization (now the World Meteorological Organization). Its objective was to promote a more consistent use of vocabulary describing clouds, in order to facilitate the communication between meteorologists about weather forecasts. As easy as it may sound, we should not overlook the challenge of finding appropriate names to describe what we see in our scientific research. Many scientific debates could have been avoided with a more appropriate choice of vocabulary (see Lévy-Leblond, 2000, for an example in quantum physics). A great challenge for atmospheric physicists – and more generally, for all scientists – is thus to find the sharpest words to describe what they study.

In particular, to separate the problem of atmospheric convection in different sub-problems, atmospheric physicists have divided the atmosphere into different components, called in the following “objects”. The term “object” originally appeared in the field of image analysis (Chien, Ma, & Chen, 2002) to describe the detection of some semantic objects of a certain class (for instance cats or dogs) in digital images and videos. The use of this term was then extended in weather and climate modeling studies to analyse 2 or 3D scalar fields (or sometimes 4D fields when the time dimension is included) and detect some localized and episodic features of interest (Brient, Couvreur, Villefranque, Rio, & Honnert, 2019; Couvreur, Hourdin, & Rio, 2010; Davis, Brown, & Bullock, 2006). Here, we will call object any coherent structure in the atmospheric flow that can be separated from an environment, and that is thought to play a role in atmospheric convection. A typical example is a localized rain shaft: this episodic event can be distinguished from an environment by the presence of a rainfall rate above a certain threshold and obviously plays a role for atmospheric convection. A cloud can also be considered as an object – in this case it is even possible to be more specific and to use more precise objects such as cumulonimbus, stratus, congestus, cumulus, etc. Clouds and rain shafts are visible to the human eye and, at first glance, seem as easy to identify as a tree, a cat or a dog.

Air movements transporting water vapor and heat in the atmosphere play a crucial role in

atmospheric convection. However, unlike clouds, they are invisible to the naked eye. We have to use specialized instruments such as lidars or numerical models to see them. Whether we study nature with the naked eye, with sophisticated instruments, or with numerical models, each scene we observe is unique. Objects help to name and recognize some similarities in this apparent diversity. They allow us to identify some recurrent phenomena behind the disordered and chaotic appearance of atmospheric flows. Finding relevant objects to describe atmospheric convection is key to understand it. In this chapter, we will see in detail that objects focus atmospheric research towards specific research question, help to formulate and test hypotheses, make the link between otherwise isolated studies, and, most importantly, are the lens through which we see atmospheric convection.

2.2 Definition and characteristics of objects

2.2.1 Splitting the atmosphere into objects

At the most fundamental level, the atmosphere is the set of molecules in motion around the Earth. The state of the atmosphere could therefore in theory be described by the position and velocity of each molecule, as well as their interactions with each other. However, this approach is inconvenient in practice due to the amount of information it requires: it is not possible to record all the data needed to describe the behavior of the atmosphere at a molecular level in a human brain or a computer, let alone use it to understand and predict atmospheric or meteorological phenomena that affect us directly, such as air temperature, wind, rain or snow, etc.

In fluid dynamics, it is common to consider instead mesoscale particles, whose volume is small enough to ensure that the characteristics of the fluid it contains (e.g. its velocity, temperature, pressure, etc.) do not vary significantly, and at the same time big enough to contain a relatively large number of molecules in order to average statistical fluctuations. The movement of mesoscale particles can be predicted using mass, momentum and energy budgets of each particle. This is the basis of numerical computation, where the fundamental equations of fluid dynamics are discretized over a given grid, at a given resolution. The smaller the grid, the more accurate the description of the fluid is. Provided a sufficient computer capacity is available, this approach is useful to predict atmospheric phenomena. On the other hand, it is more difficult to use it to understand these phenomena, i.e. to establish a limited, and thus intelligible, chain of causal links between an atmospheric phenomenon A (e.g. there is a cloud) and an atmospheric phenomenon B (e.g. it rains).

The need to work at a relevant temporal and spatial scale to understand atmospheric convection motivates the introduction of objects. The move from mesoscale particles to objects share similarities to the move from molecules to mesoscale particles. An object can indeed be seen as a group of adjacent mesoscale particles having certain characteristics in common. However, contrary to mesoscale particles which group molecules exhaustively, objects generally cover only a small part of the atmosphere. We will call the part of the atmosphere which belongs to no objects “the environment”. Another important difference is that in the move from molecules to mesoscale particles, a single spatial criterion, based on a finite volume, was used to group particles. On the contrary, the spatial and temporal limits of an object are defined by the presence of certain local heterogeneities with respect to the mean field. By local, we mean particles close to each other both in time and space.

By definition, an object has therefore both spatial and temporal coherence. This coherence is not guaranteed by the fluid particles that compose it, because when an object evolves, it can gain or lose some fluid particles. The coherence of an object is instead guaranteed by the presence of some shared characteristics in a certain region of space and during a certain time. Since an object has a spatial coherence, it is possible to define its spatial boundaries and to describe its energy and matter exchanges with its environment and other objects in terms of fluxes at its boundaries as if it was a single mesoscale particle. Since an object has a temporal coherence, it is possible to track it over time and to study its life cycle – its birth, development, and death – and how it interacts with other objects and with its environment at each stage of this life cycle.

To illustrate what it means to define an object, let's consider a very simple object, known long before it was studied scientifically: a cloud. *How do we know that a cloud is a cloud?* We are able to recognize a cloud because our past experience – in particular, the many clouds we have seen in our life – has given us a mental representation of what a cloud is, which matches the information provided by our eyes. Clouds are identified as a particular object in the sky (their environment) thanks to the concordance of what we see and certain criteria in our memory. A cloud can evolve over time: it can be transported by the wind, expand horizontally or vertically, change its shape, etc. However, it has a certain spatial and temporal coherence: it is possible to delimit it in space and follow it in time.

Many other objects used to study atmospheric convection are invisible to our eyes. Figure 2.1 gives examples of some of these objects. It shows, for example, updrafts and downdrafts, which can be defined as a set of adjacent particles with a vertical velocity anomaly with respect to the environment, positive for updrafts and negative for downdrafts. Cold pools, also shown in Fig. 2.1, are another example of objects; they can be defined as a group of particles near the ground with a negative temperature anomaly compared to the environment. More generally, in practice, any object can be defined as a set of adjacent mesoscale particles which present an anomaly with respect to one or several spatial field(s), such as temperature, wind speed, humidity, etc. In observations, the anomaly defining an object has to be detected using data from available instruments, such as thermometers, anemometers, hygrometer, etc. In computer simulations, it has to be diagnosed from model outputs.

2.2.2 On the choice of objects

Objects are in part *natural* and in part *artificial*. They are in part natural because the choice of objects is determined to some extent by the atmospheric phenomena that occur in nature. The objects are indeed based on certain regularities observed within the atmospheric flows. The different particles which constitute an updraft have for instance a common characteristic: a high vertical speed. This common characteristic is not invented, but comes from observational data, which show that certain localized regions of the atmosphere have a higher vertical speed than their surroundings.

One could argue that formally, we could invent any object X as a group of adjacent particles which present a certain property P. But if this object does not exist in nature, it is irrelevant for the study of atmospheric convection. The same holds if an object has no particular role for atmospheric convection. For an object to be useful, it must correspond to real phenomena that play an important role in atmospheric convection.

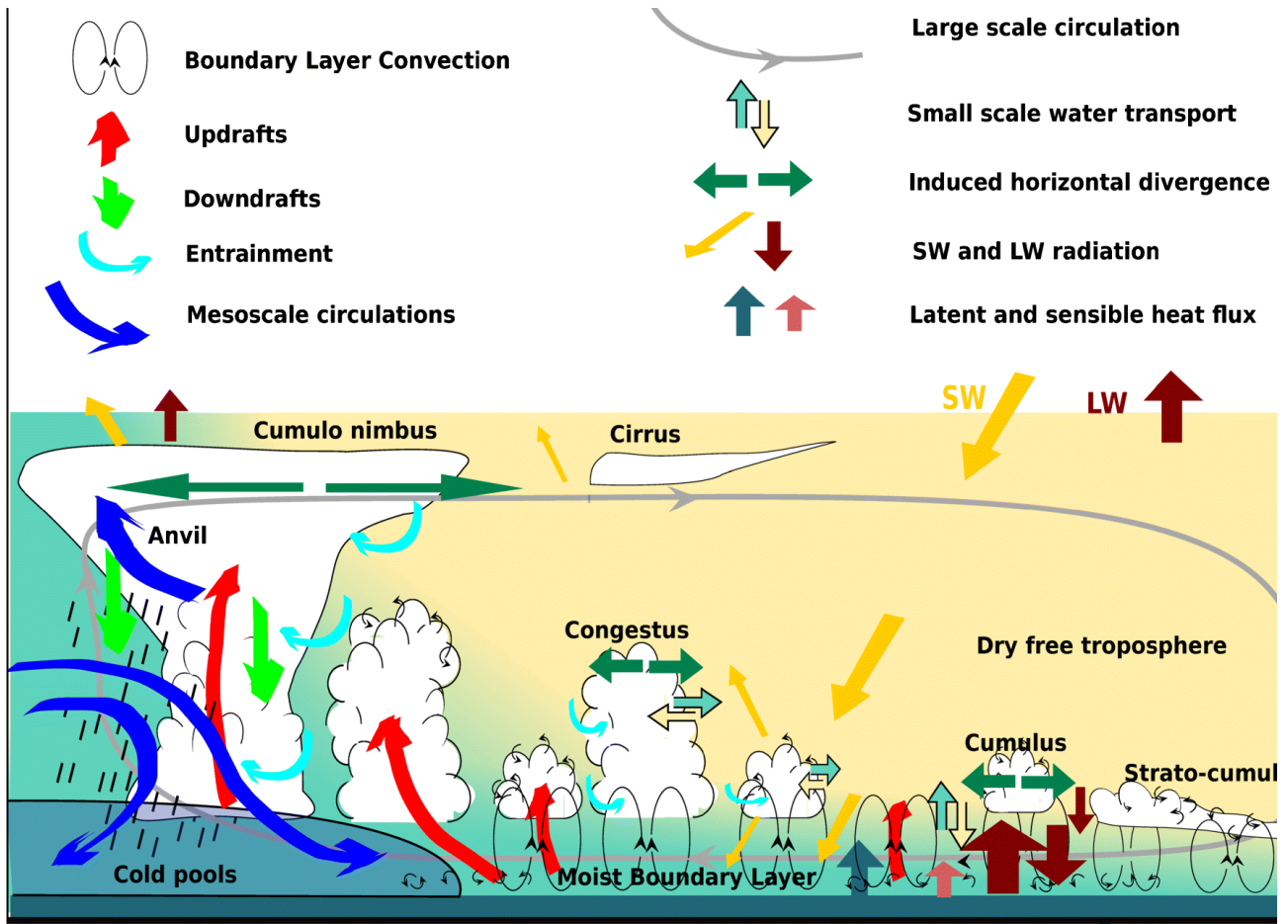


Figure 2.1: Sketch of various objects used to describe atmospheric convection and how they interact with each other and with their environment. From Rio et al. (2019).

Nevertheless, objects are not entirely determined by nature. What nature gives us is only an apparently chaotic and disorganized atmospheric flow where a few patterns emerge. We decided to name the patterns that seemed to be the most important for atmospheric convection and group them into objects. However, there are many ways to divide the atmosphere into different patterns, and the definition we use of a given object is to a large extent artificial – it reflects our subjective understanding of what is important for atmospheric convection. Objects are therefore partly constructed and express our present understanding of atmospheric convection. As we will see in Section 2.5 with the example of cold pools, the objects we use and their definition are likely to evolve as our understanding progresses. Objects are thus both upstream and downstream of our understanding: they are the prism through which we study atmospheric convection and in turn, they are continuously chosen and redefined according to what we know about it.

One could argue that what we say here about objects applies not only to the study of fluids, but to scientific research in general. Indeed, as we explained in the introduction, all scientists introduce words to name the entities they study. In essence, these words are determined in part by nature, but reflect also some kind of understanding of the phenomena at stake. However, contrary to entities used to study other physical systems, the objects used to study fluids do not have well-defined boundaries. A eukaryotic cell, for example, has a membrane which separates

it from the extracellular medium and gives it a real unity. One cannot say the same thing for an updraft. The different particles which constitute an updraft have certain characteristics in common, but are not attached to each other. An updraft has therefore less spatial coherence than a eukaryotic cell. Since it has no explicit boundaries, there are more ways to define it than a eukaryotic cell. In this sense, the objects used to study the atmosphere are more *artificial* than those used in other scientific fields.

Admittedly, the components of the atmosphere that are not in gaseous form – such as cloud droplets, rain drops, ice crystals, or dust particles – do have a well-defined boundary, and even a certain exchange surface with their environment. However, for the understanding of atmospheric convection, these small objects are analogous to air particles: they are too small and too numerous at the scale of atmospheric convection. It is for instance not possible to describe the movement and behavior of each cloud droplet to describe how a cloud interact with its environment. Larger objects have to be used and rain drops are grouped into a rain shaft, cloud droplets, into a cloud, etc. Unlike the many tiny objects that compose them, these larger objects have no clear spatial boundary: for instance, to distinguish a cloud from its environment, an artificial threshold on the concentration of liquid water droplets must be chosen.

2.2.3 Mechanistic understanding vs functional understanding

To a simple question such as “Why do deciduous trees lose their leaves in winter?”, two answers can be given. One answer would be that reduced daylight and temperatures cause some trees to decrease the production of a certain plant hormone, called auxin, which results in a weakening of the bond between the tree’s leaves and the branch. When this bond becomes too weak, the leaves eventually fall off the tree. A second and equally valid answer would be that deciduous trees shed their leaves to conserve resources during the winter and protect themselves from wind during winter storms.

The first answer gives a mechanistic explanation of the phenomenon at stake. There has been many philosophical discussions on the notion of mechanism (e.g. Bechtel & Abrahamsen, 2005; Darden, 2006; Machamer, Darden, & Craver, 2000). We take the definition of Craver (2007, 2013), which call a mechanism a set of entities and activities organized such that they give rise to a certain phenomenon. Mechanistic explanation therefore interpret the behavior of a certain system through the interactions between its components. In Craver’s definition, these components are called entities and are characterized by their structural properties and their relations with other entities.

Though Craver applies his definition mainly to physiological sciences such as neuroscience, these entities correspond to what we called objects in the study of atmospheric convection. When we say for instance that “*updrafts* bring moisture in the atmosphere and thus create *clouds*, which in turn produce *rain* and *downdrafts*, the latter spreading at the surface as *cold pools*”, we typically give a mechanistic explanation of atmospheric convection. We explain convection through the interactions between different objects.

On the contrary, a functional explanation is not so much focused on the structure of the object, rather than on its integration in a larger system. When we say that “deciduous trees shed their leaves to conserve resources during the winter”, we are not looking downward, i.e. at

the underlying mechanisms that cause leaves to eventually fall, but rather upward, i.e. at the role of leaf shedding for the tree in response to changes in its environment. As Craver noted, functional understanding is everywhere in physiological sciences, where different entities generally have complementary functions to maintain certain balances needed for organs to function properly.

This second type of explanation does not seem applicable to the study of atmospheric convection: naively, one could think that fluids are aimless and that it is therefore vain to interpret convective phenomena otherwise than by mechanistic explanations. Convective phenomena do not have their own end or purpose because contrary to biological systems, atmospheric convection has not evolved over time through natural selection to respond to its environment. The theory of evolution explains why certain entities or processes do have specific functions in physiological systems – for instance, why deciduous trees do lose their leaves in winter – but does not apply to non-living phenomena like atmospheric convection. In particular, the Gaia hypothesis, according to which the Earth forms a self-regulating system that helps to maintain and perpetuate the conditions for life on the planet has been criticized by several studies (Kirchner, 2003; Moody, 2012). The study of fluid dynamics might therefore seem fundamentally different from the study of living beings and allow only mechanistic explanation.

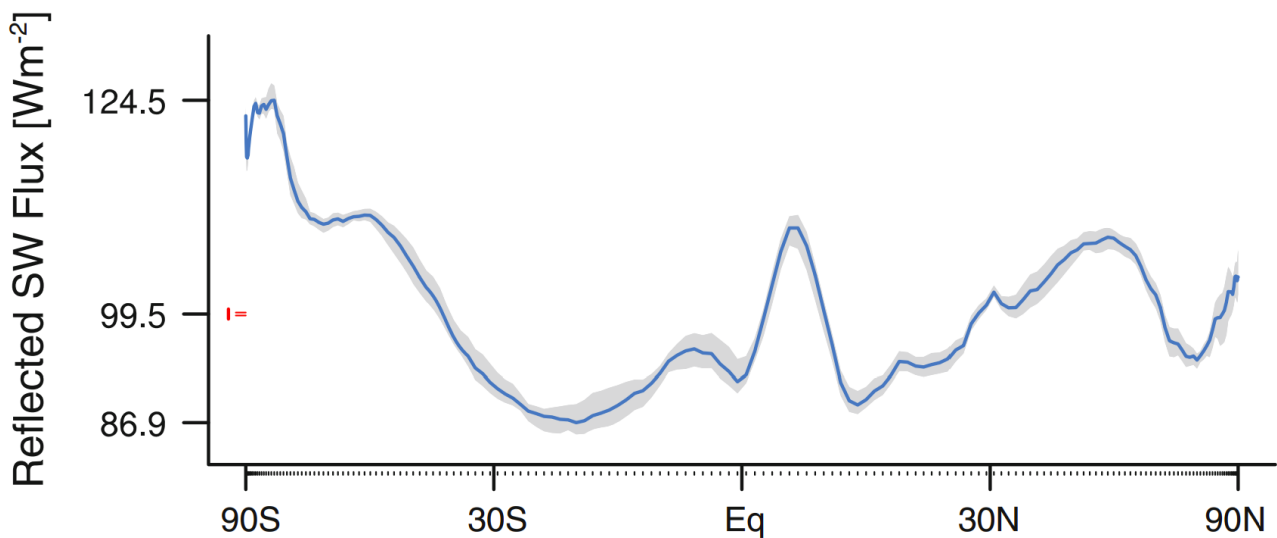


Figure 2.2: Reflected short-wave radiation at top of atmosphere from ten years of CERES EBAF data (Loeb et al., 2009) as a function of latitude. The year-to-year variability is indicated by the gray shading which spans the range of yearly values. The blue line shows the 10-year mean. The small red bar and the two small horizontal bars denote respectively the range in the global and annual means, and the 10-year means averaged over the northern and southern hemisphere separately.

Nevertheless, certain equilibria in Earth’s atmosphere are striking and suggest the presence of certain regulating feedbacks in which atmospheric convection is involved. For instance, Figure 2.2 given in Stevens and Schwartz (2012) shows the constancy of the planetary albedo over annual time scales, not only globally but also within latitude zones. The small vertical and horizontal red bars in the left side are particularly interesting: they show respectively the range of the global inter-annual variability in ten years of measurements and the two hemispherically averaged values over ten years. Despite zonal average differences of more than 40 W m^{-2} ,

the inter-annual range is only 1.16 W m^{-2} and the difference between the two hemispherically averaged values, only 0.35 W m^{-2} . As we have seen in Chapter 1, the amount of shortwave (SW) radiation reflected back to space depends on many factors including the height, depth and microphysical properties of clouds, their location, the color of the Earth's surface, etc. Despite this diversity, the planetary albedo is surprisingly constant over time, with a similar averaged value in the two hemispheres.

Another equilibrium is often invoked to interpret the behavior of atmospheric convection: the convective quasi-equilibrium (CQE). The CQE assumes that moist atmospheric convection reacts rapidly to large-scale instabilities, hence a kind of equilibrium behind convection and the large-scale (Arakawa & Schubert, 1974; A. K. Betts, 1986; K. A. Emanuel et al., 1994). Under the CQE hypothesis, convection drives the atmosphere towards certain typical temperature profiles, as was documented in various observational data sets in the tropics and subtropics (Lord, 1982; Lord & Arakawa, 1980). We discuss the quasi-equilibrium assumption and its limitations in more detail in Chapter 3. For now, we simply note that the CQE assumption typically gives a functional understanding of atmospheric convection, which can be contrasted with the mechanistic understanding given by objects: it expresses the cause of atmospheric convection in terms of its environment rather than in terms of its building blocks.

2.2.4 Sugar, gravel, fish and flowers as examples of objects

Objects determine how we study atmospheric convection. They give us words to name what we see and to give mechanistic explanations of convective phenomena. Therefore, the importance of choosing relevant objects to orient and guide atmospheric research should not be underestimated. This is best exemplified with a recent classification given in Stevens, Bony, et al. (2019), which categorizes cloud scenes from satellite imagery in the western tropical Atlantic into four main cloud types: sugar, gravel, fish and flowers. In this classification, sugar consists of many small clouds scattered across the domain, gravel is recognizable due to the presence of cloud lines or arcs, clouds in a fish have the appearance of a fishbone skeleton, and flowers refers to the presence of larger but more dispersed cloud clusters. As detailed in Stevens, Bony, et al. (2019), these patterns have been defined subjectively from the visual inspection of cloud scenes in satellite imagery. Figure 2.3 gives an example of each of these patterns.

Sugar, gravel, fish and flowers can be seen as four different objects: they are indeed defined as a set of adjacent points with a certain anomaly, here a certain appearance of a cloud field seen from space, with respect to an environment. The environment is in this case the scenes that were identified as containing none of the four patterns. We see with this example that the notion of objects is intrinsically linked to the notion of organization: it is because atmospheric flows organize themselves into characteristic structures that we can distinguish these structures and group them into objects. We would not be able to distinguish objects in a homogeneous flow: objects appear only when there are persistent heterogeneities, i.e. forms of organization.

The classification introduced by Stevens, Bony, et al. (2019) gives four new building blocks to describe and understand mechanistically atmospheric convection. The definition of sugar, gravel, fish and flowers actually expanded a whole field of research. A first step in the characterization of these patterns was for instance proposed by Bony et al. (2020a), which related the presence of each pattern to large-scale environmental conditions. The authors showed that gravel and flowers tended to form predominantly under strong surface winds, and fish and

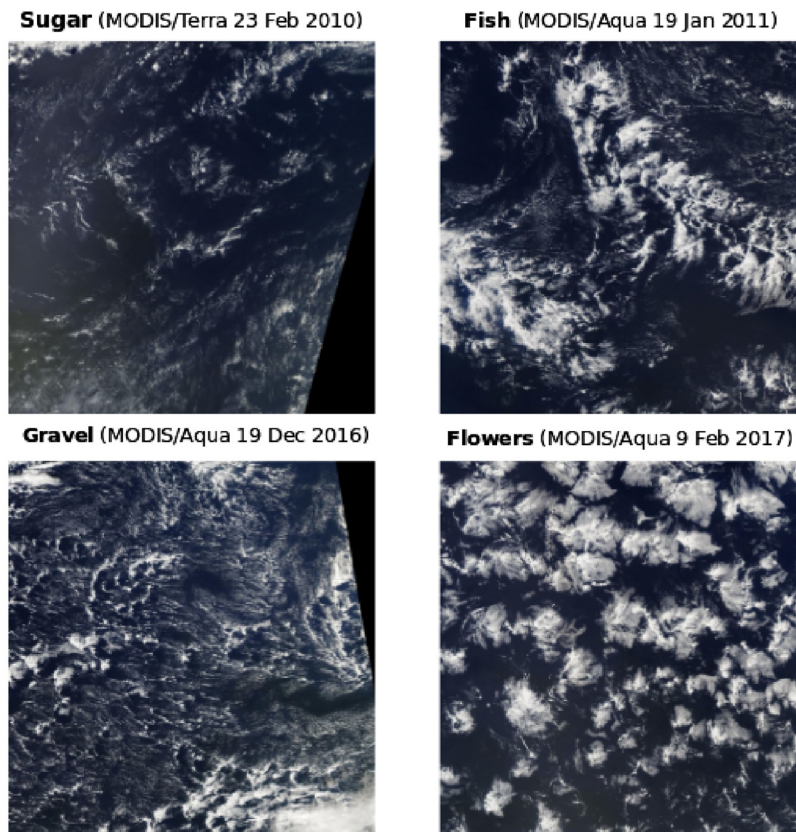


Figure 2.3: Illustration of the four prominent cloud patterns of shallow convective organization pointed out by Stevens, Bony, et al. (2019) over the tropical western Atlantic near Barbados. The four satellite images ($48\text{--}58^\circ\text{ W}$, $10\text{--}20^\circ\text{ N}$) are derived from MODIS imagery. From Bony et al. (2020a).

flowers, in situations with enhanced lower-tropospheric stability. Another recent study of Vial, Vogel, and Schulz (2021) looked at how sugar, gravel, fish and flowers evolve at a daily time scale and showed that fish and sugar appeared mainly during the day, and gravel and flowers during the night. Further studies are expected in the coming years, and the characterization and understanding of these patterns is at the core of the EUREC⁴A field campaign, which took place in January-February 2020 over the western tropical Atlantic, west of Barbados (Stevens et al., 2021, see also appendix B and C).

Sugar, gravel, fish and flowers provide an interesting example of how the definition of certain objects may focus and guide scientific research. Stevens, Bony, et al. (2019) did not discover anything new: some special kinds of shallow convective organization had long been noticed in the tropics (e.g. Warner et al., 1979). However, by naming them, Stevens, Bony, et al. (2019) made a mechanistic understanding of the associated phenomena possible.

2.3 Objects in observations

In the previous section, we defined what objects are and how they can provide a mechanistic understanding of atmospheric convection. In the next two sections, we clarify the role played by observations and computer simulations in this understanding, starting with the role and limitations of observations.

2.3.1 Defining and characterizing objects in observations

Objects are in part natural, in part artificial, that is to say they are not directly given by nature but have to be constructed. In observations, this means that one has to define what an object is, and how it can be detected. Defining an object and detecting it are two sides of the same coin. Any detection method relies directly or indirectly on a certain definition of an object. For instance, if we define a cloud as an ensemble of liquid or solid water particles in suspension in the atmosphere, we can detect the presence of a cloud with an in situ measurement – with sensors on a research aircraft for instance – of the concentration of liquid or solid hydrometeors in the atmosphere. Conversely, a detection method is in itself a particular way of defining an object. When a certain criterion is used to detect an object, it defines at the same time what that object is.

Nevertheless, in practice, the instruments and data available create a strong constraint on the detection method used. There can therefore be a gap between the definition of an object and the method used to detect it in observations. Generally, only a few features of the definition can be used in practice to detect the object. Any detection method can therefore be seen as a projection of the definition of an object on the space of available data.

Once an object is defined, it is possible to study its characteristics. For instance, if we fly with a research plane in a cloud, with appropriate sensors, we can detect the presence of liquid or solid water particles in suspension in the atmosphere. If we perform a sufficient number of research flights in a variety of clouds, we can deduce that this feature is a general characteristic of clouds. On the contrary, when we notice that certain features that are present in some clouds but not in all, we can choose separate the object “clouds” into more specific objects. For instance, differences in vertical velocity inside clouds may lead to a distinction between convective and stratiform clouds. The shape, size and altitude of clouds can create further distinctions between cumulus, strato-cumulus, congestus, and cirrus, as shown in Fig. 2.1. Each time a new object is defined, its characteristics can be studied to determine which properties are generalizable to this object and which are not and may lead to further distinctions.

Objects have played a crucial role in the interpretation of the field campaigns that have been conducted over the past 50 years. For instance, the Barbados Oceanographic and Meteorological Experiment (BOMEX) was one of the pioneering field campaigns in the tropics. It was conducted in May, June and July 1969 in the western Atlantic Ocean north and east of Barbados (Davidson, 1968; Holland, 1970). During BOMEX, clouds were sometimes observed to orient themselves in the wind direction, forming “cloud bands” or “cloud streets”. Kuettner (1971) proposed the mechanism described in Fig. 2.4 to interpret these cloud bands observed during BOMEX. In this mechanism, the observed cloud bands are attributed to helical rolls, which load water vapor in converging areas by sweeping the air laterally across the sea surface. The high humidity in the converging areas provides latent heat to support the rise of the air, which contributes to the helical circulation along the line of convergence. We see here how a certain phenomenon (the presence of cloud bands) is interpreted based on an object (helical rolls) and its interactions with its environment.

BOMEX paved the way to another, much larger field campaign in the tropical Atlantic. The Global Atmospheric Research Program’s (GARP) Atlantic Tropical Experiment (GATE) took place in the summer of 1974 and lasted approximately 100 days. GATE Observations extended from West Africa to the western Atlantic Ocean, between 10 °S and 20 °N. Considerable

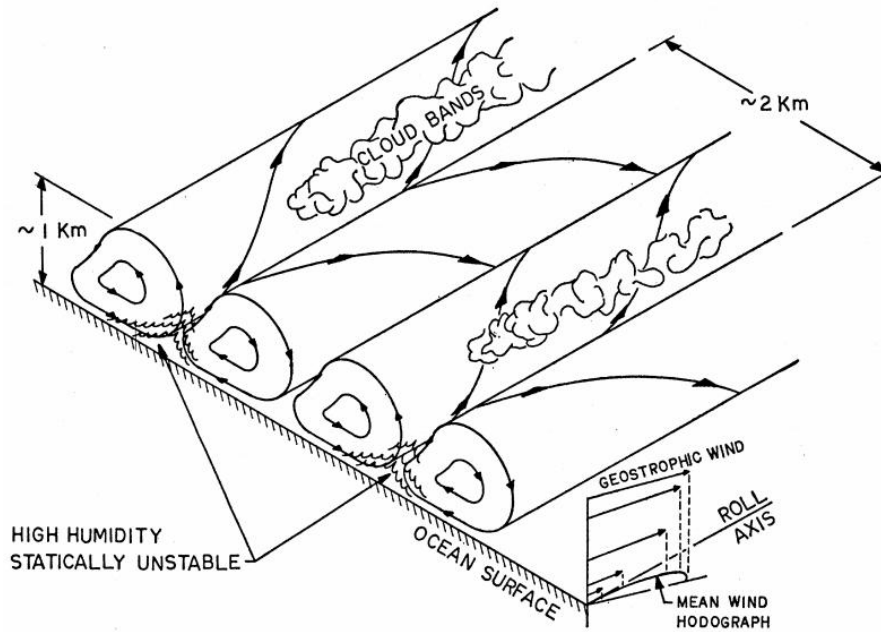


Figure 2.4: Mechanistic interpretation of cloud bands during BOMEX using helical rolls concentrating regions of high humidity along lines parallel to the wind direction (Fleagle, 1972).

resources were deployed, in particular forty meteorological and oceanographic ships, twelve research aircraft and a vast network of buoys from some twenty countries. GATE has led to numerous studies that are still references today (e.g. A. K. Betts, 1978; Houze Jr & Betts, 1981; Warner et al., 1979). In particular, E. Zipser (1977) proposed the conceptual model shown in Fig. 2.5 to describe some of the mesoscale systems observed during GATE, that he names “squall lines”. Using data from GATE and other field campaigns in the tropics, Zipser identified features that are common to different squall lines over the tropical oceans and summarized these features in a conceptual model. As we can see, this conceptual model is based on numerous objects, in particular updrafts, convective-scale downdrafts and mesoscale unsaturated downdrafts. These objects divide a squall line in different interacting blocks, which Zipser uses to give a mechanistic interpretation of the squall line phenomenon.

2.3.2 Limitations of observational measurements

Objects are 4-dimensional: they have a certain spatial extent in 3D and evolve in time. On the contrary, many instruments measure the atmosphere only in 1D. It is the case, by definition, of instruments used for in situ measurements such as radiosondes or dropsondes, weather stations, measurement probes on an aircraft or a ship, etc. Instruments taking remote measurements, whether they are active such as radars or lidars, or passive such as radiometers, spectrometers or interferometers, can measure the atmosphere either in 2D (time + 1 spatial dimension) or, in the best case, in 3D (time + 2 spatial dimensions). Remote measurements can be taken either from the ground, from an aircraft or from a satellite. These different types of instruments offer complementary visions of objects. Nevertheless, since objects are 4-dimensional and instruments are at best 3D, at least one of the dimension of the detected objects is missing. An entire object cannot be captured in observations. For instance, the sugar, gravel, fish and flowers patterns mentioned precendently were detected in satellite images in 2D only (two spatial dimensions, no time). Satellite images gave access neither to the vertical extent of these

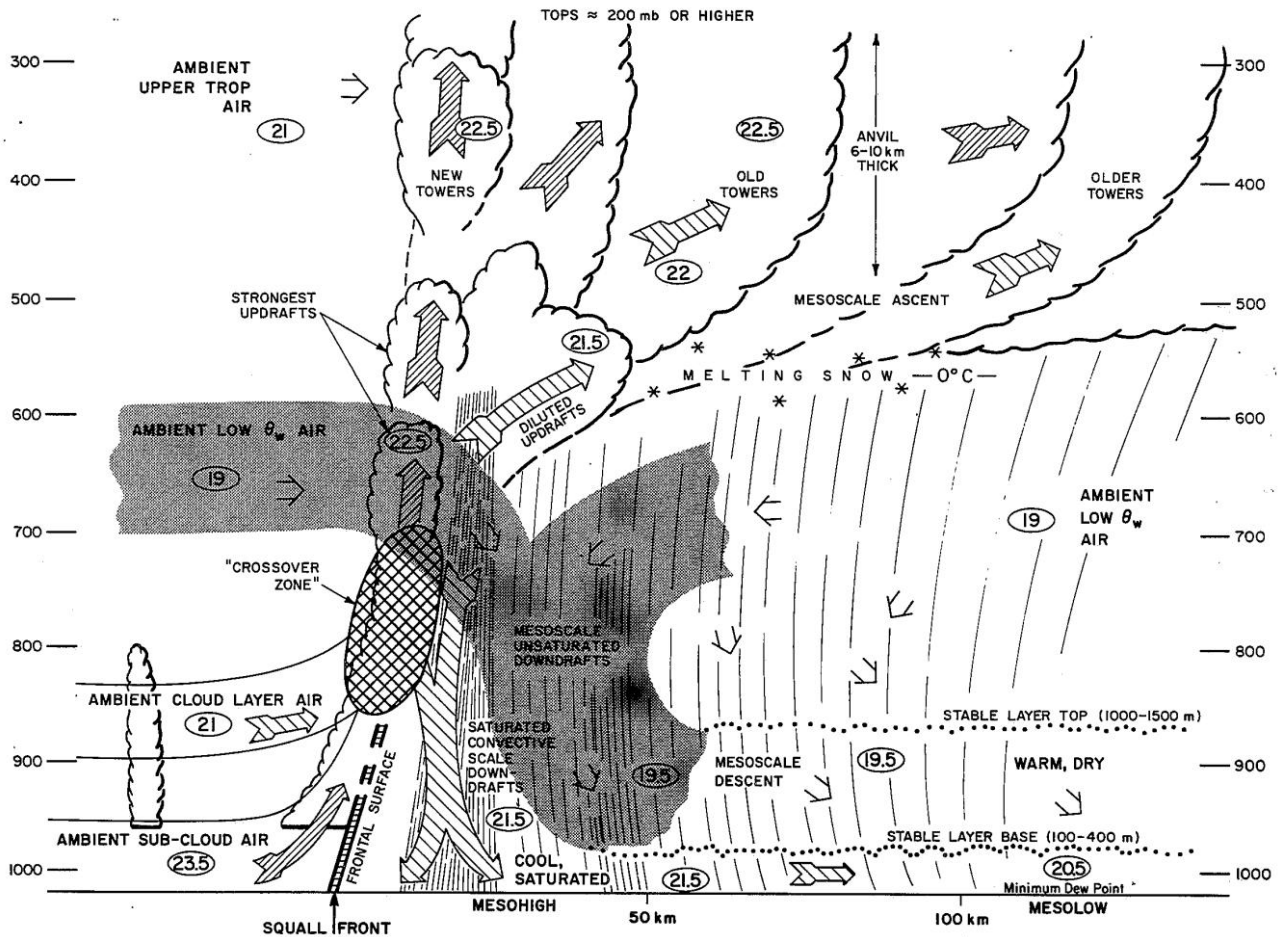


Figure 2.5: Schematic of a certain class of mesoscale convective system called squall lines, featuring different objects. The arrows indicate the flow relative to the squall line, which is moving from right to left. From E. Zipser (1977).

objects, nor to their evolution in time.

To describe the objects, it is thus necessary 1) to combine measurements from different instruments and 2) to perform statistics over several objects, as the different instruments will likely not measure the same object simultaneously. An object studied in observations is thus generally a composite of various measurements. Connecting isolated measurements from different instruments taken at different times and places is in fact an important role of objects in observations. For instance, during the field campaign EUREC⁴A, in situ measurements of sugar, gravel, fish and flowers were taken from dropsondes and radiosondes, and gave access to the vertical extent of these patterns (see appendix C, Fig. C.5). Without a definition of these four patterns, it would have been impossible to connect the properties deduced from these measurements to similar cloud scenes visible in satellite images.

When different instruments are used to study objects, each instrument will likely have its own detection method. Nevertheless, the various detection methods used have to be compatible with each other. To verify this, one can for example cross-check the measurements made by different instruments on a few well-chosen examples of objects, in order to ensure that the different detection methods used give the same results. The characteristics of a particular type of

object can then be studied statistically, by combining the measurements collected by different instruments, on different objects. For instance, in appendix B, we combine satellite, ship and sounding measures to study the properties of trade-wind cold pools.

However, as an entire object cannot be captured from instrumental measurements, observations offer only a partial view of an object: in particular, it is not possible to know accurately fluxes of energy and matter at the boundaries of an object from observations. As a consequence, causality is also difficult to establish. To analyze the causes and effects of an object – which we have defined as a group of adjacent air particles – we would like to know what goes in and out. Such budgets at the boundary of an object enable to establish formal relationships between this object and its environment, or between two objects, and thus to test mechanisms. For instance, to prove that an updraft creates a cloud, we need to know the mass flux and quantity of water vapor brought by the updraft to the cloud layer. Various approximations exist to estimate the mass flux (see for instance Vogel, Bony, & Stevens, 2020), but one should keep in mind that they are only approximations and not exact calculations. More generally, observations may be consistent with certain mechanisms linking objects, but do not prove them. This is why computer simulations – where, as we will see, formal causal relationships can be established – have to be used to complement observations.

2.4 Objects in computer simulations

We will now look at objects in computer simulations. In particular, we will make a distinction in the following between resolved objects and parameterized object and specify different uses of computer simulations to study objects.

2.4.1 Models and computer simulations in climate science

Let's first precise what we mean by a computer simulation. In a broad sense, a simulation is a method to study the behavior of a system thanks to another system. We talk about a computer simulation when this study is performed using a computer. A computer simulation involves different steps: formulating a model to represent the process, finding a way of implementing this model in a form that can be run on a computer, calculating the output of the algorithm, and visualizing and studying the resultant data. The model itself can be expressed either as a set of equations – for instance, the equations of fluid dynamics to study the behavior of a fluid – a set of rules, or both. The advent of computer simulation is relatively recent – the first examples of use of computer simulations in science comes from meteorology and nuclear physics after World War II (Edwards, 2010). Today, the use of computer simulations is nevertheless pervasive in many scientific fields: among others, particle physics, astronomy, meteorology, climate science, or even social sciences such as economics or sociology rely heavily on this tool. Some entire scientific fields, such as the study of complex systems or chaos theory, are based almost exclusively on computer simulations.

In climate science, the object of study is the behavior of the Earth system – and in particular its two main components, the atmosphere and the ocean – over long time scales, from decades to centuries and under different conditions. As both the atmosphere and the ocean are fluids, most of the models used in climate science are based in one way or another on Navier-Stokes equations, discretized over a given grid, with a given resolution. By definition, processes occurring at scales smaller than the grid used – what we will call subgrid-scale processes – will not

be resolved, as well as processes that affect the behavior of the atmosphere and the ocean, but cannot be described by fluid dynamics equations (microphysical processes associated with the formation of cloud droplets and precipitation for instance). To take their statistical effect at the scale of the grid into account, climate scientists introduce so-called parameterizations. In large-eddy simulations (LES) for instance, the resolution is typically between tens and hundreds of meters and the small-scale turbulence is not adequately resolved. The Smagorinsky-Lilly model (Lilly, 1967) is a parameterization commonly used to describe the effect of this small-scale turbulence at the scale of the LES grid. Global Climate Models (or general circulation models, GCMs) used to make projections about future climate change are much coarser than LES and typically have a resolution of tens to hundreds of kilometers. At such scales, many processes are not taken into account and several parameterizations have to be introduced. A particular difficulty comes from the representation of atmospheric convection for which many different parameterizations have been proposed. When comparing projections from different GCMs (as done in IPCC reports; see Flato & Marotzke, 2014), most of the differences between models are due to differences in these parameterizations. The importance of convection for climate is the reason why we focus on objects that are used to study atmospheric convection in this chapter – and on convection parameterizations in the rest of this thesis – although some of the points we make here could be applied to other objects used in climate science.

2.4.2 Two types of objects

Objects can be either an emergent property of a computer simulation or one of its building block. For instance, in LES, convective updrafts are explicitly simulated, whereas in GCMs, they are included in the parameterizations. In the first case, when objects are detected in the model outputs, we speak of “resolved objects”. In the second, when objects are contained in the model formulation, we refer to them as “parameterized objects”.

Let’s note that in a single computer simulation, some objects may be an emergent property of the model, and others included in the model formulation. In particular, computer simulations with a kilometric resolution typically resolve deep convection, but are too coarse to adequately resolve dry and cloudy shallow convection. The associated models sometimes include a parameterization of shallow convective updrafts (e.g. Pergaud, Masson, Malardel, & Couvreur, 2009), leading to a coexistence of resolved deep convective updrafts and parameterized shallow convective updrafts in these simulations.

In what follows, we will see that the status and role of objects are quite different depending on whether they are an emergent property of the computer simulation, or a component of the associated model.

Resolved objects

Let’s first study objects that are an emergent property of a computer simulation. It concerns mainly LES, i.e. simulations with a horizontal resolution between tens and hundreds of meters. Certain relatively large objects may also emerge in the outputs of simulations with a kilometric resolution, which we called in the introduction cloud-resolving models (CRMs). In the following, for simplicity, we use the generic term “high-resolution simulations” to refer to simulations with a grid fine enough to resolve objects, and distinguish them from GCMs with a coarser grid.

Many scientific papers aim at defining and characterizing coherent structures in the output of high-resolution simulations, even if the word “object” is not always used to designate these coherent structures (e.g. Brient et al., 2019; Dupont, Brunet, & Finnigan, 2008; Tompkins, 2001; Torri, Kuang, & Tian, 2015). Two main types of methods to detect an object in the output of a high-resolution simulation can be distinguished:

- methods that rely on an anomaly of a certain model output variable, or a combination of model output variables (e.g. Hirt, Craig, Schäfer, Savre, & Heinze, 2020; Tompkins, 2001; Torri et al., 2015). Such methods are applicable to any simulation, provided that the relevant output data are available. They only require some post-processing of the simulation data.
- methods using anomalies of the concentrations of tracers emitted at strategic locations in the simulation. In this case, the tracer that makes the objects appear is integrated in the simulation itself: it is transported with the fluid at each time step and each grid point. Then, a post-processing script detects objects based on anomalous tracer concentrations. In a simulation of marine stratocumulus, Brient et al. (2019) detected for instance updrafts and downdrafts from anomalous concentrations of two distinct tracers, emitted respectively at the surface and at cloud top.

High-resolution simulations, contrary to observations, provide a 4D view of the objects that are explicitly resolved. Regardless of the method used, it is thus possible to detect objects in their entirety in simulations. Once an object is defined, not only its characteristics can be inferred, but also the exchanges at its boundaries, thus its interactions with other objects and with its environment. By quantifying such interactions, Brient et al. (2019) for instance proved that updrafts and downdrafts contributed to most of the total transport of heat and moisture in the boundary layer, even if they covered only a small part of the domain. Another example is provided by Hirt et al. (2020), who showed in LES of different resolutions that convection is triggered much more efficiently near the edges of cold pools than in other regions, and quantified the role of cold pools for convective initiation. These two examples illustrate how objects can be used to establish causal relationships in the outputs of a computer simulation. We discuss in 2.4.3 whether such relationships can be extended to observations.

Parameterized objects

In the following, we distinguish two kinds of parameterized objects: those that are used in conceptual models and those that are used in GCM parameterizations. These two types of objects have a number of points in common, but GCMs create specific constraints that should be highlighted.

Objects in conceptual models Historically, many conceptual models have been proposed to explain phenomena in the atmospheric boundary layer, convection in particular, using a set of simple thermodynamic equations. Stevens (2006) provides a review of simple conceptual models used to study tropical and subtropical maritime atmospheric boundary layers and introduce a formalism to build such models. Using this formalism, Naumann, Stevens, Hohenegger, and Mellado (2017) interpret the behavior of a dry convective boundary layer with the conceptual model illustrated in Fig. 2.6. In this simple model, we can distinguish two main objects: the boundary layer (BL) and the free troposphere (FT). These objects are described using certain variables, in particular the potential temperature θ_{BL} , the height h of the BL and the vertical velocity in the FT w_{FT} , as well as other prescribed parameters such as radiative cooling rates

in the BL and in the FT, respectively Q_{BL} and Q_{FT} .

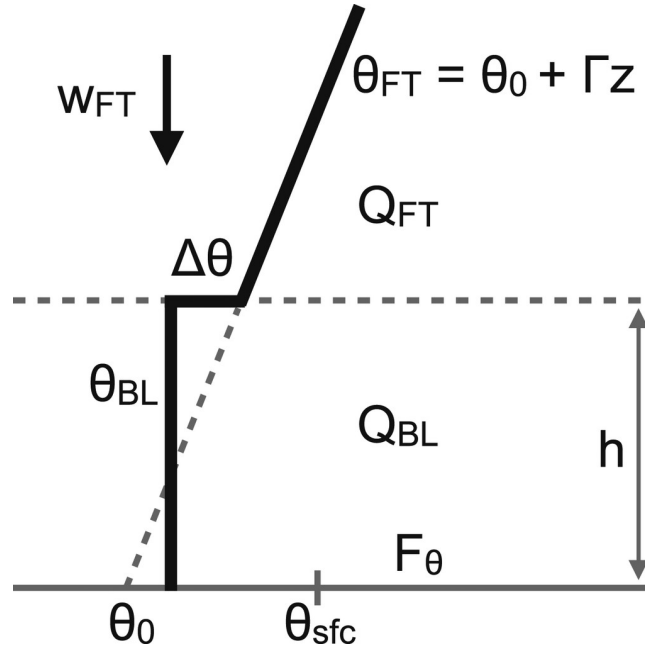


Figure 2.6: A conceptual bulk model for a dry, convective boundary layer with prescribed horizontally homogeneous cooling rates. From Naumann et al. (2017).

When we write such a conceptual model, we have to choose first some objects, characterized with certain variables, and then how we relate them. Defining objects and their interactions is not, by itself an explanation, though it may reflect our understanding or intuition of the phenomena of interest. The explanation comes later. Once the objects are defined, they have a certain autonomy. Whether the results of a model are analyzed using a pen and paper method or simulated using a computer, new causal relationships, that were not explicitly written or thought in the formulation of the model, may appear between the objects. For instance, Naumann et al. (2017) find that the height of the BL increases with the radiative cooling in the BL and prove that this behavior can be established analytically. Conceptual models can thus be used to establish causal relationships between objects, even if the latter are components rather than emergent properties of this kind of models. Once again, we leave to Section 2.4.3 the discussion on whether or not these causal relationships can be extended to the real world.

Objects in GCM parameterizations Many GCM parameterizations are based on conceptual models similar to the ones we just mentioned (see for instance Arakawa & Schubert, 1974; Deardorff, 1972). They thus share many characteristics with these conceptual models. In particular, they also rely on the definition of some objects – such as updrafts, downdrafts, clouds and cold pools – and relations between them. Nevertheless, the fact that parameterizations are implemented in operational GCMs creates some constraint on the form they can take. As illustrated in Fig. 2.7, GCMs have a dual structure. On the one hand, they are based on a particular form of fluid dynamics equations, so-called primitive equations, which are discretized over a grid spanning the whole atmosphere. In current GCMs, the typical horizontal resolution is between 50 and 300 km, with a vertical resolution varying from a few tens of meters to a few dozens of kilometers depending on the altitude – the finest resolution being near the surface. The time step of the model is typically around 10 min, which means that every ten minutes,

the state variables of the model such as the temperature, the specific humidity, the pressure, or the three components of the wind vector are calculated at each grid point.

This discretization is used to build the “dynamics” of GCMs, which computes the large-scale transport of air masses in the atmosphere. However, many phenomena occur at a smaller scale than the model grid and are ignored by the dynamics. Parameterizations are introduced to take these subgrid-scale processes into account and constitute what is called the “physics” of GCMs. Some parameterizations, for instance turbulence, occur at a subgrid scale both horizontally and vertically, but many parameterizations span multiple vertical levels and are thus discretized over the grid of the GCM in the vertical direction (see Fig. 2.7). It means that the characteristic variables of each object used in a parameterization, for instance its potential temperature or specific humidity, are calculated within each horizontal cell at each vertical level in the model. It is an important difference with conceptual models, which can freely choose the vertical grid used to represent the objects and their interactions. The conceptual model of Naumann et al. (2017), which is built on a two-level grid, cannot be directly turned into a parameterization for this reason.

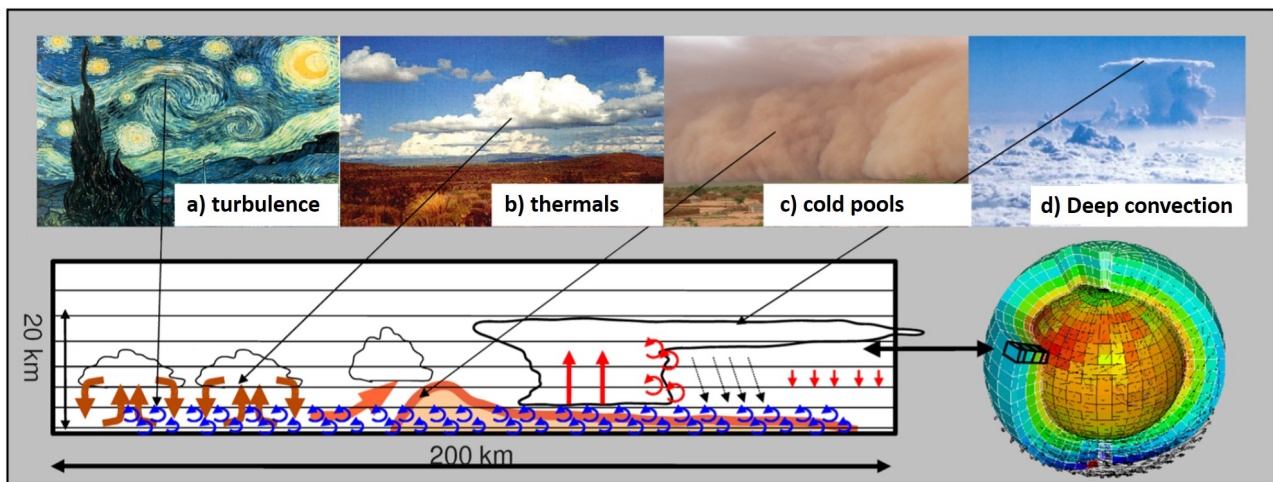


Figure 2.7: Sketch of an atmospheric column in the atmospheric GCM LMDZ. Different parameterizations are illustrated: diffusive turbulence (circular blue arrows), thermals that are used to represent shallow convection (brown arrows), deep convection, including deep convective updrafts (red upward arrows), compensating subsidence (small red arrows) and downdrafts (black downward arrows), and cold pools created below deep convective clouds (in orange). The interactions between cold pools and deep convection are also represented and materialized in particular by the orange arrow, which represents the triggering of deep convection by cold pools. Adapted from Hourdin et al. (2010).

As for conceptual models, once a parameterization is implemented in a GCM, new causal relationships between objects, that were not explicitly written in the parameterization, may appear. However, due to the complexity of GCMs, it is more difficult to diagnose such causal relationships in the outputs of a GCM simulation than in a conceptual model. Lenhard and Winsberg (2010) even accused GCMs of facing a strong form of confirmation holism, preventing to understand analytically the results of GCM simulations. In chapter 6, we nuance this claim by referring to the many tools that can be used to play with model results and interpret them. For now, we simply assume that analytical understanding of GCM results is possible, but much

more difficult to establish than for simple conceptual models. Despite that, GCMs can be used to investigate the effect of objects in a much more realistic framework than for conceptual models. In particular, the interactions between convection and the large-scale circulation, which are thought to play an important role for climate, can be studied (D. A. Randall, Dazlich, & Corsetti, 1989; S. Sherwood, Roca, Weckwerth, & Andronova, 2010).

2.4.3 Uses and limitations of computer simulations

Whether we find some causal relationships thanks to resolved or parameterized objects, the big question is of course: how well do the results found in the simulations apply to the real world? At least, two main uses of computer simulations can be distinguished. First, computer simulations may help scientists to formulate new hypotheses, which could then be tested using observations, for instance. This is what W. S. Parker (2008) called the *heuristic function* of computer simulations. Then, computer simulations might be used as *evidential resources*, i.e. to test hypotheses. In the study of atmospheric convection, both uses present epistemological challenges.

Heuristic and evidential uses of computer simulations

When computer simulations are used *heuristically*, that is to formulate hypotheses, we would like to use observations to test these hypotheses. However, as we explained in Section 2.3, observations provide only a limited view on objects. There may be some hints of certain causal relationships in observations, but it is not possible to prove them rigorously. Consequently, for the study of atmospheric convection, we cannot just ask computer simulations to serve as mere heuristic resources: they must also act as evidential resources and help test and validate certain hypotheses.

Nevertheless, to serve as an *evidential resource*, a computer simulation must itself be validated, which raises another bunch of epistemological questions, such as: how do we know that a computer simulation is good enough to serve as an evidential resource? What criteria should be used? These questions are not specific to the study of atmospheric convection and many philosophers have studied the epistemological issues associated with the validation of computer simulations (e.g. Frigg & Reiss, 2009; Klein & Herskovitz, 2005; Winsberg, 2009). The principal strategy used to validate a computer simulation consists in comparing its outputs with data from observations. In practice, observations are however often scarce and it is furthermore not clear how much agreement with observational data is necessary to have confidence in simulation results. Due to these difficulties, the expression “validation of computer simulations” by observations itself has been criticized by Oreskes et al. (1994) as it leads to overconfidence. Taking into account this criticism, the last IPCC reports changed their terminology and used the term confirmation of climate models by observations instead (Giorgi & et al., 2001; IPCC, 2014).

Trust and distrust in simulations of atmospheric convection

In computer simulations used to study atmospheric convection as in other simulations, it is difficult to guarantee the reliability of the outputs of a simulation. Nevertheless, the fact that the physical laws that describe fluid motions are well understood gives a fundamental reason to have a relative confidence in high-resolution simulations and in GCMs. It has to be contrasted

with empirical laws used to study human behaviors in economy or sociology, or the dynamics of ecosystem – a famous example being the Lotka-Volterra equations, also known as the predator-prey equations, which are often used to describe the dynamics of biological systems where two species interact, one as a predator and the other as prey. Contrary to the equations of fluid dynamics, which can be derived from Newton’s laws, the Lotka-Volterra equations are not related to more fundamental principles. Models describing fluid motions in the atmosphere are therefore based on a more solid theoretical foundation than models using empirical laws such as the Lotka-Volterra equations.

This argument concerns only the *dynamics* of high-resolution simulations and of GCMs. The parameterizations used in these models do not have such a solid theoretical basis (see Chapter 3) and are, in practice, a main source of uncertainties for GCMs (McFarlane, 2011; T. J. Phillips et al., 2004). Due to these uncertainties, if we establish a formal relation in the outputs of a GCM, this relation may not hold in the real world because of certain important phenomena absent from the simulation or inadequately represented in the parameterizations. This is also true for high-resolution simulations, which use parameterizations for microphysics, turbulence and radiation (Jansson, Edeling, Attema, & Crommelin, 2021). High-resolution simulations, however, are more accurate than GCMs because they resolve many phenomena that are parameterized in GCMs. As a consequence, despite the superiority of high-resolution simulations and to a lesser extent, GCMs, compared to simulations solely based on empirical laws, there are many reasons why certain causal relationships established in the outputs of an atmospheric simulation would not hold in the real world.

The validation of conceptual models is even more challenging. Conceptual models represent a further degree of simplification compared to parametrizations: they also contain objects, but these objects are generally more simplified than those used in parameterizations. Moreover, the domain of application of conceptual models is extremely limited: conceptual models usually target one or a few phenomena and it is not possible to test them in a variety of situations as can be done for GCMs and high-resolution models. Despite that, the main advantage of conceptual models is that they can be understood analytically: in these models, the link between different objects can be easily diagnosed.

Table 2.1 summarizes the characteristics, uses and limitations of the various tools we have mentioned to study atmospheric convection. This list of tools is not exhaustive, and we could mention in addition:

- **Direct Numerical Simulations** (DNS), in which the whole range of spatial and temporal scales of turbulence is resolved, but which requires a resolution of approximately 10^{-3} m for the atmosphere. DNS are thus very expensive and can only be run over very small domains.
- **cloud-resolving models** (CRMs), which have a coarser horizontal resolution than LES, typically between 1 and 10 km, but can encompass larger domains. Recently, global simulations using CRMs and SRMs have even been performed (e.g. Satoh et al., 2019; Stevens, Acquistapace, et al., 2020; Stevens, Satoh, et al., 2019).
- **Earth System Models of Intermediate Complexity** (EMICs), which are simplified versions of GCMs – typically with less details in the representation of processes in parameterizations and/or a coarser resolution. EMICs are sometimes seen as a bridge between conceptual models and GCMs (Claussen et al., 2002; Weber, 2010).

| Tools | Characteristics | Uses | Limitations |
|-----------------------------------|--|--|---|
| Observations | Either <i>local</i> and 1D with dropsondes, radiosondes, measurement probes on a weather station, etc. or <i>remote</i> and potentially 2 or 3D (time+2D) using radar, lidar, radiometers, etc. from the ground, an aircraft or a satellite. | Detect and characterize objects from statistics over several objects. | 1) Do not give a 4D view of the objects, thus cannot prove causal relationships between objects. 2) The characterization of the object is limited by the accuracy of the instrument used. |
| Large-Eddy Simulations (LES) | Horizontal resolution: around 100 m. Domain size: 10-100 km. Parameterizations: turbulence, radiation and microphysics mainly. | Resolve explicitly some objects that can then be detected and studied in 4D. | 1) Do not span the whole atmosphere, and are therefore inappropriate for studying interactions between convection and the large-scale circulation or climate. 2) (Small) biases associated with the parameterizations. |
| General Circulation Models (GCMs) | Horizontal resolution: between 50 and 300 km. Domain size: the whole atmosphere (+ oceans in coupled GCMs). Parameterizations: shallow and deep convection, turbulence, radiation, microphysics. | Simulate the whole atmosphere and its interactions with other components of the climate system (ocean, sea ice, vegetation). | 1) The understanding and the interpretation of GCM outputs raise difficulties. 2) (Large) biases associated with the parameterizations. |
| Conceptual models | Conceptual models consist of a set of closed equations linking different variables and free parameters, usually targeting a given phenomenon. | Understand analytically the behavior of the objects defined by the equations. | 1) Conceptual models have a limited domain of application. 2) The validation of conceptual models is even more difficult than for LES and GCMs. |

Table 2.1: Different tools to study atmospheric convection, with their characteristics, uses and limitations.

Need for a model hierarchy

These models all have some usefulness, but also some limitations. Due to these limitations, when a certain causal relationship is found in a particular computer simulation, there is some doubt whether this relationship is valid as well in the real world. Then, how can we prove that a given relationship is robust? The mathematical ecologist Richard Levins was confronted to a similar issue when using various kinds of models in population biology. In Levins (1966), he explains that “all models leave out a lot and are in that sense false, incomplete, inadequate.” The strategy he proposes to reach robust conclusions thanks to these imperfect models is to use several models, based on different assumptions, but which attempt to represent the same phenomenon. If these various models lead to similar results, he explains that these results might be considered as independent from the details of the different models, thus robust. This leads him to claim that “our truth is the intersection of independent lies.” (Levins, 1966, p. 423). Levins’ account of robustness is a strong argument for the use of various models to study atmospheric convection, or more generally any important phenomenon in the climate system. Robustness can for instance be achieved when the results diagnosed in the outputs of LES, CRMs and GCMs are consistent with what we see in observations, and can furthermore be interpreted using simple conceptual models.

In addition, the use of models of varying complexity can facilitate our understanding of the results of comprehensive GCMs, the most realistic models to make projections of future climate change. In an influential paper entitled “The Gap between Simulation and Understanding in Climate Modeling” (I. Held, 2005), the climate physicist Issac M. Held compares the progress in climate modeling and in biology, where a hierarchy of biological systems of increasing complexity have been used to understand the human genome, from bacteria, to insects, mice and finally, humans. Rapid progress in the understanding of the human genome was possible because biologists first put their focus on simple organisms, such as the bacterium *E. coli* or the fruit fly *Drosophila melanogaster*. Similarly, to understand the most complex GCM simulations, Held argues that simplified models capturing the essence of the phenomena at stake should be used. Reflecting on the progress made in climate research since the Charney Report on CO₂-induced climate change published in 1979, Bony et al. (2013) reach similar conclusions: progress in climate change projections depends critically on our understanding of the phenomena at stake, and this understanding can be achieved through the use of models of increasing complexity.

However, having a hierarchy of model of various complexity is not enough: we also need to make these different models talk to each other. Contrary to the genome, which is written in nature in a single language, the various models used in the study of atmospheric convection sometimes look like a tower of Babel: the view they give of atmospheric convection is so different that it is not easy to transfer the knowledge acquired on a given model to another (Jeevanjee et al., 2017). For instance, when we compare the conceptual model of a dry, convective boundary layer given in Naumann et al. (2017), with the parameterizations of the atmospheric GCM LMDZ shown in Fig. 2.7 or the outputs of a LES, we have three quite different pictures of atmospheric convection.

To climb the model hierarchy, objects that are consistent across models have to be defined. For instance, to compare the results of their conceptual model with LES, Naumann et al. (2017) had to define the object “boundary layer” in LES. For this, they used the height of maximum temperature variance as a proxy for the height of the boundary layer in LES. Similarly, to

compare model simulations with observations, objects that refer to the same phenomena in simulations and in observations have to be defined. Well-defined objects can make the link between the various tools we have at our disposal to study atmospheric convection: they are the *Esperanto* making the construction of our Tower of Babel possible. In the next section, we will see how a well-defined object can serve as a bridge between GCMs, LES and observations with the example of cold pools.

2.5 Objects as a bridge between computer simulations and observations: the example of cold pools

2.5.1 Definition of a cold pool

Below clouds, the partial evaporation of precipitation may cool the air sufficiently to generate unsaturated downdrafts, which spread horizontally when reaching the surface under the form of density currents. These density currents are called “cold pools”. When spreading at the surface, cold pools can reach a size between 10 and 200 km in diameter and typically last less than a day (Zuidema, Torri, Muller, & Chandra, 2017). The edge of a cold pool is called a gust front. While the air inside a cold pool is relatively cold, thus less able to support buoyancy-driven convection, in some situations, the gust front has the ability to lift surrounding air masses and create new convective cells (Craig Goff, 1976; Warner et al., 1979). This triggering has been observed to be particularly effective when two or more cold pools collide (Droegemeier & Wilhelmson, 1985) or when the vorticity created by a cold pool counteract that from the low-level wind shear (Rotunno, Klemp, & Weisman, 1988). In addition, since they induce strong gusts near the surface, cold pools are suspected to enhance surface fluxes, and thus, to modify the thermodynamic properties of the subcloud layer (Langhans & Romps, 2015; Tompkins, 2001). For all these reasons, cold pools are thought to play an important role for the organization and the propagation of convection (Kurowski, Suselj, Grabowski, & Teixeira, 2018; Schlemmer & Hohenegger, 2014; Tompkins, 2001).

This description shows that today, a cold pool is a relatively well-defined object: it is created by a precise process, the evaporation of precipitation, is delimited spatially by a gust front, has a certain lifetime, and plays an important role for convection. This object can be detected in a variety of environments: on land and over oceans, in deep or shallow convective regimes, at various latitudes, etc. But making cold pools such a universal object and one of the building blocks of convection took time.

2.5.2 The construction of the object “cold pools”

Early views of cold pools in observations and in numerical simulations

Early studies on cold pools in observations were made as part of the Thunderstorm Project, a major meteorological study conducted by the United States just after World War II to study the nature and causes of thunderstorms (Byers & Braham, 1949; Newton, 1950). At that time, cold pools were seen as one of the properties of so-called squall lines, that is lines of thunderstorms forming along or ahead of a cold front. Newton (1963) referred to them as “pseudo-cold front” and already identify their capacity to regenerate convection by lifting the unstable warm air ahead of the squall line. A few years later, E. J. Zipser (1969) came to similar conclusions on the role of cold pools when analyzing data from the Line Islands Experiment conducted in

the eastern Indian Ocean in February-April 1967. Once again, in this study, cold pools as not yet seen as a separate object, but as an emerging features of squall lines – E. J. Zipser (1969) identified them as a particular kind of circulation induced by downdrafts (see also Fig. 2.5, taken from his subsequent paper E. Zipser, 1977). Analyzing data from a large tropical cloud cluster during GATE, Houze Jr and Betts (1981); Leary and Houze Jr (1979) later employed the terms outflows (or wakes) of convective-scale downdrafts to speak about cold pools. They both noted that such “outflows” exerted a strong control on where future convection broke out. These studies suggested that cold outflows should be represented in parameterizations of convection along with updrafts and downdrafts (Houze Jr & Betts, 1981). Although the term “cold pools” did not become widespread until the 1990s (see Fig. 2.8), this marks the birth of cold pools as an *object*.

In parallel, similar features were detected in the first numerical simulations of storm dynamics, first in two dimensions (Hane, 1973; Schlesinger, 1973; Takeda, 1971), and later in three dimensions (Klemp & Wilhelmson, 1978; Redelsperger & Lafore, 1988). These simulations were used to better understand the role of cold pools in triggering convection. In particular, Rotunno et al. (1988), based on two- and three-dimensional squall line simulations, concluded that the interactions between cold pools and the low-level shear were crucial to explain the formation of new convective cells and the longevity of squall lines. A key point in their demonstration concerns the generation of vorticity near cold pool edges, which can counteract the vorticity induced by low-level wind shear in certain situations. Following this study, D. J. Parker (1996); Q. Xu (1992) looked at cold pools as a particular kind of gravity (or density) current. They borrowed a fluid dynamics formalism developed by Benjamin (1968); Prandtl (1952) and other early studies to investigate the production of vorticity near the edge of a cold pool. In these studies, the properties of cold pools are studied independently of the storm that generates them, which underlines the emergence of cold pools as a distinct convective object.

Detection methods of cold pools

The apparition and the definition of cold pools as an object is also reflected in the appearance of the first methods to detect cold pools in high-resolution simulations and observations. As we have seen in section 2.2, there is a close link between detecting an object and defining it. As a consequence, the first attempts to detect cold pools in simulations led to define them more precisely (D. J. Parker, 1996; Skamarock, Weisman, & Klemp, 1994). In model simulations, cold pools were defined as a set of connected points fulfilling some conditions. Since cold pools are essentially density currents, a natural candidate to identify them was the buoyancy, defined as:

$$b = \frac{g(\theta_\rho - \bar{\theta}_\rho)}{\bar{\theta}_\rho} \quad (2.1)$$

where g is acceleration due to gravity, the overbar represents the horizontal domain mean, and θ_ρ the density potential temperature, defined following K. A. Emanuel et al. (1994) as

$$\theta_\rho = \theta(1 + 0.608 q_v - q_l) \quad (2.2)$$

with q_v and q_l are respectively the mixing ratios of water vapor and liquid water. In Tompkins (2001) for instance, cold pools were identified as regions where the buoyancy is lower than a given threshold. Since, many other methods have been proposed to detect cold pools in LES or CRMs, relying mostly on buoyancy, temperature or wind anomalies (see Drager & van den

Heever, 2017, for an overview).

In observations, Young, Perugini, and Fairall (1995) provided one of the early attempt to detect cold pools. Analyzing data from the Tropical Ocean Global Atmosphere (TOGA) Coupled Ocean-Atmosphere Response Experiment (COARE; Webster & Lukas, 1992), they defined the beginning of a cold pool time period through the onset of any rain shaft of at least 2 mm h^{-1} and the end of a cold pool time period by the end of the subsequent surface temperature recovery. Other methods were then used to detect cold pools in observations, using for instance abrupt wind shifts at the surface (Engerer, Stensrud, & Coniglio, 2008; Provod, Marsham, Parker, & Birch, 2016) or temperature drops in the surface timeseries (de Szoeke, Skillingstad, Zuidema, & Chandra, 2017; Vogel, 2017). The choice of a method rather than another depends on the observational data, the environmental conditions, and potentially the characteristics of the observed cold pools and/or the question of interest. As for models, there is therefore no consensus about the “best” method to detect cold pools from observations.

This lack of consensus reminds us that the construction of an object is always an idealization: what nature – and to some extent high-resolution simulations – gives us is only a single atmospheric flow, with some persistent heterogeneities. We have decided to call some of these heterogeneities “cold pools”, but depending on the situation, weather conditions or available data, what we call “cold pools” may correspond to different phenomena. The hope is that progressively, the various methods used by atmospheric physicists to define cold pools will become increasingly consistent, in order to guarantee that all the tools available to study cold pools speak the same language.

Cold pool parameterizations

Cold pool parameterizations are another aspect of the emergence of cold pools as an object. One of the early attempt to parameterize the effect of cold pools in GCMs is given in Qian, Young, and Frank (1998). Later on, Grandpeix and Lafore (2010) and Park (2014) proposed their own cold pool parameterizations for use in GCMs. Like any parameterization, these cold pool parameterizations are highly idealized. Hence, to build a parameterization, it is necessary to choose which characteristics of cold pools are important to represent, and which can be ignored. In Grandpeix and Lafore (2010), for instance, cold pools are represented as circular objects with vertical frontiers, whereas in Qian et al. (1998) there are represented as rectangular objects. Choices must also be made regarding the way cold pools interact with the other parameterizations. For example, Grandpeix and Lafore (2010) has been implemented in the atmospheric GCM LMDZ, which has two separate parameterizations for shallow and deep convection. In this implementation, the cold pools are generated and cooled by the deep convective downdrafts. In turn, cold pools trigger and feed the parameterization of deep convection. In Grandpeix and Lafore (2010), cold pools are therefore an internal feedback of convection.

Due to these different choices, parameterizing an object is a distinct epistemic activity than detecting it in observations or in simulations. In both cases, there is a construction: as we have seen, objects are not given as such in observations and we need to find a way to see them, that is to say a detection method and thus a certain definition of the object. Similarly, objects in a parameterization have to be constructed inside the frame of a given GCM. Nevertheless, detecting an object involves choosing only one aspect of the object – the one used in the detection method. The other characteristics of the object emerge freely from the observations or the simulations. On the contrary, by design, a parameterization defines both what an object is

and how it interacts with other objects and with the large-scale. In the process of defining an object, the formulation of a parameterization thus represents one step further compared to the choice of a detection method.

The above story, telling the emergence of cold pools as an object, can be summarized in three steps: 1) first, certain patterns that will be later called “cold pools” were seen in observations. 2) Then, cold pools were identified as one of the building blocks of convection, detected and studied both in observations and in high-resolution simulations. 3) Finally, the knowledge gained about cold pools was used in a parameterization and incorporated in a GCM. At each step, a stone was laid in the construction of cold pools as an object.

2.5.3 The role of cold pools as an object

The role of an object like cold pools encapsulates three main dimensions that we will detail in the next subsection: focusing atmospheric research, summarizing our understanding, and serving as a guide to analyze observations and high-resolution simulations.

Focusing atmospheric research

Naming an object focuses atmospheric research on that object. Cold pools were first named and identified as a building block of atmospheric convection in the 90s. Since, the number of studies on atmospheric cold pools exponentially increased, as shown in Fig. 2.8. Today, cold pools are one of the central objects around which research in atmospheric sciences is organized. A well-defined object such as cold pools leads to the definition of specific research questions, shared within certain research communities, which can be addressed using the many tools mentioned above. With these questions, objects break complex problems, for instance “understanding atmospheric convection” into a serie of simpler problems that can be treated separately, such as “in which conditions do cold pools trigger new convective cells?”. In other words, objects orient atmospheric research on some specific and solvable issues. Precise research questions motivate the formulation and test of hypotheses, thus focus and guide atmospheric research. This coordination brought by objects links otherwise isolated studies. In fact, in the introduction of many studies on cold pools, the relevance of the study to some recognized ongoing issues on cold pools is often cited as a main motivation for the work presented (see for instance the introduction of Drager & van den Heever, 2017; Gentine et al., 2016; Torri et al., 2015).

By focusing atmospheric research on some specific questions, objects make it possible to study a given hypothesis using a variety of tools, thus to assess its robustness. For instance, in the past decade, different studies focused on the role of cold pools in the diurnal cycle of convection over land. In particular, when the parameterization of Grandpeix and Lafore (2010) was implemented in the atmospheric GCM LMDZ, the diurnal cycle of convection over land was shifted to the late afternoon. This corrected a bias shared by many GCMs, where continental thunderstorms tend to peak with insolation, in contrast to observations where they tend to peak a few hours later. This GCM result led to the hypothesis that cold pools play a fundamental role in the diurnal cycle of convection over land. The role of cold pools in the diurnal cycle of convection was further tested in Haerter, Böing, Henneberg, and Nissen (2019), who showed in an idealized LES that cold pools played a role in the self-organization of convection, and thus potentially on its diurnal cycle, a result confirmed by Hirt and Craig (2021) in a realistic CRM simulation over Germany, and interpreted in Haerter (2019) using a simple conceptual model.

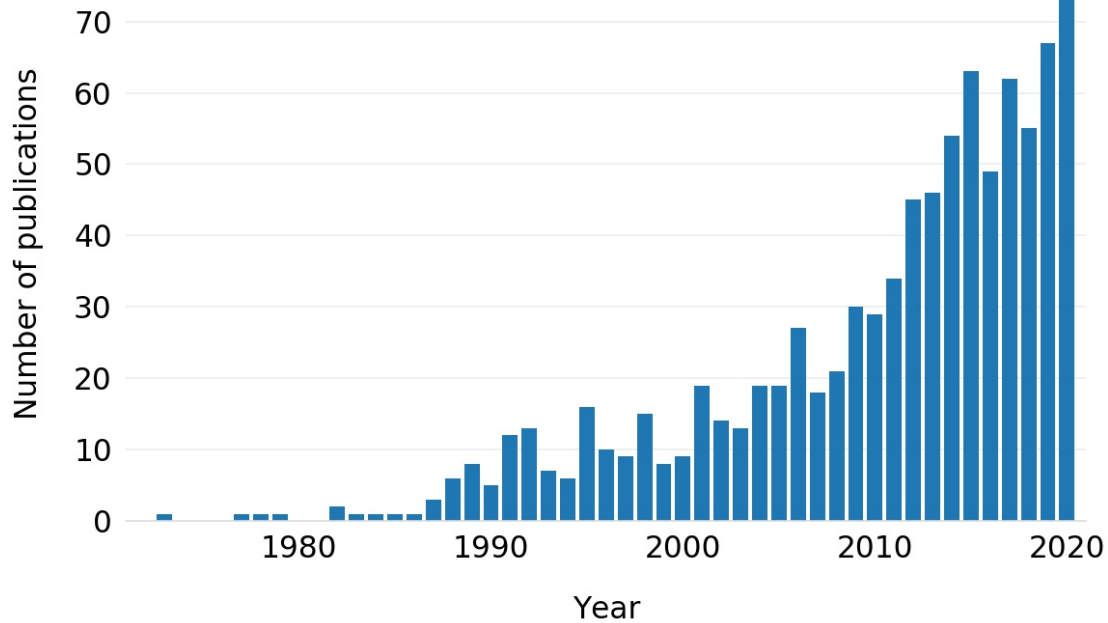


Figure 2.8: Occurrence of the term “cold pools” in the abstract or title of papers published in atmospheric sciences journals from 1970 to 2020. Data source: Google Trends.

With this example, we see how a given hypothesis – the role of cold pools in the diurnal cycle of convection – could be addressed and thus proven robust using a variety of tools. It illustrates how objects can serve as a catalyst to tackle certain questions and robustly test hypotheses.

Summarizing our understanding

Objects connect different studies at two levels: first, as we have just seen, they relate studies dealing with the same research questions, thus allowing us to study the robustness of the hypotheses proposed to answer them. Then – and this is what we will see here – they connect isolated hypotheses about a certain phenomenon, and thus enable incremental progress in the understanding of this phenomenon. Objects give us words to which we can attach our understanding of certain phenomena which share similar characteristics. For instance, if two studies A and B successfully prove two different hypotheses about the same phenomenon, we will be able to connect the results of A and B only because we have a word to name this phenomenon, thus to recognize it both in A and in B. More generally, objects help us to connect studies that all give a partial vision of one phenomenon in order to understand this phenomenon globally.

The conceptual model given in Rochetin et al. (2021) and shown in Fig. 2.9 best illustrates this power of object. Indeed, it summarizes many studies about cold pools, which all give a partial understanding about what a cold pool is: for instance, observations from the field campaign GATE (Houze Jr & Betts, 1981) led Rochetin et al. (2021) to represent a cold pool as a mixed layer near the surface and a stratified layer higher up, numerical studies from Hirt et al. (2020); Romps and Jeevanjee (2016) and Torri et al. (2015) supported the presence of lifting ahead of the gust front, etc. Such a conceptual model also helps to formulate and thus motivate unresolved research questions: in particular, Rochetin et al. (2021) explain that the subsidence and the “capping inversion” above cold pools are more speculative and should

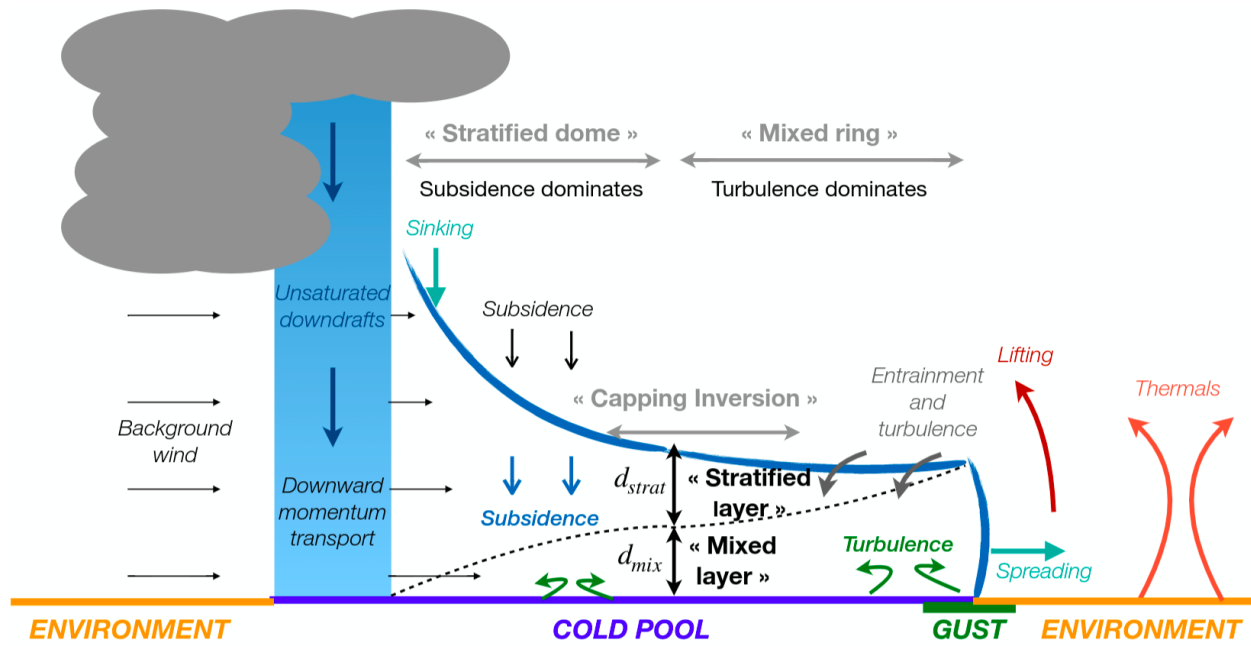


Figure 2.9: Conceptual diagram of a cold pool and its interactions with clouds and its environment. From Rochetin et al. (2021)

be studied further. By summarizing what we know, objects help to assess the limits of our knowledge and thus to push back these limits.

Analyzing computer simulations and observations

“If, while contemplating phenomena, we did not immediately link them to some principles, not only would it be impossible for us to combine these isolated observations, and, consequently, to draw any fruit from them, but we would even be entirely unable to remember them.”¹ wrote Auguste Comte in 1830 in his *Course of Positive Philosophy* (Comte, 1835). According to the philosopher, we need some principles to study phenomena. These principles determine what we can learn from observations. In the study of atmospheric convection, and more generally of climate, we have seen that some of these principles can be summarized into objects. In particular, we explained previously how the object “cold pool” has been progressively constructed from observational and numerical studies. These studies led to an increasing understanding of the characteristics of cold pools and of their role for convection; the state of the art of what is known today about cold pools is illustrated in Fig. 2.9. This process suggests a linear road from observations to understanding. Here, we mean observations in a broad sense, that is both observations of real phenomena and high-resolution simulations where cold pools are resolved and can also be “observed”. Comte opens a different path: our accumulated knowledge on cold pools could help us to better observe them. The journey from observations to understanding would then be more like a series of round trips than a straight line.

In fact, our understanding of an object has an influence on how we observe and see these objects in high-resolution simulations or observations. We illustrate it with two recent studies

¹ “Si en contemplant les phénomènes, nous ne les rattachions point immédiatement à quelques principes, non seulement il nous serait impossible de combiner ces observations isolées, et, par conséquent, d’en tirer aucun fruit, mais nous serions même entièrement incapables de les retenir”, Auguste Comte, *Cours de philosophie positive*, première leçon. Traduit par nos soins.

on cold pools in which we have been involved.

Rochetin et al. (2021) – A physically-based definition of convectively generated density currents: detection and characterization in convection-permitting simulations This first study proposes a method to detect and study cold pools in CRMs and LES. This method relies on the conceptual model given in Fig. 2.9. As in this conceptual model, cold pools are defined horizontally as a cold area adjacent to at least one gust, and vertically, as the combination of a mixed layer and a stratified layer. Rochetin et al. (2021) explains how this definition is used in practice to build a detection method. This method is “physically-based” in the sense that it relies on a physical understanding of what a cold pool is.

We tested the use of the method proposed by Rochetin et al. (2021) in LES simulations over the western tropical Atlantic, using the ICOSahedral Non-hydrostatic (ICON) model (Zängl, Reinert, Rípodas, & Baldauf, 2015), with 313 m horizontal grid spacing (see Dipankar et al., 2015; Heinze et al., 2017; Vial et al., 2019a, for more details about this simulation). The identification of cold pools provides access to quantitative and statistical information about them, such as their number N_{dens} and their fractional coverage f_{dens} . U_{mix}^{gust} – the mean horizontal wind velocity in gusts over the height of the mixed layer – is also used as a proxy for estimating the average cold pool propagation speed c_{dens} . Finally, the mean equivalent radius r_{dens} of cold pools can easily be deduced from their spatial density: $D_{dens} = N_{dens}/D$, with D , the domain size:

$$\langle r_{dens} \rangle = \sqrt{\frac{f_{dens}}{\pi D_{dens}}} = \sqrt{\frac{f_{dens} D}{\pi N_{dens}}} \quad (2.3)$$

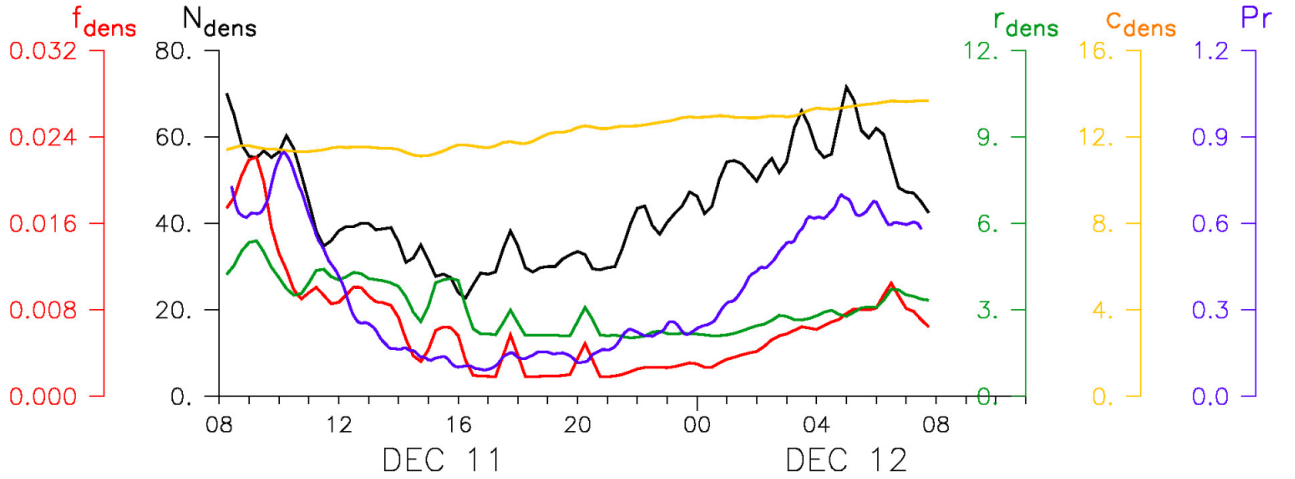


Figure 2.10: Time series of cold pool fractional coverage f_{dens} (red), cold pool number N_{dens} (black), domain-mean cold pool radius r_{dens} (km, green), domain-mean cold pool gust velocity c_{dens} (m s^{-1} , yellow) and domain-mean surface rain rate Pr (mm hr^{-1} , blue).

Figure 2.10 shows the time series of these different variables over one simulation day, starting at 8 a.m. (local time) on December, 11 2013. We see that the number of cold pools N_{dens} follows qualitatively the precipitation rate at the surface, whereas the cold pool mean radius is relatively constant across the simulation. This suggests that more precipitation at the surface leads to an increasing number of cold pools, but not necessarily larger cold pools, an hypothesis that could be tested in other simulations and observations. As explained in Rochetin et

al. (2021), N_{dens} , f_{dens} , c_{dens} , and r_{dens} can also be used to evaluate the realism of the parameterization of Grandpeix and Lafore (2010), as the latter uses these different variables as parameters (for N_{dens}) or prognostic variables (for f_{dens} , r_{dens} and c_{dens}). In fact, having data from LES or CRMs to test Grandpeix and Lafore (2010) parameterization is an important motivation behind Rochetin et al. (2021). This shows that theory – here, under the form of a parameterization – determines not only how we detect cold pools, but also which features of cold pools we characterize once cold pools have been defined. Once again, this illustrates the power of objects to create multiple interactions between different research tools.

Touzé-Peiffer et al. (2021) – Detecting cold pools from soundings during EUREC⁴A

We illustrate the role of our understanding of an object on how we can observe it with a second study, whose preprint version is given in appendix B. In this study, we propose a detection method to analyze cold pools from data of the field campaign EUREC⁴A (Elucidating the role of clouds-circulation coupling in climate; Bony et al., 2017), which took place in January and February 2020 over the Atlantic ocean east and south of Barbados. During EUREC⁴A, more than 2000 atmospheric profiles were measured using radiosondes and dropsondes launched from aircraft, research vessels and a surface observatory (Stevens et al., 2021). This unprecedented data set represents a unique opportunity to study the properties of cold pools over tropical oceans. However, to make sense of this large amount of data, a method to detect cold pools from atmospheric soundings is needed.

As in Rochetin et al. (2021), to build this detection method, we relied on the conceptual model shown in Fig. 2.9. The challenge was however to project the definition of cold pools given by this conceptual model on the space of available observations, that is 1D vertical profiles. Most cold pool detection methods, including Rochetin et al. (2021), use the fact that the air inside a cold pool is colder than the air *around* it. This is unapplicable to soundings, which are by definition pointwise in the horizontal plane. Instead, we decided to use the fact that cold pool air is colder than the air *above* it. In terms of virtual potential temperature, a cold pool is indeed colder, thus denser than the subcloud layer air on top of it. Consequently, a sharp increase in virtual potential temperature θ_v is expected at the top of a cold pool, as suggested by the presence of a stratified layer in Fig. 2.9. Over tropical oceans, in convective regimes, this contrasts with the subcloud layer outside cold pools, which tends to be well-mixed in θ_v up to cloud base (e.g. Cuijpers & Duynkerke, 1993; Pennell & LeMone, 1974). This led us to use a cold pool detection method based on the height of the mixed layer (H_{mix}).

Following Canut, Couvreur, Lothon, Pino, and Saïd (2012) and Rochetin et al. (2021), we defined H_{mix} as the lowest altitude Z above $Z_{min} = 100$ m where the virtual potential temperature θ_v is higher than its mass-weighted average from Z_{min} to Z by a fixed threshold $\epsilon = 0.2$ K:

$$\begin{aligned} \theta_v(Z) &\geq \tilde{\theta}_v + \epsilon \\ \text{with } \tilde{\theta}_v &= \frac{\int_{Z_{min}}^Z \rho(z) \theta_v(z) dz}{\int_{Z_{min}}^Z \rho(z) dz} \end{aligned} \quad (2.4)$$

ρ being the density of the air. Setting Z_{min} at 100 m is necessary due to the presence of unphysical temperature peaks below 100 m for a few radiosondes. The virtual potential temperature is calculated assuming that the air of the lowest layers is not saturated, so that the mixing ratio of liquid water in the air can be neglected. It is then approached as: $\theta_v = \theta(1 + 0.61r)$, r

being the mixing ratio of water vapor. The calculation of H_{mix} thus requires only the vertical profiles of pressure, temperature and humidity at a single point and is directly applicable to soundings.

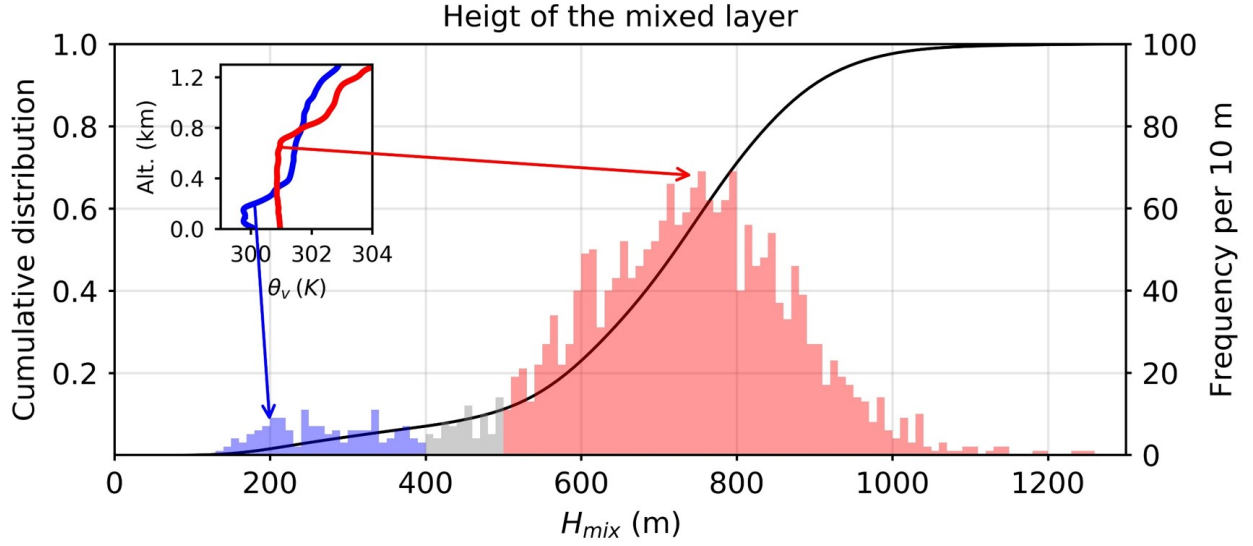


Figure 2.11: Cumulative distribution function (black) and histogram for the height of the mixed layer for all radiosondes and dropsondes launched during EUREC⁴A. The bins of the histogram are 10 m wide. Colors indicate cold pool soundings (blue), unclassified soundings (grey) and environmental soundings (red). The upper left panel shows two examples of θ_v profiles, one in a cold pool and the other in the environment.

Figure 2.11 shows the H_{mix} distribution for EUREC⁴A soundings. The histogram reveals a negatively skewed distribution, with a median of 720 m. Assuming that the left tail of the distribution is due to cold pools, we choose to define “cold pool soundings” as those with H_{mix} lower than 400 m (7% of soundings, in blue), and “environmental soundings” as those with H_{mix} higher than 500 meters (90% of soundings, in red). With this definition, only 3% of soundings (in grey) are neither in cold pools nor in the environment.

To show that H_{mix} provides a robust detection of cold pools, we tested our detection method in a high-resolution simulation over the Atlantic Ocean, upstream of Barbados. This simulation was the same as the one used in Rochetin et al. (2021) and described previously, that is the LES version of the ICON model, with realistic boundary conditions and 313-m horizontal grid spacing (Dipankar et al., 2015; Zängl et al., 2015). Figure 2.12 shows that our detection methods gives consistent results with the surface temperature and precipitation fields. It represents a snapshot of the mixed layer height H_{mix} (left) and the temperature near the surface ($z \approx 50$ m, right) on 12 December 2013 0930 UTC. In the right panel, we see that regions with negative temperature anomalies are co-located with significant surface precipitation (red dots) and also with important wind shifts (not shown), suggesting that these cold regions are in fact convective cold pools. In the left panels, we apply our detection method by circling in yellow regions where the mixed layer is less than 400 meters. Qualitatively, there is a really good agreement between cold pools detected with our method and regions strongly cooled by rainfall. This gives us confidence to choose H_{mix} to detect cold pools from soundings during EUREC⁴A.

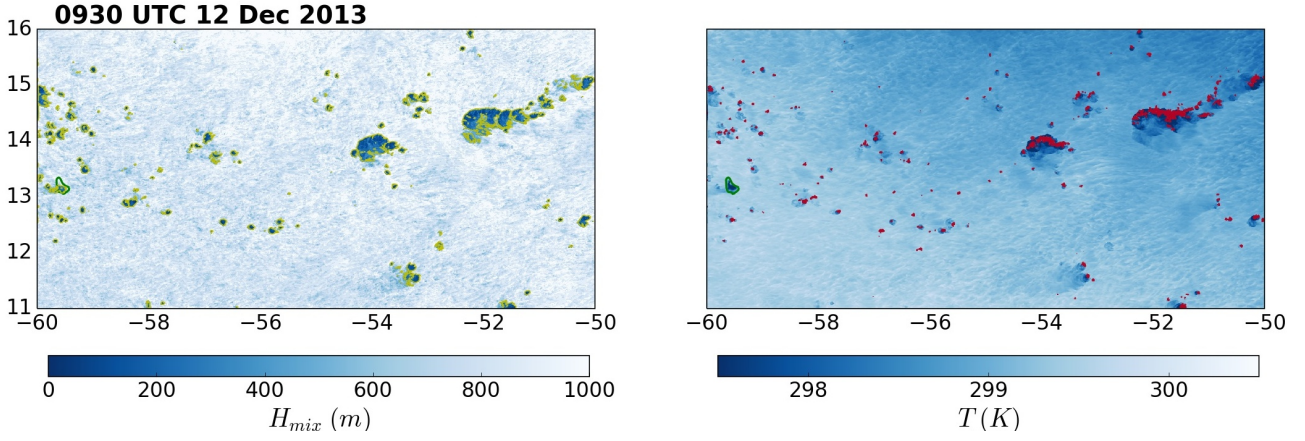


Figure 2.12: Height of the mixed layer H_{mix} and surface temperature T at $z \approx 50$ m in an area of $5^\circ \times 10^\circ$ upstream of Barbados (circled in green) on December 12, 2013 at 0930 UTC. In the left-hand panel, cold pools are circled in yellow as regions where H_{mix} is lower than 400 meters. In the right-hand panel, the red dots represent locations where surface rainfall is greater than 10 mm day^{-1} .

This example is interesting for our discussion on objects, because we were able to assess the credibility of our method to detect cold pools from atmospheric soundings using 1) a LES, which illustrates the complementarity between various tools to study an object and 2) some understanding of what a cold pool is, in order to check whether the objects identified with our method in LES are consistent with what we would expect. Without such understanding, we would not have been able to build our detection method and test it. This shows that **how we understand an object determines how we see it in observations**.

In turn, our detection method has enriched our understanding of cold pools. The full study is provided in Appendix B, and we give here only an example of its results. This result concerns the impact of the wind shear on the spreading of a cold pool. It is shown in Fig. 2.13, where $\Delta_{SCL}(u)$ and $\Delta_{SCL}(v)$ refers respectively to the difference of zonal and meridional wind speed in the subcloud layer between 100 and 500 m, that is in and above cold pools when there is one. In the environment (in red), there is little difference in terms of wind speed between the two layers, consistent with what we would expect for a relatively well-mixed layer. On the contrary, in cold pools, the average wind difference is about 3 m s^{-1} between the two layers. Furthermore, zonal and meridional wind distributions show that cold pools spread in all directions, consistently with the conceptual picture of a density current propagating over a solid boundary. In the presence of vertical wind shear, theories and numerical simulations nevertheless predict that cold pool spreading may not be perfectly isotropic, as vertical momentum transport might favor the propagation of cold pools downshear (Grant, Moncrieff, Lane, & van den Heever, 2020; Mahoney, Lackmann, & Parker, 2009; Moncrieff, 1992). Fig. 2.13 middle and right panels test this hypothesis by showing the zonal and meridional wind difference between 100 and 500 m ($\Delta_{SCL}(u)$ and $\Delta_{SCL}(v)$) for three zonal and meridional wind shear ranges in the cloud layer, between 1 and 2 km: $]-\infty; -1]$, $]-1, 1]$ and $]1, +\infty[$ (m s^{-1}). The box and whisker diagrams reveal that $\Delta_{SCL}(u)$ is gradually shifted to higher values as the zonal wind shear in the cloud layer increases. On the contrary, in environmental soundings, the zonal wind shear in the cloud layer has little, if any, influence on $\Delta_{SCL}(u)$. The same result stands for the meridional wind, suggesting significant momentum transport by downdrafts in cold pools in all directions. Such

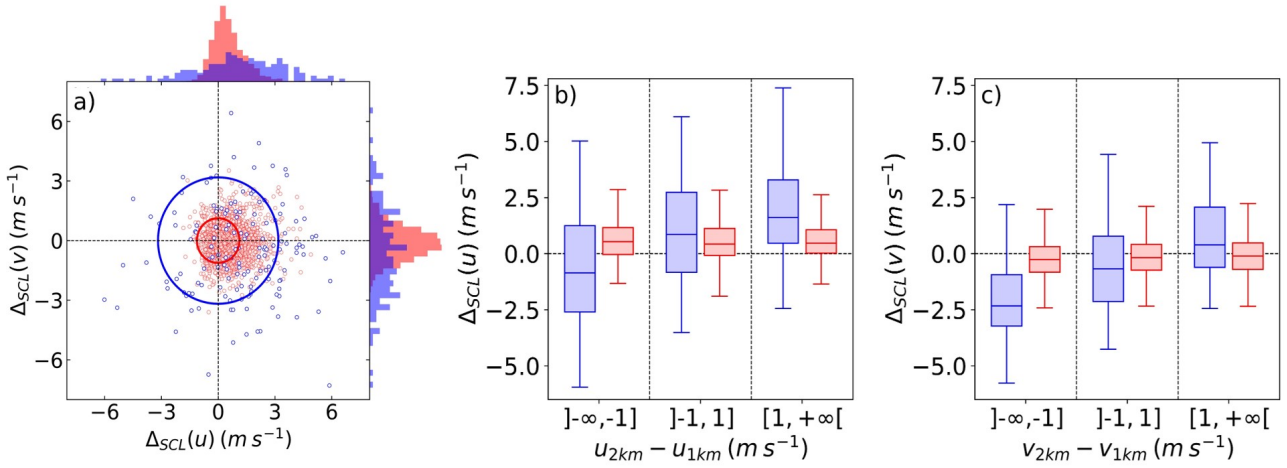


Figure 2.13: (a) Scatter plot of the difference of meridional wind speed at 100 m and at 500 m $\Delta_{SCL}(v)$ for cold pool (blue) and environmental (red) soundings. On the upper and right sides of the graph, marginal distributions of each variable are also reported for cold pools and their environment, with a bin width of 0.3 m s^{-1} . The circles represent the mean of the absolute wind speed distributions, that is the mean of $\sqrt{(\Delta_{SCL}(u))^2 + (\Delta_{SCL}(v))^2}$ for cold pool and environmental soundings. (b) Box and whisker plots for the zonal wind difference between 100 and 500 m $\Delta_{SCL}(u)$ for three zonal wind shear ranges in the cloud layer. The wind shear ranges are defined by calculating for each sounding the difference between $u_{2km} = \int_{1.8km}^{2.2km} u(z)dz$ and $u_{1km} = \int_{0.8km}^{1.2km} u(z)dz$. Horizontal line within each box represent the median, box bottom and top are 1st (Q1) and last (Q3) quartile of the distributions, and the whiskers extend up to 1.5 interquartile range above Q3 and below Q1. (c) Same plot, but for the meridional wind difference between 100 and 500 m $\Delta_{SCL}(v)$ for three meridional wind shear ranges in the cloud layer.

momentum transport by downdrafts in cold pools had been suspected in conceptual models and numerical simulations, but never shown nor quantified in observations.

This new result would not have been possible without a robust detection method applicable to sounding measurements. This demonstrates in practice the importance of having relevant tools, based on well-understood physical principles, to dissect observations. Our understanding is in fact incremental: what we understand gives us an angle of attack to analyze observations that allow us to understand even more.

More generally, Rochetin et al. (2021) and Touzé-Peiffer et al. (2021) illustrate the many interactions between observations, numerical simulations, and our theoretical understanding in the case of cold pools. Well-defined objects such as cold pools appear as bridges between different tools to study atmospheric convection: they focus atmospheric research, connect different studies, help formulate specific hypotheses and assess their robustness, summarize our understanding and give us reading grids to analyze observations and numerical simulations.

2.6 Conclusion

In this chapter, we defined what objects are and what role they play in our understanding of atmospheric convection. Objects are a way to cut the atmosphere in different components that present similarities. The study of objects and their interactions provides a mechanistic understanding of atmospheric convection. We explained that such mechanistic explanation is not specific to the study of atmospheric convection, but is frequently used in physiological sciences such as neuroscience. Many physical systems can indeed be understood by being broken down into different components. However, the specificity of the objects used to study convection is that they are in part artificial: their frontiers are not given by nature, but have to be constructed. In observations, it means that some criteria have to be chosen to detect an object, i.e. to isolate it from its environment. Once an object is detected in observations, its characteristics can be studied. Nevertheless, since an object is by definition four dimensional (3 spatial dimensions + time) and observations are at best three dimensional (2 spatial dimensions + time), they give only a partial view of an object. The limitations of observations motivate the use of numerical simulations to track objects in time, and study their interactions with their environment and with other objects. We distinguished simulations in which objects are simulated explicitly from simulations in which they are simulated implicitly. In each case, we precised the kind of understanding given by objects and why this understanding is necessarily limited. We then argued that well-defined objects can link these different tools in order to test the robustness of certain assumptions. The various roles an object can have for atmospheric research were finally illustrated with the example of cold pools.

At the end of this chapter, we showed that well-defined objects such as cold pools determine how we observe natural phenomena, and we interpreted this role of objects in light of Auguste Comte's positivism. This discussion on objects reminds us of another one, which took place in the first half of the XXth century between the founding fathers of quantum mechanics. Quantum physicists, destabilized by the many paradoxes of quantum physics, felt that they had to build a new science on solid epistemological foundations. In this context, Heisenberg chose to follow the guiding principle: "any good theory must be based only on quantities which can be directly observed" (Heisenberg, 1970). The famous author of the uncertainty principle had indeed been influenced by the positivism of Mach, according to which "the goal of physics is the simplest and cheapest abstract expression of the facts" (Mach, 1882). For the Austrian physicist and philosopher, the formalism used in physics had to be as close as possible to what we can observe in the real world. The theory developed by Heisenberg was thus immediately linked to experience: his matrices were simply the translation in mathematical terms of quantities that could be observed experimentally (Heisenberg, 1970). On the contrary, influenced by the positivism of Auguste Comte, Einstein considered that we should not neglect the importance of theory on our observations. According to him, it is the theory alone that decides what we can observe. Any experiment needs a theory to be interpreted and a quantity is therefore only observable in relation to an old theory. This debate, at first sight anecdotal, had profound implications on the development of quantum physics and inspired in particular Heisenberg's famous uncertainty principle.

Likewise, we hope that our reflection on objects will have a positive influence on any atmospheric scientist who reads it. We hope to have convinced him that objects are not just given by Mother Nature, but are the result of an artificial construction of our human mind. As a consequence, we are free to choose new objects or to redefine them, which often means

opening new fields of research. The classification of sugar, gravel, fish and flowers given in Stevens, Bony, et al. (2019) is probably one of the best examples. Let's take a step back and reflect: are the objects used today the most relevant? Do they allow the formulation of specific research questions that can be answered with the available research tools? Do they act as a bridge between different studies? If they do, let's continue to use these objects. If they don't, let's be creative and imagine new ways to talk about what we observe. For it is our language that defines what we understand.

Chapter 3

The semi-empirical world of convection parameterizations

All models are wrong, but some are useful.

George Box

Contents

| | | |
|------------|--|-----------|
| 3.1 | Introduction | 66 |
| 3.2 | Why do we need to parameterize atmospheric convection? | 67 |
| 3.2.1 | The need for GCMs to study climate change | 67 |
| 3.2.2 | The need to parameterize atmospheric convection in GCMs | 68 |
| 3.3 | In pursuit of a theoretical foundation for mass-flux convection parameterizations | 71 |
| 3.3.1 | The mass-flux formalism | 71 |
| 3.3.2 | Quasi-equilibrium filtering – Arakawa and Schubert (1974) parameterization | 73 |
| 3.3.3 | Statistical interpretation – Craig and Cohen (2006) stochastic parameterization | 77 |
| 3.3.4 | Segmentally constant approximation – Yano (2014) | 79 |
| 3.4 | Object-based interpretation of LMDZ convection parameterizations | 82 |
| 3.4.1 | Overview of LMDZ convection parameterizations | 82 |
| 3.4.2 | Interpreting LMDZ multi-object parameterizations | 86 |
| 3.5 | Conclusion | 90 |

3.1 Introduction

In atmospheric general circulation models (GCMs), a distinction is often made between the “dynamics” or “dynamic core”, representing the large-scale transport of heat, mass and moisture based on well-established equations of fluid dynamics and thermodynamics, and the “physics”, representing the collective influence of subgrid processes on grid-scale conditions using diverse parameterizations (Guillemot, 2017; W. S. Parker, 2006). Contrary to the dynamics, parameterizations are not solidly anchored in the theory. They rely both on some theoretical principles and some empirical relationships derived from observations or process studies using high-resolution models. Parameterizations are thus often described as “semi-empirical” – they stand between theory and observations (Edwards, 2010; Katzav, 2013).

Many GCM biases have been attributed to the poor representation of subgrid processes in parameterizations. In particular, convection parameterizations constitute a major source of uncertainty for climate change projections (Stevens & Bony, 2013) and were sometimes described as “the Achilles’ heel” of climate modeling (Stone & Risbey, 1990). Moreover, despite significant progress were made since the first convection parameterizations were introduced in GCMs 50 years ago (Rio et al., 2019), the rate of progress associated with parameterizations has been considered too slow, and parameterizations were sometimes described as a “deadlock” (T. Palmer & Stevens, 2019; D. Randall et al., 2003). To break this deadlock, some scientists have proposed to replace parameterizations with high-resolution models using super-parameterizations (D. Randall et al., 2003; D. A. Randall, 2013) or global cloud-resolving models (GCRMs, Satoh et al., 2019; Stevens, Acquistapace, et al., 2020), or with machine-learning algorithms (Gentine et al., 2018).

Criticisms against parameterizations do not only concern their effect in GCMs, but also their scientific basis. In particular, the semi-empirical nature of parameterizations is often seen as a weakness. For instance, Edwards (2001) argues that the “fuzzy relationship” of parameterizations with observational data is contradictory with the reductionist imperative of the physical sciences, which attempt to “explain large-scale phenomena as an outcome of smaller-scale processes”. Since they are based partly on observations, parameterizations would be less scientific grounded than for instance GCRMs or super-parameterizations, which represent subgrid processes using first principles. However, Edwards (2001) does not explain why parameterizations should meet the reductionist standards. More generally, while many philosophers of science have mentioned the semi-empirical nature of parameterizations, few studies have defined precisely in which sense parameterizations are partly determined by theory, partly determined by observations, thus whether and why their semi-empirical nature is legitimate, or not.

Answering these questions is the objective of this chapter. We adopt a partly historical approach to explain the context in which parameterizations have emerged and how the parameterization problem has been originally framed. With different examples, we illustrate two extreme positions that have been proposed to build parameterizations: deduce logically parameterizations from first principles – what we called theory-based parameterizations – or rely mostly on observations – observation-based parameterizations. We underline the limits of each approach and motivate the simultaneous use of observations and theory to build parameterizations. We argue that the semi-empirical world of convection parameterizations can be best understood in terms of objects, as we defined them in the previous chapter. We give examples of object-based parameterizations used in the atmospheric model LMDZ and justify why

objects are a useful framework to make the best use of both theory and observations.

3.2 Why do we need to parameterize atmospheric convection?

To understand why we need to parameterize atmospheric convection, we need first to understand why the models in which parameterizations are used, GCMs, have appeared historically as a natural way to study climate change. This is the objective of the following subsection (3.2.1), where we trace the origins of GCMs, and explain how they gradually emerged as a prominent tool for studying the general circulation of the atmosphere, and then climate and climate change.

3.2.1 The need for GCMs to study climate change

Early climate change studies using 1D radiative-convective models

Many of the early studies on the temperature rise caused by an increase in atmospheric CO₂ concentration were conducted using simple, one-dimensional model: so-called Radiative Convective Models (RCMs). RCMs allow to simulate energy transfer in one vertical atmospheric column. At each vertical level, RCMs calculate shortwave and longwave radiation fluxes by taking into account various parameters such as surface albedo, cloud amount and atmospheric humidity. In such models, the effect of atmospheric convection is represented in a simplified way called “convective adjustment”: when the vertical temperature profile becomes unstable, i.e. when density starts to increase with height, convection is assumed to take place and the temperature profile is adjusted towards a stable profile. This convective adjustment is designed to move heat from near-surface levels upwards (Manabe & Strickler, 1964).

Despite their simplicity, RCMs proved valuable to isolate important feedbacks, such as water vapor feedback. In particular, using a RCM, Manabe and Wetherald (1967) represented a breakthrough in our understanding of climate change induced by an increase of CO₂. They tested a RCM under present-day conditions and under a doubling of CO₂ concentration in the atmosphere. Assuming the relative humidity would not change under an increase of temperature, they calculated a 2.4°C warming under a doubling of CO₂. This result is still consistent with the most recent statements of the IPCC, which estimate that the climate sensitivity is “likely between 1.5°C and 4.5°C” (Stocker et al., 2013). Many other results given by Manabe and Wetherald (1967) are equally valid today. In particular, the two authors predicted a significant stratospheric cooling under an increase of CO₂, a result that has been confirmed by observations in the past 50 years (e.g. Ramaswamy et al., 2006; Randel et al., 2009). They also tested changes in clouds at different levels and found that an increase in low clouds would cool surface temperatures, while changes in high clouds could instead warm them, again consistent with what we know today (e.g. Voigt et al., 2021).

These results were promising in terms of our ability to use numerical models to better understand various climate feedbacks. Nevertheless, Manabe and Strickler (1964) pointed out that RCMs were limited by their inability to describe horizontal transports in the atmosphere. They suggested that numerical models able to simulate the large-scale movement of air masses would be useful to solve these limitations, opening the door to the use of GCMs in climate studies.

The genesis of GCMs

GCMs were derived from numerical models used for weather predictions. The early 1950s marked indeed the advent of numerical weather prediction in the United States. The first weather forecast by means of an electronic computer was run by Jule Charney and his team at the Institute for Advanced Studies (IAS) in Princeton in 1950. This forecast relied on fluid dynamics equations ingeniously simplified into a so-called “barotropic vorticity equation”, easier to run on a computer. The horizontal grid used for the first weather prediction covered North America and had 270 points about 700 km apart. In the following years, other forecasts were run and proved the feasibility of using numerical models to simulate the large-scale movement of air masses. This paved the way to the more ambitious goal of simulating the entire atmosphere, and thus climate.

Norman Phillips, a young meteorologist working in Charney’s team at the IAS, was the first to simulate the general circulation of the atmosphere using a numerical model. The numerical model he considered was similar to the ones used for numerical weather prediction, but contained two vertical levels and covered a larger area. Phillips chose a modeling surface of 10000 km by 6000 km – 10000 km being approximately the distance between the equator and the poles on the Earth’s surface, and 6000 km a width deemed large enough to capture large-scale eddies. To construct the grid of his model, he divided this surface into rectangles of 625 km along the y-axis (latitudes) and 375 km along the x-axis (longitudes). In addition, he imposed periodic conditions along the x-axis, so that his domain was cylindrical. Phillips launched his model for a duration of 31 days, with a time step of one hour. The results were convincing. Even if Phillips had not taken into account moist phenomena and in particular had ignored the influence of clouds, a circulation similar to the Hadley-Ferrel circulation appeared in this model, as well as large-scale eddies comparable to the cyclones and anticyclones typically observed in the atmospheric circulation (N. A. Phillips, 1954).

Phillips’ work intrigued American meteorologists and triggered further work on GCMs. A specialized research unit on GCMs was created in 1955 a few months after the publication of Phillips’ results and placed under the responsibility of Joseph Smagorinsky. In 1963, this unit became the Geophysical Fluid Dynamics Laboratory (GFDL) and hosted in particular Manabe and Wetherald, who developed their RCM to study the effect of an increase in CO₂ on global temperatures. The use of GCMs to study climate change was thus born at the crossroads of two successes: 1) on the one hand, the first tests of GCMs, and Phillips’ model in particular, had shown the ability of these models to qualitatively predict large-scale atmospheric motions. 2) On the other hand, RCMs had proven successful in understanding climate feedbacks, but were limited because, unlike GCMs, they were one-dimensional and therefore did not capture the effects of large-scale horizontal transport in the atmosphere.

3.2.2 The need to parameterize atmospheric convection in GCMs

Moist convective adjustment

When the first atmospheric GCMs were developed in the 1960s, a major issue appeared: the atmospheric state given by GCMs in the tropics was sometimes unstable, leading to an exponential growth of intense vertical motions at certain grid points – so-called “grid-scale atmospheric convection” (J. Charney & Ogura, 1960). This was due to the fact that the grid used in GCMs was far too coarse to resolve atmospheric convection. As discussed in chapter 1, convection acts

as a negative feedback for large-scale instabilities, but takes place at a much smaller scale than the GCM grid: resolving atmospheric convection requires numerical models with a horizontal resolution of one kilometer or less. Grid-scale atmospheric convection that appeared when the atmospheric profiles became unstable were thus very unrealistic. When they developed, errors exponentially increased and GCMs were not able to predict accurately the observed large-scale atmospheric circulation (Kasahara, 1961; Lilly, 1962).

To get a realistic large-scale atmospheric circulation using GCMs, it was necessary to find a way to account for the presence of atmospheric convection within each grid box of a GCM, hence to prevent the simulated atmospheric profile from becoming unstable. This was the initial motivation behind parameterizations of atmospheric convection. The basic idea of the first parameterization schemes (Manabe, Smagorinsky, & Strickler, 1965; Mintz, 1968; Smagorinsky, Manabe, & Holloway Jr, 1965) was to adjust the temperature lapse rate when it exceeded the moist adiabatic value. This procedure had already been used with success in RCMs and referred to as “moist convective adjustment” (Manabe & Strickler, 1964). This simple strategy helped early GCMs to overcome the pathological difficulty of predicting realistic vertical thermodynamic profiles.

The authors of this first parameterizations of atmospheric convection were well aware of the limitations of the solution they proposed. As Manabe et al. (1965) explained:

Since the grid-scale convection cannot be resolved by the grid itself, the computation quickly deteriorates. Therefore, it is desirable to design a scheme of convection such that the grid-scale convection does not develop. [...] In view of our ignorance in this matter, we used a very simple scheme of convective adjustment [...] and successfully avoided the abnormal growth of grid-scale convection. It is hoped that the results obtained can provide the basis for the computation of a more exact and better scheme of convection. (Manabe et al., 1965, p. 770)

Despite these limitations, the early moist convective adjustments were sufficient to make GCMs realistic enough to be useful.

On the use of ad hoc hypotheses in GCMs

The first parameterizations of atmospheric convection, moist convective adjustments, appeared for practical reasons – correcting the large-scale instability observed in the GCMs – rather than for well-established scientific grounds. According to the Merriam-Webster dictionary, something is “ad hoc” if it is “formed or used for specific or immediate problems or needs” (Merriam-Webster, 2021). Moist convective adjustment could be seen as a typical ad hoc hypothesis, because it was specifically added to the GCM formulation to solve the problem of convective instability. Is the addition of such an ad hoc hypothesis to correct GCM results legitimate?

The history of sciences is full of ad hoc hypotheses that were added to scientific theories to make them consistent with observational data. Some of these ad hoc hypotheses turn out to be true – a famous example being the ad hoc introduction by Albert Einstein of the cosmological constant to the Theory of General Relativity in order to explain why the universe is static. Nevertheless, the use of ad hoc assumptions to correct a scientific theory is often considered questionable. The philosopher of science Popper explained that the increasing addition of ad hoc hypotheses to a scientific theory makes it more and more complex and less and less falsifiable, which means that it cannot be contradicted by empirical data. According to him, one

of the characteristic of pseudo-sciences such as Marx's theory of history and Freudian psychoanalysis is to rely on ad hoc assumptions that allow them to justify almost any empirical fact (Popper, 1982). If each time observational data disagree with a certain theory, a new ad hoc hypothesis is added to it, there is indeed a problem of circular reasoning and it will be difficult to use the predictive power of the theory to test it.

Nevertheless, the use of ad hoc hypothesis in numerical models such as GCMs does not have the same status as in scientific theories. Indeed, in a GCM, when Navier-Stokes equations are discretized over a given grid to simulate the large-scale circulation of the atmosphere, the objective is not to test Navier-Stokes equations. It is taken for granted that these equations are well established, and on the other hand, it is known that the way these equations are simplified and discretized over a grid with a resolution of several hundred kilometers introduces significant errors. Despite these errors, the hope is that GCMs will be sufficiently realistic to be useful, i.e. to help understand certain features of the atmospheric circulation, or certain climate feedbacks when used to simulate climate change. If some corrections can make GCM results more realistic and therefore more useful, it seems legitimate to use them.

A second fundamental reason for the use of a moist convective adjustment in GCMs is that it is attributed to a specific phenomenon: convection. It is possible to interpret qualitatively what is represented with such adjustment. As such, it is not entirely arbitrary – it is rooted in some physical understanding of the system under study. The cause of GCM failure to represent accurately large-scale atmospheric circulation is attributed to convection, and the convective adjustment is a way to correct this specific failure.

As a consequence, methodologically, using ad hoc moist convective adjustments to correct GCMs seems legitimate for two main reasons: first, it allows a better agreement between GCM results and the observations, thus making GCMs more useful. Second, it takes into account the effect of a process that cannot be simulated by GCMs: atmospheric convection. The introduction of convective adjustments therefore represents an acknowledgment both of the usefulness of GCMs to study the general circulation of the atmosphere and of their main limitation – not resolving atmospheric convection.

Convective quasi-equilibrium hypothesis

A further justification for the use of convective adjustments in GCMs can be found in A. K. Betts (1986), one of the only study which developed a moist convective scheme since the early attempts of Manabe et al. (1965); Mintz (1968) and Smagorinsky et al. (1965). A. K. Betts (1986) relies on the convective quasi-equilibrium (CQE) hypothesis, originally introduced by A. Betts (1973) for shallow convection and Arakawa and Schubert (1974) for deep convection. According to this hypothesis, as interpreted by A. K. Betts (1986), the collective effect of clouds strongly constrains the temperature and moisture structure in the atmosphere. CQE thermodynamic structures would therefore be pervasive in the atmosphere, both in shallow and deep convective regimes. This hypothesis has been partially verified by observations, which have shown that over large domains, and sufficiently long time scales, characteristic temperature and humidity structures were observed (Lord, 1982; Lord & Arakawa, 1980).

If entirely verified, this hypothesis would be a strong support for the claim that it is not necessary to represent atmospheric convection in GCMs. Only its effect could be taken into account. If cloud fields do on average always the same thing, unstable atmospheric profiles pro-

duced by GCMs could be adjusted towards the typical equilibrium states observed in nature, without needing to enter the complexity of the convective phenomena that lead to these equilibrium states. Just as, when we study a gas, the laws of thermodynamics are enough to study states of equilibrium without taking into account the individual movement of gas molecules.

The CQE motivated A. K. Betts (1986) to propose an adjustment scheme based on equilibrium structures found in observations. In this scheme, the atmospheric temperature and moisture structures given by the GCM at each grid point are adjusted with a characteristic time scale towards two distinct reference thermodynamic structures, one for shallow convection and the other for deep convection. The altitude of cloud top is computed by the model and helps to distinguish between shallow and deep convection. The reference profiles in each case is determined from observations.

To this day, however, A. K. Betts (1986) represented an isolated attempt. Although the usefulness and simplicity of this scheme has often been recognized (e.g. Arakawa & Jung, 2011), parameterizations based on certain theoretical principles were preferred to this observation-based parameterization. Convective adjustment schemes were judged unphysical, as “there is little physical basis for the universality of the water vapor profile” (K. A. Emanuel, 1991, p. 2331). The fear was that the reference profiles chosen by A. K. Betts (1986), although representative of the tropical atmosphere over oceans, could not be easily extended to the whole globe: in fact, numerous studies have shown some departure from quasi-equilibrium over lands as well as in mid-latitudes (e.g. Bechtold et al., 2014; Zimmer, Craig, Keil, & Wernli, 2011). These studies suggested that atmospheric convection is too diverse to be represented by a few empirically determined reference profiles. In order to make parameterizations more generalizable, and also easier to interpret and justify, climate modelers tried to base parameterizations on a certain theoretical framework rather than on empirical profiles. It led to the development of the mass-flux representations of atmospheric convection.

3.3 In pursuit of a theoretical foundation for mass-flux convection parameterizations

In this section, we present what has become the standard approach to represent atmospheric convection in GCMs: mass-flux convection parameterizations. Different variations of mass-flux convection parameterizations have been proposed and used in GCMs. Most of them rely on a similar formalism, that we expose in 3.3.1. We then focus on three interpretations that have been given to this formalism. Contrary to moist convective adjustments, which were based on observations, these interpretations are theory-based: each of them gives a self consistent framework, derived by analogy with a well-established theory, to interpret the mass-flux formalism.

3.3.1 The mass-flux formalism

The replacement of adjustment schemes by mass-flux parameterizations roots in the hot-tower hypothesis proposed by Riehl and Malkus (1958) in their study of the tropical heat budget. As Riehl and Malkus showed, it is not possible to understand the heat budget associated with the Hadley circulation, unless we assume that the apparent large-scale ascent near the equator is actually the sum of vigorous ascent in a few isolated “hot-towers” and descent in their environment. In this hypothesis, the hot-towers cover only a small fraction of the tropics and mix

with their environment only at their base and top. This allows them to transport heat directly from the surface to the top of the troposphere.

The hot tower hypothesis considerably influenced research in the early 1960s on tropical cyclone. A major problem of cyclone research at that time was indeed to represent the role of cumulus clouds in cyclone formation. In this respect, the early phases of cumulus parameterizations in GCMs are intrinsically linked with the development of the first numerical models to simulate cyclone development. Numerical models used to represent the atmospheric general circulation and those used to understand the formation of tropical cyclones were faced with the same conceptual issue of accurately representing moist convection. Influenced by Riehl and Malkus (1958) and subsequent work, Ooyama (1964) formulated a dynamical model for the study of tropical cyclone development in which the heating effects of clouds was represented in a mass-flux form. A few years later, Arakawa (1969) adapted the mass-flux approach to GCMs.

Here, we sketch briefly this approach, which will then become the standard way of representing atmospheric convection in parameterizations. The mass-flux form assumes that atmospheric convection happens in “plumes” covering a small part of the grid box, consistently with the hot-tower hypothesis of Riehl and Malkus (1958). The collective effect of these plumes is represented through the use of a single bulk plume. Each variable can then be decomposed as:

$$\bar{\phi} = \alpha\phi_u + (1 - \alpha)\phi_e \quad (3.1)$$

where α – the fractional area of the bulk plume – is much smaller than 1, $\bar{\phi}$ is the mean value of the variable in the grid box, and ϕ_u and ϕ_e refers respectively to its value in the bulk plume and in the environment.

The vertical transport in the plume is represented by a mass flux defined as $f = \rho\alpha w_u$, where w_u – the vertical velocity in the plume – is assumed to be homogeneous horizontally. Similarly, the vertical transport in the environment – the compensating subsidence – is assumed to have a horizontally homogeneous velocity w_e and a corresponding mass flux $f_e = \rho(1 - \alpha)w_e = -f$. In the first parameterizations such as the one proposed by Ooyama (1964), the convective and environmental mass fluxes (and thus w_u and w_e) are assumed constant from the base of the plume to its top. Therefore, the air is entrained into the cloud base mass flux only at the base of the plume and detrained only at its top. In this simple scheme, the number of unknowns is decreased to only one, the cloud base mass flux f , which has to be specified depending on the large-scale state. The problem of specifying the convective mass flux – and therefore the intensity of convection – as a function of the large-scale state of the atmosphere is called “closure”. The closure usually relies on some equilibrium assumption and is a core aspect of any convection parameterization.

In subsequent, more complex versions of mass-flux parameterizations, f is not necessarily constant vertically: in particular, when $\partial f/\partial z > 0$, environmental air entrains into convection, and when $\partial f/\partial z < 0$, convective air detrains out in the environment. More generally, the vertical variation of the mass flux f is defined in terms of entrainment rate e and detrainment rate d as:

$$\frac{\partial f}{\partial z} = e - d \quad (3.2)$$

With these definitions, the vertical transport of any conserved quantity in the updraft is given

by:

$$\frac{\partial f \phi_u}{\partial z} = e \phi_e - d \phi_u \quad (3.3)$$

As $\alpha \ll 1$, ϕ_e is usually approximated as $\bar{\phi}$. The convective tendencies, that is the temporal evolution of large-scale variables due to convection, can finally be approximated as:

$$\left(\frac{\partial \bar{\phi}}{\partial t} \right)_{conv} = -\frac{1}{\rho} \frac{\partial}{\partial z} f(\phi_u - \bar{\phi}) \quad (3.4)$$

This general framework is shared by many convection parameterizations. Differences between schemes appear in the specification of the vertical velocity w_u (and therefore, the strength of convection), the fractional area α , and the entrainment and detrainment rates, as a function of grid-box mean quantities. Despite these differences, all mass-flux cumulus parameterizations produce a negative feedback on large-scale destabilization (Arakawa, 2004). They thus have initially the same function as convective adjustment schemes, but were judged more satisfying as they explain it *mechanistically*. As we have seen in 2.2.3, a mechanistic explanation interprets the behavior of a certain system in terms of the interactions between its components. This is precisely what mass-flux parameterizations do. They give a mechanistic understanding of atmospheric convection by splitting it into two objects, the plume and its environment. On the contrary, moist convective adjustments impose the effects of convection rather than deducing them, and express at best some kind of *functional understanding* of atmospheric convection, i.e. some understanding of what convection does in the atmosphere. As a result, the hot-tower hypothesis was preferred to convective adjustment schemes to construct parameterizations not because it led to better results, but because it provided at the same time a mechanism to describe atmospheric convection.

We present in the following three interpretations of the mass-flux formalism, each based on an analogy with a certain theory. These analogies are not justifications. They motivate parameterizations given some theoretical grounds, but these grounds are too weak to justify them rigorously. Nevertheless, each theoretical analogy provides a coherent framework to define how the mass-flux formalism is used. As we are going to see, each framework specifies in a strict way the form parameterizations can take.

3.3.2 Quasi-equilibrium filtering – Arakawa and Schubert (1974) parameterization

The Arakawa (1969) parameterization, which was later expanded in Arakawa and Schubert (1974) represents the first attempt to represent atmospheric convection using a mass-flux and the concept of CQE. Schubert (2000) interprets the CQE used in Arakawa (1969); Arakawa and Schubert (1974) as a filtering procedure similar to the quasi-geostrophic theory. This first theoretical analogy is interesting for understanding the basis of one of the earliest mass-flux parameterizations of convection, which is still considered as a reference today.

Sketch of Arakawa (1969) parameterization

In Arakawa (1969), at each grid point, one of three types of convection can occur: what he calls “middle-level convection”, “penetrating convection” and “low-level convection”. For simplicity,

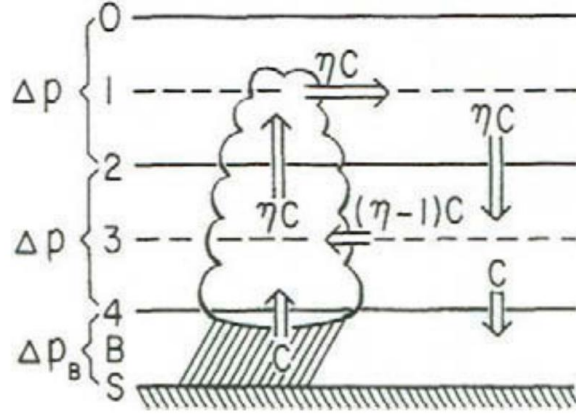


Figure 3.1: Diagram of penetrating convection used in the parameterization of Arakawa (1969). C represents the cloud base mass flux, ηC the cloud mass flux at level 2, and $\eta(1 - C)$ the entrainment.

we will focus here on the problem of penetrating convection and explain how the CQE hypothesis helps to close the corresponding parameterization, that is to determine the convective mass flux C .

As illustrated in Fig. 3.1, Arakawa (1969) is implemented in a three level GCM. Penetrating convection occurs when there is conditional instability between the boundary layer and level 3, but not between level 3 and 1. Mathematically, this condition can be expressed using the moist static energy defined as $h = s + Lq$, with $s = c_p T + gz$ and the saturation moist static energy defined as $h^* = s + Lq^*$. Penetrating convection occurs when $h_B > h_1^*$ and $h_1^* > h_3$. Middle-level convection and low-level convection are defined using similar criteria of stability.

The CQE hypothesis is expressed in the equation governing the tendency of $h_B - h_1^*$ (Schubert, 2000):

$$\begin{aligned} \frac{\partial(h_B - h_1^*)}{\partial t} + v_B \cdot \nabla h_B (1 + \gamma_1) v_1 \cdot \nabla s_1 + \omega_4 \left(\frac{h_B - h_4}{\Delta p_B} \right) - (1 + \gamma_1) \omega_2 \left(\frac{s_2 - s_1}{\Delta p} \right) \\ = -gC \left[\left(\frac{h_B - h_4}{\Delta p_B} \right) + (1 + \gamma_1) \left(\frac{s_1 - s_2}{\Delta p} \right) \eta \right] + \frac{g}{\Delta p_B} \rho_S C_E |v_s| (h_s^* - h_B) \quad (3.5) \end{aligned}$$

In the left side of this equation, the first term is the time evolution of $h_B - h_1^*$, and the other terms express its large-scale horizontal and vertical advection. In the right side, the first term is the tendency due to cumulus convection, and the last term, the effect of surface fluxes.

As cumulus convection decreases $h_B - h_1^*$, and large-scale horizontal and vertical advective terms, as well as surface fluxes, tend to increase it, Arakawa (1969) assumes a kind of equilibrium between the convection on the one hand, and the large-scale advection and surface fluxes on the other. Due to this equilibrium, the core assumption of Arakawa (1969) is to consider that the first term, $\partial(h_B - h_1^*)/\partial t$ is negligible compared to the others. This constraint leads to a diagnostic equation for C .

The CQE has thus a very precise meaning in the parameterization of Arakawa (1969). In the convective adjustment scheme proposed in A. K. Betts (1986), CQE was described as a certain

balance between convection and the large-scale circulation that led to typical atmospheric profiles. Arakawa (1969) relies on stronger assumptions. First, convection is assumed to be governed entirely by the large scale. Second, the characteristic time of convection adjustment, i.e. the time required for C to become constant, is assumed to be negligible compared to the characteristic time of large-scale evolution.

Comparison with quasi-geostrophic theory

Schubert (2000) points out the parallel between the CQE used in Arakawa (1969); Arakawa and Schubert (1974) and the quasi-geostrophic theory. This parallel, which we present in the following, is a first way to frame the parameterization problem by analogy with an existing theory.

The concepts of geostrophic balance and quasi-geostrophic balance have been introduced more than half a century ago and widely applied in atmospheric and oceanic studies, such as weather prediction and ocean circulation research. These concepts help to simplify in a standard way the so-called primitive equations given in Holton and Hakim (2012) as:

$$\frac{Du}{Dt} - fv = -\frac{1}{\rho} \frac{\partial p}{\partial x} \quad (3.6)$$

$$\frac{Dv}{Dt} + fu = -\frac{1}{\rho} \frac{\partial p}{\partial y} \quad (3.7)$$

with $f = 2\Omega \sin\theta$ and g the gravity acceleration. These two equations express Newton's second law in rotating coordinates projected on the horizontal plane.

Jule Charney, in his famous 1948 paper "On the scale of atmospheric motions" (J. Charney, 1948) distinguishes different scales to simplify these equations, in particular a typical horizontal length scale $L \approx 10^6 \text{ m}$, a typical horizontal velocity $U \approx 10 \text{ m s}^{-1}$ and angular velocity $f \approx 10^{-4} \text{ s}^{-1}$ for mid-latitudes. With these notations, the first two terms in equations 3.6 and 3.7 are on the order of U^2/L and fU respectively. Their ratio $Ro = U/fL$, called the Rossby number, is about 10^{-2} in the mid-latitudes, thus negligible compared to 1. It means that the advection terms Du/Dt and Dv/Dt are relatively small compared with the Coriolis force fu and fv in the mid-latitudes. The geostrophic balance consists in neglecting these two terms entirely in equations 3.6 and 3.7, so we have:

$$v = \frac{1}{f\rho} p_x \quad (3.8)$$

$$u = -\frac{1}{f\rho} p_y \quad (3.9)$$

This balance has many applications and explains for example the fact that in the mid-latitudes, the direction of flow is often aligned with the isobars – the lines of constant pressure. However, for many applications, the geostrophic balance is too idealized as it neglects entirely the time evolution of the fluid. Equations 3.8 and 3.9 are indeed those of a static system. The quasi-geostrophic theory introduced by Charney relaxes these assumptions slightly by keeping first-order terms in the equations. It brings back the time evolution and non-linear terms in the equations, while assuming them of higher order.

The geostrophic and quasi-geostrophic approximations can be thought of as a filtering procedure applied to the primitive equations. Indeed, $1/f$ can be interpreted as the time scale of

Earth’s motion and L/U as the time scale of fluid motion we are interested in. When Ro is small, the time scale of fluid motion is therefore much smaller than the timescale of Earth’s rotation. Based on these two time scales, the quasi-geostrophic approximation can be seen as a filtering of high frequencies – that is short timescales – and the atmospheric waves they correspond to. Schubert (2000) summarised the role played by the quasi-geostrophic theory by recalling the musical metaphor used by Charney himself:

About the time he was developing quasi-geostrophic theory, Jule Charney visualized the atmosphere “as a musical instrument on which one can play many tunes”. He thought of the high notes as the sound waves and gravity waves, and the low notes as the Rossby waves, with Mother Nature being “a musician more of the Beethoven than the Chopin type” in that she prefers “the low notes and only occasionally plays arpeggios in the treble and then only with a light hand”. If [the primitive equations] can be thought of as a whole piano, the quasi-geostrophic equations might be thought of as a piano that has been saved in half, with only the low notes remaining usable. Even though its dynamic range is limited, it can still play some beautiful music. (Schubert, 2000, p. 186, 188)

According to Schubert (2000), there is a parallel between the filtering of transient inertia-gravity waves in quasi-geostrophic theory and the filtering of the transient adjustment of a cloud ensemble in CQE. The filtering in CQE is justified by the relatively slow evolution of large-scale forcing compared to the characteristic life-cycle of convection, just as, in quasi-geostrophic balance, the time scale of fluid motion is small compared to the characteristic time scale of the Earth’s motion (when $Ro \ll 1$).

Several criticisms could be made to the CQE hypothesized by Arakawa (1969). In particular, as shown by T. R. Jones and Randall (2011), the hypothesis that the convective time scale is much smaller than the time scale associated with large-scale processes is questionable. It is now well known that convection in nature often develops mesoscale organization (Molinari & Dudek, 1992) and that the various interactions between the convective, mesoscale and large scales are crucial to predict the subsequent evolution of large-scale air masses (Heggen et al., 1998; Ishida, 1989; Vinnichenko, 1970). However, one could argue that the conceptual basis of geostrophic theory has not been better established than CQE: there is no clear separation in the atmosphere between the relatively long time scale of Earth’s rotation (f) and a much shorter time scale associated with fluid motion. It has not prevented the quasi-geostrophic theory to be used in a range of applications, in particular in the first atmospheric numerical weather prediction (NWP) models.

Arakawa and Schubert (1974) recognizes that their CQE hypothesis is not rigorously established in nature. Nevertheless, they assumed that such an assumption would help represent atmospheric convection in GCMs – just as quasi-geostrophic theory helped represent large-scale atmospheric motions in early numerical weather prediction models. This reveals a certain conception of the role of models, and a certain faith in their usefulness despite their imperfections. As R. S. Plant and Yano (2016) argues, “the usefulness of the principle must then be judged a posteriori from its applications, such as the performance of parameterizations” – which echoes the epigraph of this chapter “all models are wrong, but some are useful” (Box, 1976).

3.3.3 Statistical interpretation – Craig and Cohen (2006) stochastic parameterization

Formally, Arakawa (1969) can be seen as a filtering procedure: fast convective motions are filtered in order to guarantee a steady state balance between convection and the large-scale forcing. However, mass-flux convection parameterizations were sometimes interpreted rather as an attempt to represent the mean effect of an ensemble of cumuli. In the literature, parameterizations are indeed often described as a statistical representation of a cumuli field. For instance, K. A. Emanuel (1983) writes:

The foundation of most parameterizations of cumulus convection is the two-scale hypothesis, which may be thought of consisting of two parts. The first is the assumption that the time scale of individual cumuli is small compared to that of the larger-scale circulation we wish to explicitly resolve, so that the cumulus activity may be regarded as being in statistical equilibrium with the larger scale flow. Without this assumption, it is very difficult to treat cumulus activity in a statistical manner. The second part of the two-scale hypothesis is similar: it states that it is possible to define an average over an area which is large enough to encompass a statistically significant number of individual cumuli, but small enough to be regarded as infinitesimal compared to the larger-scale flow we wish to explicitly resolve.

One might think that there is little difference between the CQE used in Arakawa (1969) and the statistical equilibrium discussed here. However, these two equilibrium require different justifications. In Arakawa (1969), as we have seen, the main assumption is that the convective adjustment time, i.e. the time needed for the mass flux C to become constant, is negligible in front of the characteristic time of the large-scale circulation. In the statistical interpretation, what is considered is not the adjustment time of convection, but the time scale of individual cumuli (as well as their dimensions). The fundamental assumption of the statistical interpretation is that there is a sufficient number of cumulus clouds in the model grid at each time step so that their mean properties converge. More precisely, the statistical interpretation assumes that it is possible to choose a grid size Δx much smaller than the typical large-scale L , but much larger than the time scale of individual cumuli l and similarly, that the time step Δt has to be much smaller than the typical timescale of the large-scale T , but much larger than the typical cumulus scale τ . Mathematically, these conditions can be written:

$$l \ll \Delta x \ll L \quad (3.10)$$

$$\tau \ll \Delta t \ll T \quad (3.11)$$

In this second interpretation, an analogy can be made with statistical mechanics. Indeed, as noted by Craig and Cohen (2006), this interpretation of convection parameterizations reminds some typical problems in statistical mechanics, such as the ideal gas model. Statistical mechanics attempt to explain macroscopic physical properties such as temperature, pressure or heat capacity in terms of microphysical parameters that fluctuate around average values and are characterized by a certain probability distribution. A basic application of statistical mechanics to explain, for instance, the properties of an ideal gas requires the following assumptions. First, the gas is assumed to be in equilibrium and there is supposed to be a scale separation between the measurement scale and the molecular scale, so that a large number of particles with well-defined averages are subject to the measurement. Then, the gas particles are supposed to be independent, i.e. their interactions are neglected. Finally, all states of the system consistent with these assumptions are assumed to be equally probable. This last assumption – which might

seem questionable at first sight – is fundamental in the formalism of statistical mechanics and was successful in explaining many natural phenomena. The strength of statistical mechanics is to derive thermodynamic macroscopic formulas from a statistical description of microscopical states under these three assumptions.

Convection parameterizations similarly attempt to derive large-scale tendencies from small-scale processes. Moreover, the first two hypotheses mentioned above – scale separation and particle independence – are analogous to hypotheses generally made in convection parameterizations. In order to consider a kind of equilibrium between the large-scale and the convective scale, the grid-box area is indeed much larger than the surface occupied by a single convective system. Moreover, clouds are assumed to be independent because their interactions are not taken into account. Concerning the last assumption – the equiprobability of all possible states – most convection parameterizations do not make any a priori assumptions on the distribution of convective states in the grid, but consider instead a mean, idealized convective system as representative of all convective phenomena. With this view, conventional convection parameterizations implicitly assume that the number of convective cells in the grid is infinite and that all convective processes are linear, so that only the mean values can be considered.

This hides the quantitative differences that would appear when considering a finite number of convective clouds. As R. S. Plant and Yano (2016) explain, this is a strong assumption. Assuming for instance one convective system every 10 km would yield 100 convective systems within a typical grid cell of 10^4 km^2 . Since the convergence of a number N of random and independent processes towards their mean value scales as $1/\sqrt{N}$, assuming that the mean over a grid cell is representative of convective processes is a very rough approximation, even if some large-scale equilibrium is satisfied. It tends to smooth out the variability of convection, which is problematic as convective fluctuations might interact strongly and in a non-linear way with the resolved large-scale dynamics.

This is why the statistical interpretation of parameterizations naturally leads to the introduction of stochastic components to better account for subgrid-scale variability. R. Plant and Craig (2008) provide an example of such a stochastic scheme. Following Craig and Cohen (2006), they assume that there is indeed a large-scale equilibrium governing the average properties of convection, but that this equilibrium occurs at a scale potentially much larger than the typical GCM grid size. The average of convection over a given grid box will then be a sub-ensemble of the large-scale equilibrium drawn from a distribution determined by the large-scale forcing. This approach does not require a grid size much larger than the typical spacing between convective elements. The grid scale only has to be small enough to resolve the large scale, that is:

$$\Delta x \ll L \tag{3.12}$$

$$\Delta t \ll T \tag{3.13}$$

and in addition there has to be a scale separation between the convective scale and the large scale so that:

$$l \ll L \tag{3.14}$$

$$\tau \ll T \tag{3.15}$$

It is therefore only the conditions between the grid size Δx and the convective scale l on the one hand, and the grid timestep Δt and the convective time τ on the other, that are relaxed –

in particular, we can now have $\Delta x \leq l$ and $\Delta t \leq \tau$ without any contradiction.

Quasi-convective equilibrium and statistical equilibrium are two different ways of formulating the convection parameterization problem. It is not clear which interpretation is more theoretically founded – both interpretations are based on an analogy with a given theory, but this analogy does not prove them completely. However, they give rise to parameterizations of convection with a different formulation and different results – one again, the usefulness of one interpretation or the other must be evaluated according to the results it allows.

3.3.4 Segmentally constant approximation – Yano (2014)

The quasi-equilibrium filtering of Arakawa and Schubert (1974) and the statistical interpretation of Craig and Cohen (2006) both propose an interpretation of the collective behavior of a field of updrafts. Each interpretation relies on an analogy with a given theory, which determines how the closure of the mass-flux formulation is defined. Conversely, with the segmentally constant approximation (SCA), Yano (2014) interprets the mass-flux formalism itself rather than its closure. His objective is not to relate the cloud base mass-flux to some theoretical principles, but to justify why it is legitimate to use a parameterization of atmospheric convection based on the mass-flux formalism described in 3.3.1 – regardless of the formulation chosen for its closure. Despite this difference, we choose to present this third interpretation here because, once again, it is based on an analogy with a certain theory, and gives a coherent framework for formulating convection parameterizations.

The objective of SCA is to systematically derive the mass-flux formalism and its potential extensions from the following set of anelastic equations:

$$\frac{\partial}{\partial t}\phi + \nabla \cdot \mathbf{u}\phi + \frac{1}{\rho} \frac{\partial}{\partial z} \rho w \phi = F \quad (3.16)$$

$$\nabla \cdot \mathbf{u} + \frac{1}{\rho} \frac{\partial}{\partial z} \rho w = 0 \quad (3.17)$$

with ϕ a general prognostic variable, \mathbf{u} the horizontal wind components, w the vertical velocity and F a source term. The first equation expresses the conservation of any general prognostic variable ϕ (for instance momentum or water vapor) and the second equation expresses mass continuity under the anelastic approximation, which filters acoustic waves. For more details about the anelastic approximation and how these equations can be derived, see Gough (1969) and Masmoudi (2007).

The equations 3.16 and 3.17 could be used to run a high-resolution model such as a CRM or a LES within each grid box of the GCM – this is what is done in super-parameterizations. The refinement brought by CRMs or LES is mainly horizontal, as the vertical resolutions used in GCMs and CRM/LES are comparable. Similarly, the SCA can be seen as a horizontal refinement of the grid of the GCM. The main assumption behind SCA is that at each vertical level of a GCM subgrid-scale processes (such as those simulated by CRMs or LES) can be merged into different “segments” with size S_j , such that the grid-box domain with area S can be expressed as $S = \sum_{j=1}^n S_j$. These segments are homogeneous, i.e. the physical variables that describe them are constant horizontally. Vertically, the segments form a “plume” (for example, an updraft, a downdraft or a cold pool) as illustrated in Fig. 3.2. At each vertical level z , we

thus have:

$$\phi = \sum_{j=1}^n \delta_j(x, y) \phi_j \quad (3.18)$$

where $\delta_j(x, y)$ is defined by:

$$\delta_j(x, y) = \begin{cases} 1, & \text{if } (x, y) \in S_j \\ 0, & \text{if } (x, y) \notin S_j \end{cases} \quad (3.19)$$

With these definitions, the average of variable ϕ over the j th segment is defined by:

$$\phi = \frac{1}{S_j} \int_{S_j} \phi dx dy \quad (3.20)$$

and it is then possible to deduce the prognostic equation and the mass continuity equation under SCA from eq. 3.16 and 3.17 for ϕ_j (see Yano, 2014, for more details).

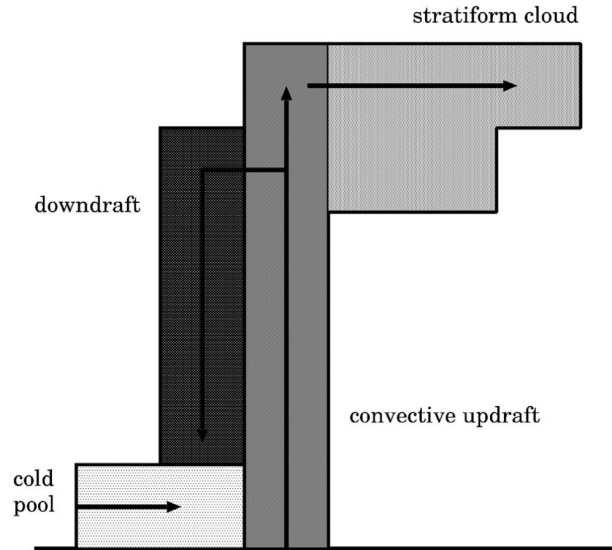


Figure 3.2: A schematic view of a grid box in the SCA. The main hypothesis of SCA is that the different subgrid-scale components (updraft, downdraft, cold pool, stratiform cloud) are constant horizontally. From Yano (2014).

The framework given by Yano (2014) is much more general than the standard mass-flux formulation. As illustrated in Fig. 3.2, it is possible to use it to describe not only updrafts, but also other subgrid-scale components such as downdrafts or cold pools. The fundamental idea of SCA is to represent with homogeneous variables what is sufficiently similar horizontally. Yano (2014) makes an analogy between this approximation and a finite volume approach:

From a purely numerical algorithmic point of view, SCA is nothing other than a finite volume approach [...]. This correspondence becomes more evident when a sufficiently large number of segments are introduced [...]. However, under SCA, the number of elements may be radically reduced in an analogous manner as an image *compression* by wavelet [...]: retaining a high local resolution only where a high

variability is found, but keeping a much lower resolution where less variability is found. As a result, SCA provides a compresses-CRM, and introduction of SCA in place of a standard CRM into GCM leads to compressed super-parameterization (Yano, 2014, p.10)

SCA can be seen as a “selective refinement” of the grid of the model. As such, it can be used to interpret both a CRM – when there is a large number of segments – and the standard mass-flux formulation, which can be recovered using three additional assumptions:

- **entrainment-detrainment hypothesis:** to describe the interactions between different segments, an entrainment-detrainment hypothesis is introduced. Entrainment and detrainment rates $E_{j,i}$ and $D_{j,i}$ describe the lateral exchanges at each vertical level between a pair of segments i and j . This avoids the need to estimate the horizontal winds crossing the segment boundary, and therefore to take into account the position of the segment boundary. Under this simplification, the total entrainment and detrainment rates for the segment j are given respectively by $E_j = \sum_{i=1, i \neq j}^n E_{j,i}$.
- **environment hypothesis:** an environment is introduced and different segments are assumed to interact with the environment exclusively. The total entrainment and detrainment rates for each segment j thus become $E_j = E_{j,e}$ and $D_j = D_{j,e}$ where the subscript e stands for the environment.
- **scale separation hypothesis:** the fractional area σ_j of each convective segment is taken to be much smaller than that of the environment σ_e : $\sigma_j \ll \sigma_e$ or equivalently $\sigma_j \rightarrow 0$ and $\sigma_e \rightarrow 1$. Under this approximation, grid-box means are approximated by their environmental values, i.e. $\bar{\phi} \approx \sigma_e \phi_e \approx \phi_e$ (except for the vertical velocity).

With these assumptions, Yano (2014) explains how the standard mass-flux formulation can be systematically derived from the set of fundamental equations given in 3.16 and 3.17. This method is self-consistent and makes explicit all the successive assumptions that are made. It also provides a guide to extend the standard mass-flux formulation in order to take into account other objects such as downdrafts or cold pools, which are known to play an important role for convection in high-resolution simulations (Brient et al., 2019; Torri et al., 2015). Such a clean derivation of convection parameterizations from fundamental principles seems satisfying. Nevertheless, we disagree with Yano, Bengtsson, Geleyn, and Brozkova (2016) when they argue that any parameterization should be based on a full deduction from basic physical laws similar to SCA:

A parameterization is, by definition, a parametric representation of the full physics on the subgrid-scale [...]. Thus, a certain process of deduction from the full physics is required in order to arrive at a parametric representation. Such a deduction process must be self-consistent and logical: a simple moral dictum. (Yano et al., 2016, p. 426)

For us, the fact that a system is described by certain physical laws does not imply that any parametric representation of this system has to be deduced from the same physical laws. This is not the only way to proceed, and in many cases, not the most efficient one. Let’s take a simple example: the latent heat of vaporization of water, that is the amount of energy that must be added to a given quantity of liquid water to transform it into vapor. The latent heat of vaporization depends, among other things, on the temperature. In theory, we could try to deduce this dependence from first principles, based on quantum dynamics. Yet, to our knowledge,

this approach has not been successful yet. As a consequence, the temperature dependence of heat vaporization of water vapor is based on tabulated values from experiments. This does not prevent the empirical relationship obtained from being useful for various applications.

Similarly, we argue that even if analogies with theories are interesting and provide a basis to construct parameterizations, in many cases, they are not enough. Therefore, we should not rely exclusively on theory and accept to use ad hoc assumptions to make our parameterizations more realistic. The observation-based convection schemes and the three theory-based parameterizations described here should be taken as two extreme stances. In the following, we will see that the parameterizations used in the atmospheric model LMDZ are “object-based” and in between these two antagonist positions. We will argue that objects provide a flexible framework to combine information from observations and theory in convection parameterizations.

3.4 Object-based interpretation of LMDZ convection parameterizations

In this section, we focus on the convection parameterizations used in the atmospheric model LMDZ and defend the “object-based” approach that is used in this GCM – and in many other GCMs, though we limit our study to LMDZ. The construction of parameterizations in LMDZ was done incrementally. Progressively, parameterizations were modified or added in part for theoretical reasons, in part to correct some of the identified deficiencies in model results. Most of these parameterizations were based on an object, as we have defined it in chapter 2. As we will see in the following, the representation of objects in LMDZ parameterizations is based partly on theory, and partly on observations.

3.4.1 Overview of LMDZ convection parameterizations

We first give an overview of the convection parameterizations used in LMDZ, focusing on three central aspects of these parameterization: the representation of shallow convection by the thermal plume model (Hourdin, Couvreux, & Menut, 2002; Rio & Hourdin, 2008), the deep convection parameterization (K. A. Emanuel & Živković-Rothman, 1999; Grandpeix, Phillips, & Tailleux, 2004), and the cold pool parameterization (Grandpeix & Lafore, 2010).

The thermal plume model

The original thermal plume model of Hourdin et al. (2002) combines a diffusion scheme for small-scale turbulence and a mass-flux model to represent dry convection. We focus on the mass-flux part of the model and explain first its conceptual basis as originally formulated in Hourdin et al. (2002). The formulation of Hourdin et al. (2002) of the updraft mass flux is similar to the standard mass-flux formulation exposed in 3.3.1. Hourdin et al. (2002) relies on a simple updraft – so called, a “thermal plume” – homogeneous horizontally and which entrains air along its ascent. The closure of the parameterization, however, is less conventional. As illustrated in Fig. 3.3, the assumption is made that convection happens through stationary convective rolls. From observations and LES, Hourdin et al. (2002) choose a certain value for the typical ratio between the width of a convective cell L and its height z_{max} . Based on this particular geometry, the authors relate the cloud base mass flux to the convergence of air in the subcloud layer, and deduce a certain expression for this mass flux.

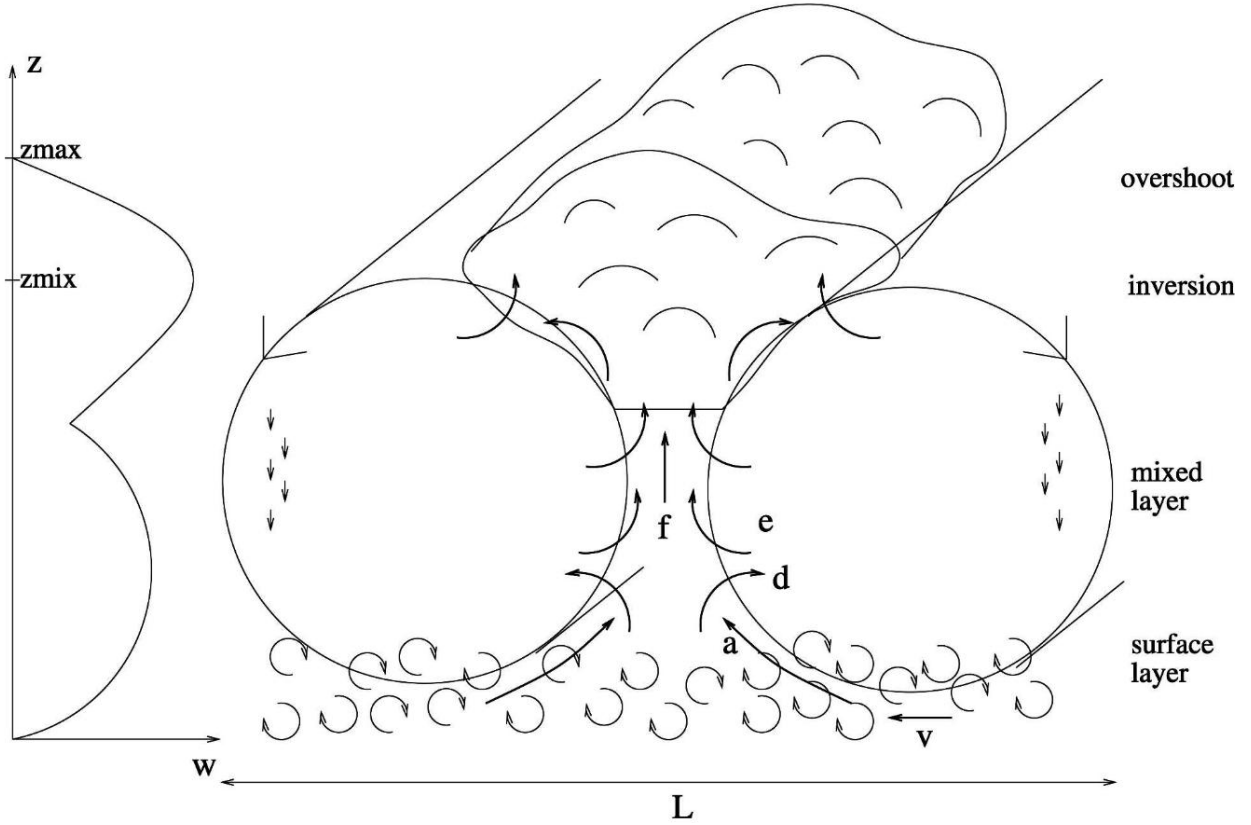


Figure 3.3: Physical image sustaining the thermal plume model and corresponding vertical velocity w , relying on diffusive turbulence in the surface layer and coherent “convective rolls” in the mixed layer. The mass flux f depends on the entrainment of air inside the thermal from the surface layer a , and also on the entrainment e and detrainment d in the mixed layer. From Rio and Hourdin (2008).

There is some similarity between Hourdin et al. (2002) and the standard convection parameterization of Arakawa. However, the arguments used to justify the closure are not of the same order: Arakawa (1969) does not make any assumption on the geometry of convection. The assumption he makes concerns the collective effect of convection, whose adjustment time is assumed to be much faster than the characteristic time of the large scale. Even if Hourdin et al. (2002) is also based on a certain equilibrium assumption between the small scale and the large scale, the justification used to construct the closure is instead based on the geometry of some *objects* – convective rolls – as observed in the atmosphere.

Various developments have been made since the original formulation of the thermal plume model in Hourdin et al. (2002), initially developed to represent only dry convection: in particular Rio and Hourdin (2008) extended the scheme to the representation of cloudy convection, Jam, Hourdin, Rio, and Couvreux (2013) refined the representation of cloud processes by adding a bimodal distribution of subgrid-scale water, and Rochetin, Couvreux, Grandpeix, and Rio (2014) proposed a stochastic triggering of deep convection by thermals by introducing a thermal size distribution. We will not detail these various developments here, but simply note they are also based on geometric pictures of “objects” that can be observed in the atmosphere or in LES.

For instance, the stochastic triggering parameterization of Rochetin et al. (2014) is based

on a LES diagnostic of the geometry of cloudy thermals. The authors show that in LES, two populations of thermals can be distinguished: the smallest thermals, not able to trigger convection, and larger thermals, which can turn into cumulus or cumulonimbus. They calculate that the cross section spectrum of each population can be described using an exponential distribution and use these distributions to calculate the probability of having a cumulus hosting an undiluted updraft strong enough to trigger deep convection. As in Hourdin et al. (2002), the geometry of the objects considered in Rochetin et al. (2014) plays a crucial role in the formulation of the parameterization.

Deep convective scheme

The deep convective scheme in LMDZ is another interesting example of a parameterization that is neither theory-based, nor observation-based, but in between. Its original formulation is described in K. A. Emanuel (1993), and its actual implementation in LMDZ in Grandpeix et al. (2004). This scheme has been developed to correct important limitations of mass-flux convection parameterizations and in particular account for the fact that 1) the entrainment of air in clouds is not continuous as in the standard mass-flux formulation, but episodic and heterogeneous, 2) mixing in clouds crucially depends on microphysical processes, a phenomenon ignored by convection parameterizations so far. The main steps of this scheme are detailed in Fig. 3.4. The adiabatic ascent of a parcel starting from the subcloud layer and ending at its level of neutral buoyancy is the core of this scheme. As the thermal plume used in Hourdin et al. (2002), this parcel transports a certain mass flux. The main difference is that at each level, the parcel is ‘peeled’ according to two processes. First, a fraction of the condensed water is converted to precipitations and used to feed an unsaturated downdraft. Second, part of the parcel is shed and mix with the environmental air to build a set of mixtures, which rest at their level of neutral buoyancy after further removal of precipitation and evaporation of cloudy water.

The actual implementation of this scheme in LMDZ, described in Grandpeix et al. (2004), defines how the mixing is performed according to a certain probability distribution, and how the cloud base mass flux of the parcel is calculated. We will not detail it here. By giving the backbone of the scheme, our idea is just to raise the question: how such a scheme can be interpreted? The three interpretations given above attempt to give a precise theoretical framework to parameterize atmospheric convection. Obviously, the parameterization of K. A. Emanuel (1993) does not fit in any of these frameworks. Indeed, the latter is not primarily motivated by theoretical reasons, but by an observation-based understanding of the role played by different phenomena in atmospheric convection.

As a consequence, the whole scheme proposed by K. A. Emanuel (1993) contains many more ad hoc assumptions than for instance Arakawa and Schubert (1974). Apparently, the latter is thus more theoretically based than the former. But this strength is actually a weakness. Paradoxically, the strict theoretical framework chosen by Arakawa and Schubert (1974) prevents them from taking into account some important processes in the atmosphere, microphysical processes in particular. On the contrary, K. A. Emanuel (1993) is not so much based on theory than on the understanding of an object – here, an air parcel that mixes inhomogeneously with the environment. An object, as we have seen in chapter 2, is a natural way of observing and understanding convective phenomena in the atmosphere. Object-based parameterizations do contain ad hoc hypotheses, but such hypotheses allow them to be easily improved and, if necessary, made more complex to take into account new observational data. In particular, the many ad hoc hypotheses used in K. A. Emanuel (1993) helped this object-based parameterization to

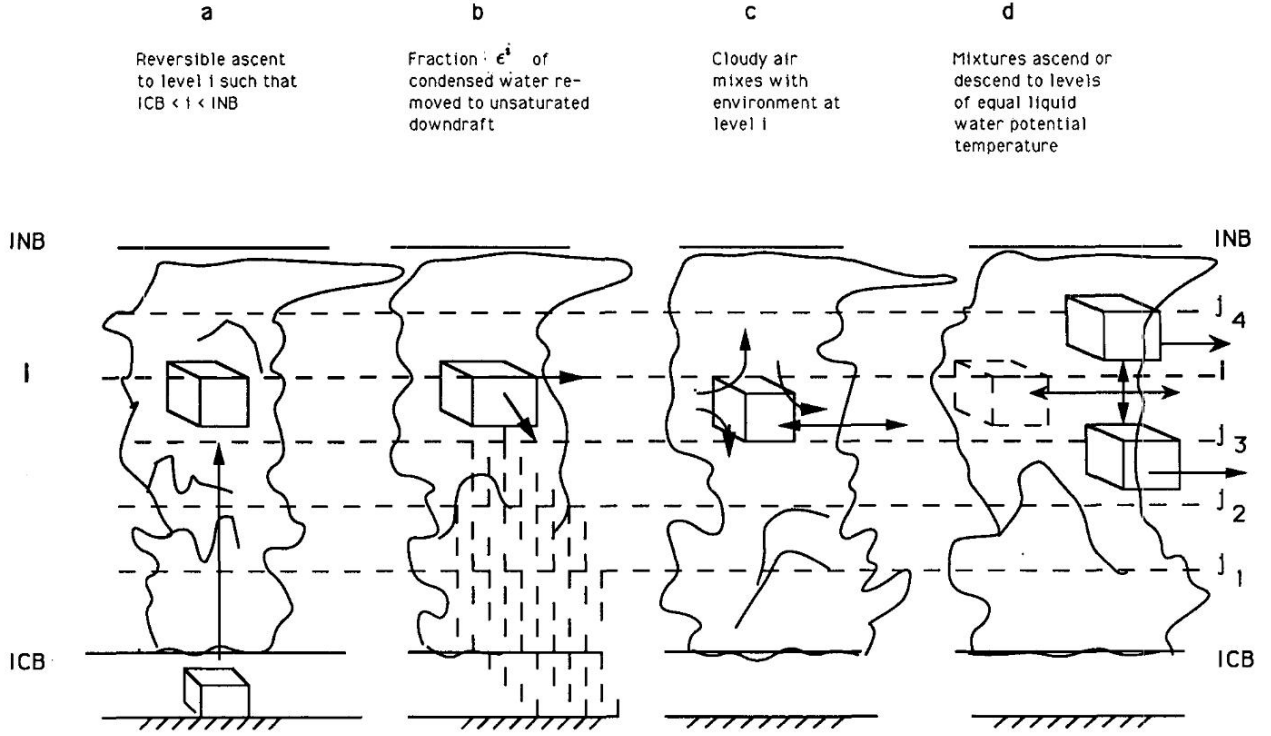


Figure 3.4: Idealized model of the convection parameterization of K. A. Emanuel (1991). a) Reversible ascent from subcloud layer to arbitrary level (i) between cloud base (ICB) and level of neutral buoyancy (INB). b) A fraction ϵ^i of condensed water is converted to precipitation which is added to a single unsaturated downdraft. c) Remaining cloudy air is mixed according to an equal probability distribution with the environment at level i . d) Mixtures then ascend or descend to levels at which their liquid water potential temperature is equal to that of their environment.

be closer to real phenomena, and thus more realistic than theory-based parameterizations such as Arakawa and Schubert (1974), deduced logically from a few justified hypotheses.

Cold pool parameterization

Before further developing our object-based interpretation of LMDZ parameterizations, we present a last parameterization used in LMDZ, which attempt again to represent an object: the cold pool parameterization of Grandpeix and Lafore (2010), already mentioned in Chapter 2. As illustrated in Fig. 3.5, the parameterization consists in circular cold pools with vertical frontiers. The cold pools originate from the downdrafts of the deep convective scheme. Three prognostic variables are used to compute the characteristics of cold pools and their evolution: 1) their fractional coverage σ 2) the potential difference $\delta\theta(p)$ between cold pools and their environment and 3) the specific humidity difference δq_v between the two. As cold pools have vertical frontiers, σ is constant vertically, but $\delta\theta(p)$ and $\delta q_v(p)$ may vary along the pressure axis. Since cold pools are denser than their environment, they spread at the surface. Their spreading rate is calculated following Von Karman et al. (1940) as the square root of their potential energy:

$$EP = -g \int_0^{h_w} \frac{\delta\theta_v}{\theta_v} dz \quad (3.21)$$

where θ_v is the virtual potential temperature and h_w the mean cold pool depth (at which $\delta\theta_v$ vanishes). Heat, mass and energy budgets then allow to calculate σ , $\delta\theta(p)$ and $\delta q_v(p)$ at each vertical level.

An important aspect of the cold pool parameterization is to trigger and feed the deep convective scheme. The feedback of cold pools on the deep convective scheme takes into account the spreading rate of cold pools and is parameterized in the closure of the deep convective scheme. When deep convection is activated, the cloud base mass flux is enhanced in the presence of cold pools.

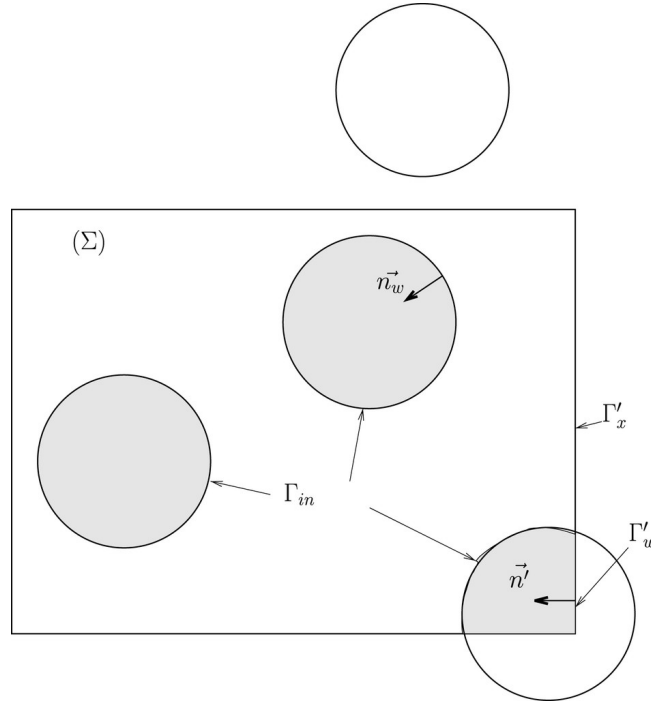


Figure 3.5: Representation of cold pools in a GCM grid cell, as seen from above, in the parameterization of Grandpeix and Lafore (2010).

The representation of cold pools used in Grandpeix and Lafore (2010) is partly based on theory. For instance, the spreading of cold pools as well as the power available to feed convection are determined following simple equations of fluid dynamics and conservation laws. However, it is also partly based on observations and includes ad hoc assumptions motivated by observational data of cold pools – in the original version of the parameterization, the density of cold pools is for instance imposed with a value representative of observations. The parameterization of Grandpeix and Lafore (2010), like the other LMDZ parameterizations exposed previously, is therefore semi-empirical in nature: it is based on a certain theoretical framework, but is not entirely deduced from this theoretical framework. On the contrary, numerous ad hoc assumptions help to ensure that the cold pools represented in this parameterization are sufficiently close to the cold pools studied in observations or in high-resolution simulations.

3.4.2 Interpreting LMDZ multi-object parameterizations

The previous description of LMDZ parameterizations reveals that they are based on some objects: convective rolls for the thermal plume model, an undiluted parcel for the deep convective

scheme, and cold pools for the parameterization of Grandpeix and Lafore (2010). Object-based parameterizations might seem more ad hoc and less solidly founded than theory-based ones. As Edwards (2010) put it, they live in a “semi-empirical world” in between theory and observations. Yet, what matters is not how solidly founded they are, but how useful they are to develop a GCM and improve it. We argue that objects provide a useful and flexible framework to guide the development of parameterizations.

Object-based vs theory-based parameterization

Quasi-equilibrium, statistical equilibrium and SCA attempt to deduce logically the formulation of convection parameterization or its closure from a coherent theory. In each case, an analogy with an existing well-established theory is used to justify the assumptions made: Schubert (2000) uses an analogy between CQE and quasi-geostrophic theory, Craig and Cohen (2006), an analogy with statistical thermodynamics, and Yano (2014) describes its schemes as a particular case of finite volume approach. These analogies should not be confused with demonstrations: although interesting, their explanatory power remains limited. They give a coherent, but incomplete description of phenomena as they occur in nature.

On the contrary, the thermal plume model, the deep convective scheme or the cold pool parameterization used in LMDZ are not so much theory-based as object-based. Their objective is not to apply a certain theory to the parameterization problem, but to describe an object: convective rolls in Hourdin et al. (2002), an air parcel in K. A. Emanuel (1993) and cold pools in Grandpeix and Lafore (2010). The description of each object is neither entirely empirical like the reference profiles used in moist convective adjustment schemes, nor entirely determined from first principles: it is based both on observations and some theoretical considerations. Theoretical equations provide a basis to interpret the observed behavior of the object and construct a parameterization based on some fundamental principles. Ad hoc assumptions are used to guarantee that this theoretical basis is sufficiently close to the real world. Object-based parameterizations are thus semi-empirical: the assumptions they use are motivated in part for theoretical reasons, in part for empirical ones.

Chapter 2 provides arguments for the use of such object-based parameterizations. Indeed, we saw that objects are a natural and convenient way to analyze both observations and high-resolution simulations. When we observe clouds, we do not see a quasi equilibrium, a statistical equilibrium or SCA equations. We see objects interacting with each other. Analyzing these interactions enables first to characterize the objects, and second to understand them. Thinking parameterizations with objects helps to use this understanding both to construct and test parameterizations. For instance, the rate of spreading of cold pools in Grandpeix and Lafore (2010) can directly be compared with the rate of spreading of cold pools in observations or high-resolution simulations. If there is a mismatch between what is predicted in the parameterization and observations, it is possible to try to adjust the specific hypothesis involved. In this way, it is possible to ensure that the parameterized objects are as close as possible to the real world.

Conversely, a theory-based parameterization takes more distance from the observed phenomena. It attempts to deduce a coherent representation of convection from fundamental principles. This logical deduction gives a strict framework for representing atmospheric phenomena. If this framework does not give realistic results, it is difficult to use observations to interpret the failure of a theory-based parameterization and improve it. A new coherent theory

has to be proposed to solve the discrepancy. It is of course impossible to formulate a new theory for each bias identified in the model. As a consequence, even if theory-based parameterizations are more coherent than object-based one, they are likely to be further from observed phenomena.

On the coherence of parameterizations

Unlike theory-based parameterizations that attempt to derive successive assumptions from first principles in a logical and consistent manner, the ad hoc assumptions used in object-based parameterizations can be partially incoherent. For instance, the thermal plume model and the deep convective scheme used in LMDZ express two visions of convection, the first one based on convective rolls and the second on an air parcel. These visions are in part incoherent: when the thermal plume model and the deep convective scheme are used one after the other in LMDZ, we assume successively that convective instabilities are dissipated through convective rolls of Fig. 3.3 and through the undiluted parcel of Fig. 3.4 – two representations of convection that are radically different and in part contradictory. One might prefer to have instead a single unified scheme to express both shallow and deep convection. However, is the lack of coherence an issue when developing parameterizations?

To answer this question, we will first wonder: why does a scientific theory need to be coherent? Nature itself is coherent: physical laws are universal, i.e. they apply in the same way everywhere and at all times. Since the vocation of a scientific theory is to explain nature, a scientific theory should be coherent as well. A theory that would not be coherent would be described as incomplete. For instance, time has a different meaning in quantum physics and general relativity. For over half a century, many theoretical work has been done to attempt to unify general relativity and quantum physics in a single coherent framework.

However, parameterizations are not theories. Their objective is to represent approximately atmospheric convection in GCMs, not to provide an explanation for atmospheric convection from first principles. In addition, coherence does not guarantee accuracy. A parameterization may be perfectly coherent and at the same time produce very unrealistic results. Therefore, for parameterizations, coherence is not a panacea. It is only one criterion, among others, to take into account to evaluate the quality of a parameterization (see chapter 5 for other criteria to assess a parameterization from a conceptual point of view).

The semi-empirical world of object-based parameterizations is thus not a weakness, but a strength. The framework it gives is more flexible than the strict coherent framework imposed by theory-based parameterizations. It helps to recognize the usefulness of certain empirical hypotheses, alongside other assumptions that can be justified on the basis of theoretical principles. By using different objects, in part incoherent, we recognize that the representations used in parameterizations are imperfect, so that a single coherent representation of convection is not sufficient. Each object has a limited explanatory and predictive power. Thinking with objects allows us to represent in parameterizations physical images which are certainly partly contradictory, but also complementary to give a realistic picture of atmospheric convection.

Limits of object-based parameterizations

Despite the advantages of objects to parameterize atmospheric convection, there are some limits inherent to the convection parameterizations that cannot be solved using objects. The two main

issues we identified are the artificial separation of scale and the artificial separation of processes introduced by parameterizations.

Separation of scales Objects used in parameterizations represent subgrid-scale phenomena. Convection parameterizations thus introduce a separation between two scales: the large-scale, resolved by the grid of the model, and the convective scale, which has to be parameterized. Regardless of the parameterization chosen, the large scale is taken to be much larger both in terms of spatial and time scale, so that fast convective processes react quickly to the large-scale state, hence a kind of equilibrium between the two. Parameterizations aims at representing this reaction and the associated feedback of convection on the large-scale state. In this idealized vision, atmospheric processes at other scales either do not exist or interact in a negligible way with the convective scale and the large scale.

In the real world, the large scale and the convective scale are of course further divided into many different scales: different large scale disturbances are best characterized with different spatial and temporal timescales and the same can be said for individual cumuli. A range of scales seems more realistic to describe typical large-scale or convective phenomena than a single value. It does not necessary make the parameterization problem more complicated if we assume a clear separation between the two ranges of scales, i.e. if we assume that the kinetic energy in the atmosphere is concentrated in two distinct windows, the large scale and the convective scale, the latter being much smaller than the former, and that there is a so-called “mesoscale gap” between the two so that the large scale and convective scale interact exclusively with one another.

The problem is that there is no such mesoscale gap in the atmosphere. It is now well known that convection in nature often develops mesoscale organization (Molinari & Dudek, 1992) and that the various interactions between the convective, mesoscale and large scales are crucial to predict the subsequent evolution of large-scale air masses (Heggem et al., 1998; Ishida, 1989; Vinnichenko, 1970). The atmospheric power spectra is more and more seen as a continuum and there is no mesoscale gap that clearly separates the convective and large scale (Yano, 2016). Due to this absence of clear-scale separation, we miss something when thinking only with subgrid-scale objects. It seems difficult to take into account what happens in the mesoscale with an object-based parameterization of atmospheric convection.

Separation of processes Thinking with objects helps to separate the problem of atmospheric convection into different sub-problems. As a consequence, object-based parameterizations typically have a modular structure, with different components describing different physical processes: shallow or deep convection, precipitation, cold pools, radiation, turbulence, etc. In practice, the different modules are generally coupled to the dynamics more or less independently from the others. Most small-scale interactions between processes are not taken into account.

This absence of coupling is a serious limitation of parameterizations because in the atmosphere, the coupling between different processes is the norm rather than the exception. A famous example is given by the very strong and fast interactions among cloudiness, turbulence and radiation in stratocumulus clouds. Turbulence driven by radiative cooling near cloud top in stratiform clouds leads to the entrainment of warm and dry air which might produce holes in the cloud layer, thus affect the fractional cloudiness (Lilly, 1968; Stevens, 2002). Such effects are rarely taken into account in GCMs because turbulence, radiation and microphysics are

generally described in separate modules interacting with the dynamics only.

This limitation is of a more practical than theoretical nature. In theory, the coupling between the different modules of a parameterization can indeed be envisioned. For instance, in LMDZ, the deep convective scheme is coupled to the cold pool scheme of Grandpeix and Lafore (2010): deep convective downdrafts may generate cold pools, which in turn trigger and feed the deep convective scheme. Such attempts are a nice step forward, but a lot of work still has to be done before the various modules of a parameterization act in concert rather than as separate processes.

3.5 Conclusion

In this chapter, we clarified and justified the semi-empirical nature of convection parameterizations used in GCMs. Recalling the context in which convection parameterizations emerged, we first mentioned that parameterizations appeared because of an operational need to make GCMs work: the motivation behind the first parameterizations was thus more technical than scientific. However, in a second time, some theoretical formulations of atmospheric convection were formulated, notably the Arakawa and Schubert (1974) parameterization. Due to the imperfections of these theoretical formulations, hybrid representations, based partly on theoretical principles and partly on observations, have finally been adopted in most GCMs. These hybrid representations have proven to be the most useful for parameterizing atmospheric convection and can be best understood in terms of objects. We relied on chapter 2 to justify why objects can act as a bridge between observations and theory and as such, are a useful way to think the convection parameterization problem. Some inherent limitations of convection parameterizations were finally clarified, namely the fact that parameterizations introduce an artificial separation of scales and of processes in GCMs.

In which sense does the previous discussion inform the parameterization debate, that is whether or not parameterizations should be replaced by approaches based on high-resolution models or machine learning? To a certain extent, GCRMs (or super-parameterizations) bear similarities with what we called theory-based parameterizations – they attempt to deduce a representation of convective phenomena from first principles. In contrast, machine learning algorithms are closer to observation-based parameterizations – they also rely on learning from observations (or from high-resolution models used as observations), with the difference that they potentially use a much larger sample of reference profiles. We explained in this chapter that parameterizations should not be considered inferior to these alternatives just because they are “semi-empirical”, i.e. based on both theory and observations. On the contrary, in some cases, the simultaneous use of theory and observations is a more useful framework to make models consistent with the real world. The comparison between parameterizations and competing approaches must therefore be judged solely on the basis of their applications – and it remains to be proven that there exists an alternative more useful than parameterizations for making climate projections.

Chapter 4

Epistemology of climate model tuning

Contents

| | | |
|------------|---|------------|
| 4.1 | Introduction | 92 |
| 4.2 | Principle and methods of climate model tuning | 93 |
| 4.2.1 | Typology of parameters used in GCMs | 93 |
| 4.2.2 | Need for climate model tuning | 97 |
| 4.2.3 | Tuning strategies | 99 |
| 4.3 | Climate model tuning vs calibration | 100 |
| 4.3.1 | Calibration – definition and characteristics | 101 |
| 4.3.2 | Similarities and differences between climate model tuning and calibration | 103 |
| 4.4 | Some terminology – Compensating errors, structural error and overtuning | 104 |
| 4.4.1 | Non-accuracy of parameterizations and consequences for tuning | 104 |
| 4.4.2 | Compensating errors | 106 |
| 4.4.3 | Structural error vs structural limitations | 106 |
| 4.4.4 | Overtuning | 107 |
| 4.5 | Automatic tuning methods: a shift in tuning paradigm? | 108 |
| 4.5.1 | Limitations of climate model tuning by hand | 108 |
| 4.5.2 | Automatic tuning methods could make climate model tuning more similar to a traditional calibration procedure, | 109 |
| 4.5.3 | facilitate the interpretation of GCM results, | 110 |
| 4.5.4 | and guide the improvement of parameterizations | 110 |
| 4.6 | Conclusion | 111 |

4.1 Introduction

As their name suggests, the behavior of parameterizations depends crucially on parameters, whose values are often poorly constrained in observations. During the development stage of a parameterization, parameter values are set from a mixture of theoretical or observational reasons, as well as modeling results obtained in simplified case studies used during the development of the parameterization, such as single column model (SCM) case studies. When parameterizations are integrated in a full GCM, parameters can also be adjusted in order to improve the quality of simulation results and guarantee in particular that the temperatures predicted by the coupled GCM are sufficiently close to present-day values. This process is generally referred to as “tuning” (Hourdin et al., 2013; Mauritsen et al., 2012; D. A. Randall & Wielicki, 1997).

Most coupled GCMs are tuned before each phase of the Coupled Model Intercomparison Project (CMIP, see chapter 6). As part of CMIP, every 5 or 6 years, the main coupled GCMs used around the world are submitted to standardized simulation protocols. This coordinated effort represents an opportunity to compare the state-of-the-art coupled GCMs under similar conditions, and use them to study anthropogenic climate change and provide statements for the IPCC reports. The results of coupled GCMs depend to a large extent on their tuning process. Yet, until recently, this process was not well documented, and therefore not taken into account in the many analyses based on CMIP results. Contrary to the calibration of scientific instruments, which is typically conducted following a strict protocol (e.g. Manrique et al., 2020), GCM tuning is not standardized at the level of the climate modeling community. Tuning practices vary among modeling centers, and the criteria used for tuning are in part subjective and depend on the personal expertise of climate modelers. Only recently have climate modelers felt the need to systematically document and compare the tuning procedures they used (Hourdin et al., 2017a; Mauritsen & Roeckner, 2020; Mignot et al., 2021).

Conversely, over the past twenty years, the tuning of climate models has attracted the attention of many philosophers of science. One of the main issues they addressed concerned the interpretation of the agreement between GCM results and observations. If GCMs are tuned against observable data, what does it mean when GCM simulations and observations agree? In particular, should this agreement support confidence in future climate projections? For some philosophers, the agreement between GCM results and past observations may well be just a consequence of the tuning process, especially if the same observations are used to tune a model and evaluate its results (W. S. Parker, 2011; Stainforth, Downing, Washington, Lopez, & New, 2007). Others, however, are more optimistic and highlight the various lines of evidence used to test climate models as an argument for their empirical adequacy (Lloyd, 2009, 2010).

Philosophers of science, however, have generally been more interested in the consequences of tuning on the results of the model, than in the tuning process itself. In particular, there is little discussion in the philosophical literature about the variety of strategies used in the tuning process. The parameters used for tuning are rarely specified, nor are the needs that tuning addresses. Moreover, we found no attempt to define what makes climate model tuning unique compared to calibration procedures used in other scientific fields. Such comparison would be valuable in defining the basis, scope, and purpose of climate model tuning and constraining it with what is done in other fields.

We believe that the fact that climate modelers, in the past few years, have decided to better

document their tuning process is an opportunity for philosophers of science. The purpose of this chapter is to initiate a discussion on the specific kind of scientific practice tuning represents. In the following, we are not so much interested in the consequences of tuning on the interpretation of climate models – a subject much discussed by philosophers of science – as in the tuning process itself.

4.2 Principle and methods of climate model tuning

The tuning process is confronted with many constraints, which take various forms. In particular, there are constraints on the parameters that can be used for tuning, constraints on the objectives that tuning must satisfy, and constraints on the strategies that can be used to tune a climate model. Clarifying these different constraints is the purpose of the following subsections.

4.2.1 Typology of parameters used in GCMs

We first look at the parameters that can be used for tuning. Modern coupled GCMs typically contain hundreds of parameters. To understand the role of tuning in GCMs, we first need to understand what these parameters control in climate models. We propose four criteria to distinguish on different levels the role played by parameters in a GCM:

1. What does the parameter represent?
2. Is it observable, i.e. does it have an empirical equivalent?
3. How is it used mathematically?
4. What is the effect of the parameter on GCM results?

These criteria are partly arbitrary and other classifications could be proposed. We have chosen them because they are in continuity with what we have presented elsewhere in this thesis, particularly in Chapters 2 and 3. Regardless of the classification chosen, we argue that it is important not to lump parameters together: there is a wide variety of parameters used in GCMs, and not all parameters are equally important for climate model tuning. We hope that the distinctions we propose in the following can help in understanding what types of parameters can be used for climate model tuning and how they can be used.

1) What does the parameter represent?

In chapter 3, we have seen that most parameterizations are based on some objects. Parameters in such object-based parameterizations can represent three main phenomena: the interaction between an object and its environment, the interaction between two objects, or the characteristics of an object.

The interaction between an object and its environment This first type of representation concerns the majority of parameters used in GCMs. With a few exceptions, parameterizations describe objects evolving independently of each other in the same environment. Most objects used in parameterizations interact with this environment exclusively. Direct interactions between objects are theoretically possible, but difficult to include in parameterizations in practice. We mentioned in section 3.4.2 that this artificial separation of processes was a serious

limitation of convection parameterizations.

As a consequence, how an object interacts with its environment is a core aspect of any object-based parameterization. To specify this interaction, semi-empirical formulas involving parameters are often used. In particular, as we have seen in 3.3.1, mass-flux convection parameterizations specify the exchanges between a bulk plume and its environment through entrainment and detrainment rates. The expression of entrainment and detrainment rates typically involves some parameters. For instance, in the thermal plume model of Rio and Hourdin (2008), the entrainment and detrainment rates in the cloud layer, respectively $e(z)$ and $d(z)$, are proportional to the mass flux inside the cloud $f(z)$:

$$e(z) = \epsilon_* f(z) \quad (4.1)$$

$$d(z) = \delta_* f(z) \quad (4.2)$$

with ϵ_* and δ_* two parameters whose values must be determined empirically¹.

Certain formula parameterizing the effects of cloud and precipitation microphysics also belong to this category. For instance, the following formula given in Sundqvist (1988) is used in LMDZ to calculate the evaporation of precipitation at each vertical level:

$$\frac{\partial P_{l,s}}{\partial z} = \mathbf{EVAP} \left(1 - \frac{q_t}{q_{sat}}\right) \sqrt{P_{l,s}} \quad (4.3)$$

In this formula, the parameter **EVAP** specifies how a given object – here, a precipitation shaft – interacts with its environment. In the standard version of LMDz, the precipitation shaft is described with a liquid or solid precipitation flux $P_{l,s}$.² The environment plays a role in the formula through the mean relative humidity over the grid-box at each vertical level q_t/q_{sat} .

The interactions between two objects Since interactions between objects are limited, parameters representing the interactions between two objects are less frequent than those representing the interactions between an object and its environment. In LMDZ, they concern mostly the activation of the deep convective scheme by the thermal plume model or by the cold pool parameterization (see 3.4.1). For instance, in the cold pool parameterization of Grandpeix and Lafore (2010), the power transferred by cold pools to the deep convective updraft P_{lift} is proportional, for simple dynamical reasons, to the cube of the cold pool spreading speed C_* , the height of cold pools h , the density of the air ρ , and the total perimeter of cold pools per unit area L :

$$P_{lift} = \epsilon_{lift} \rho C_*^3 h L \quad (4.4)$$

The parameter ϵ_{lift} specifies that only part of the power of cold pools can be used for lifting, therefore to enhance the deep convective mass flux. As explained in Grandpeix and Lafore (2010), this parameter accounts for different phenomena that prevent the full conversion of cold pool energy to lifting power for deep convection, and that were ignored in the formulation of the cold pool parameterization.

¹More precisely, the parameters that are imposed in Rio and Hourdin (2008) are δ and $\beta = e/d = \epsilon_*/\delta_*$. The advantage of using β instead of ϵ_* is that β is also used to estimate the entrainment as a function of the detrainment in the *subcloud layer*, where the detrainment is calculated following a more complex expression than 4.2.

²In chapter 5, we propose to change formula 4.3 in order to make the object “precipitation shaft” more realistic by introducing clear and cloudy precipitation fractions.

The characteristics of an object Finally, parameters can be used to specify the characteristics of an object, in particular its geometry. For instance, in the thermal plume model, the aspect ratio of the convective rolls is prescribed, i.e. the ratio between the width L of the convective rolls and their height z_{max} (see Fig. 3.3):

$$r = \frac{L}{z_{max}} \quad (4.5)$$

Another example is given in Grandpeix and Lafore (2010), where the parameter k_* specifies the link between the cold pool spreading speed and the square root of their potential energy EP (as defined in 3.21):

$$C_* = k_* \sqrt{2EP} \quad (4.6)$$

Such parameters precise the properties of the objects used in parameterizations.

2) Is the parameter observable, i.e. does it have an empirical equivalent?

This second criteria is related to the first one. Parameters representing the characteristics of an object are usually the easiest to measure empirically. For instance, the value of the aspect ratio of convective rolls in the thermal plume model $r = L/z_{max}$ was set to 2 in Rio and Hourdin (2008) based on the geometry of convective rolls typically observed in LES. In this case, **the empirical equivalent of the parameter is well defined and can be directly diagnosed in observations or high-resolution simulations**. Nevertheless, we will see in 4.4.1 that even parameters that have a direct equivalent in observation are in fact weakly constrained by this equivalent.

It is often more difficult to calculate the empirical value of a parameter representing an interaction between two objects, or between an object and its environment. Two types of situations can be distinguished:

- **the parameter has, in theory, an equivalent in the real world, but this equivalent is difficult to diagnose in practice.** Formula 4.3 is a good example. In theory, it is possible to calculate the parameter **EVAP** by measuring the evaporation rate of different precipitation fluxes in different environments. However, such an empirical calculation of **EVAP** requires extensive observational data in various environments, and is therefore difficult to achieve in practice.
- **the parameter has no direct equivalent in the real world.** In fact, the empirical equivalent of many parameters is often not well defined. For instance, the cloud and precipitation overlap parameterization introduced in chapter 5 contains a parameter that guarantees a linear decrease of the fractional precipitation area when the precipitation intensity falls below a certain threshold. This threshold was introduced to ensure a linear behavior of the parameterization but has no direct analogue in observations.

3) How is it used mathematically?

GCM parameters can be distinguished on another level: how is the parameter expressed mathematically, i.e. what is the form taken by the equation in which this parameter is written? It is useful to keep this third criterion in mind when tuning a GCM. Here again, a few broad categories can be distinguished:

- **threshold parameters**, used to trigger or strongly enhance a certain parameterized process when a given variable goes above or below their value. To guarantee a smooth behavior, the corresponding equations are often expressed using exponential functions. For instance, the following equation defines the sink associated with the formation of liquid precipitation in LMDZ:

$$\frac{dq_l}{dt} = -\frac{q_l}{\text{CLTAU}}(1 - e^{-(\frac{q_l}{\alpha_C})}) \quad (4.7)$$

This equation depends on the local concentration of cloud liquid droplets averaged over the grid box q_l . When the local concentration of cloud liquid droplets in clouds q_l/α_C goes above the parameter **CLC**, the sink associated with the release of liquid precipitation increases exponentially.

- **statistical parameters**, that define how certain variables are distributed over the grid box. For instance, the water vapor distribution is crucial to determine cloud properties, and the equations defining the shape of this distribution generally depend on some parameters. Typically, in LMDZ, in the presence of shallow convection, the cloud fraction and cloud water content are calculated following a bi-Gaussian distribution of the saturation deficit. The variance of each distribution is related to variables of the thermal plume model through different parameters as specified in Jam et al. (2013).
- **linear parameters**, from which depend more or less linearly a certain parameterized process. These parameters include those that are used to specify the efficiency of a certain process such as ϵ_{lift} in 4.4, a characteristic time constant such as **CLTAU** in 4.7, or the relative importance of two processes.

The mathematical expression of parameters also allows to distinguish them depending on their **unit**: for instance, threshold parameters such as **CLC** have the unit of the variable for which they set a threshold, whereas parameters expressing an efficiency such as the parameter ϵ_{lift} used in 4.4 are dimensionless. The unit of a parameter helps determine its range: for example, a dimensionless parameter expressing an efficiency is usually between 0 and 1, while the value of a threshold parameter must be within the range of values taken by the variable it controls.

4) What is the effect of the parameter on GCM results?

To determine the effect of a parameter on GCM results, thus how useful it will be for tuning, different things have to be taken into account:

- The effect of threshold parameters, statistical parameters and linear parameters will be significantly different on GCM results. In particular, threshold parameters can have a very large effect or on the contrary no effect at all depending on the value of model variables compared to the threshold. On the contrary, changing the value of a linear parameter systematically increases or decreases the intensity of the corresponding parameterized process.
- some parameters are only read in certain cases in the model – for example, only when ice clouds are present – while other parameters always have an influence. As we are going to see in chapter 5, it allows to focus on some parameterizations – parameterizations involved in shallow convection for instance – in part of the tuning process.

- certain parameters control directly the same model results. For instance, in equation 4.7, **CLTAU** controls the speed at which precipitation is formed, whereas in equation 4.3, **EVAP** controls the evaporation rate of precipitation. If we decrease **CLTAU**, more precipitation will form, but if at the same time we increase **EVAP**, more precipitation will evaporate. Therefore, the two effects might cancel each other out in terms of the precipitation rate at the surface.
- Last but not least, the effect of a given parameter on GCM results depends obviously on the importance of the process it represents. In particular, since clouds are crucial for climate, parameters that have a direct effect on clouds are considered very important and generally kept for tuning (Hourdin et al., 2017a).

4.2.2 Need for climate model tuning

The type of parameters that can be used for tuning provides a first fundamental constraint on the tuning process. The parameters of a GCM indeed define the degrees of freedom on which a climate scientist can act to tune his model. One might think that another constraint is provided by the value that model parameters take in observations. However, we will see that GCM parameters are most of the time weakly constrained by their empirical equivalent, even when the latter is well defined. The observations do give some constraints, but on the model results, rather than on the model parameters directly. We will mention in particular that simulating a correct Earth's energy budget is one of the main objectives of climate model tuning.

The values of GCM parameters are weakly constrained by observations

A first basic reason to tune GCM parameters is that the value of many parameters are not known. As we have seen, some parameters have no equivalent in the real world and others have in theory an equivalent, but this equivalent is difficult to calculate in practice. For such parameters, tuning seems to be a necessity.

Does it mean that the observable parameters, i.e. those which have an equivalent that can be easily measured in observations, should be set once and for all before the model is run? The values of such parameters could indeed be determined empirically, by analyzing observations and high-resolution simulations. However, the way parameters are computed in observations and high-resolution simulations must be consistent with their definition in parameterizations. For instance, in mass-flux convection parameterizations, detrained air is assumed to have the properties of the plume, and entrained air, the properties of the environment. On the contrary, in LES, Romps (2010) has shown that it is not the most humid parcels that detrains, neither the most buoyant ones, so that detraining air has properties between the average properties of the plume and those of the environment. Similar results were found for the entrainment. Local estimates of entrainment and detrainment rates at the boundary of a plume then give erroneous flux predictions when used in a parameterization. To measure entrainment and detrainment rates in LES consistent with those used in parameterizations, Couvreux et al. (2010) take instead the definition of entrainment and detrainment rates in mass-flux convection parameterizations, namely (with the same notations as 3.3.1):

$$\frac{\partial \phi_u}{\partial z} = \epsilon(\bar{\phi} - \phi_u) \quad (4.8)$$

with ϵ the fractional entrainment rate defined as $\epsilon = e/f$. This equation can be easily deduced from eq. 3.2 and 3.3 given in 3.3.1. Such bulk calculation underestimates entrainment and detrainment rates by a factor of 2 compared to their direct calculation in Roms (2010). Yet, the diagnosed values by Couvreux et al. (2010) are a much more relevant constraint for convection parameterizations.

Nevertheless, even the definition of Couvreux et al. (2010) could be questioned. Indeed, it implicitly assumes that the fluxes $\bar{\phi}$ and ϕ_u calculated in the LES are equivalent to those parameterized in the GCM, whereas the latter are highly idealized. As a result, the values found in Couvreux et al. (2010) are a relatively weak constraint for the entrainment and detrainment rates used in the thermal plume model: a certain agreement is expected between the two, but this agreement is not supposed to be perfect. More generally, care should be taken in comparing observable parameters to their real-world equivalent. Parameterizations involve simplifications and idealizations: they are inspired by the real world, but do not reproduce it accurately. Due to these approximations, defining the true equivalent of an observable parameter in observations or in high-resolution simulations is virtually impossible. We will come back to the consequences of the non-accuracy of convection parameterizations in 4.4.1.

Simulating a correct Earth’s energy budget

The need to simulate a correct Earth’s energy budget is a second fundamental reason to tune GCM parameters. Earth’s energy budget describes the balance between the energy that the Earth receives from the sun, and the energy it sends back to space. If the Earth’s energy budget is not satisfied in a coupled GCM, the climate simulated by the model will drift towards a climate state quite different from the present climate (Marotzke & Stone, 1995; Murphy, 1995). When a GCM is initialized with present day conditions, it means that there will be an increase or decrease of sea surface temperatures (SSTs) until the model reaches a new equilibrium, warmer or colder than the present climate. This increase or decrease may be stronger than the model simulated transient response to an increase of CO₂ concentration in the atmosphere. If GCMs are to be used to study climate change, they must therefore correctly simulate the Earth’s energy budget.

Historically, when the first coupled GCMs were developed, climate modelers used so-called “flux adjustments” to guarantee that the climate state produced by coupled GCMs was similar to the present climate. Flux adjustments were ad hoc corrections on the heat and moisture fluxes between the atmosphere and the ocean – their objective was to ensure that the atmosphere and ocean components receive the same fluxes as those they need to reach equilibrium when run separately with observed boundary conditions. These flux adjustments have given rise to much debate both within and outside the climate modeling community, even if climate modelers generally preferred to avoid discussing the issue outside the academic circle, so as not to fuel climate-skepticism (Shackley, Risbey, Stone, & Wynne, 1999). For instance, in 1994, a critical comment entitled “Climate Modeling’s Fudge Factor Comes Under Fire” was published in *Science* and starts with the following statement:

In climate modeling, nearly everybody cheats a little. Although models of how the ocean and the atmosphere interact are meant to forecast the greenhouse warming of the next century, when left to their own devices they can’t even get today’s climate right. So researchers have tidied them up by “adjusting” the amount of heat and moisture flowing between a model’s atmosphere and ocean until it yields something

like the present climate. (Kerr, 1994, p. 1528)

A few years later, however, climate modelers managed to get rid of flux adjustments through improved parameterizations and tuning (Kerr, 1997). The strategy used in many coupled GCMs consists in tuning the top-of-atmosphere energy budget in atmospheric simulations forced with SSTs to a value slightly different from observations, in order to guarantee realistic surface temperatures in coupled mode. For instance, in the atmospheric model LMDZ, while the net flux at the top of the atmosphere should in theory be tuned to the observed value of about $0.6 \pm 0.4 \text{ W m}^{-2}$ when imposing present-day SSTs and forcing, it happens that it should be rather tuned to a value of $\approx 2 \text{ W m}^{-2}$ in order to get the present-day global mean surface temperatures in the coupled GCM (Mignot et al., 2021).

Increasing the realism of GCM results

In addition with the global constraint of simulating correctly the Earth’s energy budget, tuning can be used to make some aspects of GCM results more realistic, i.e. to make them more consistent with selected observations. Some modeling groups consider indeed that it is necessary to constrain a few key aspects of climate in order to have coupled GCMs sufficiently realistic to be useful for studying future climate change. As a consequence, they use tuning to guarantee for instance a realistic sea ice volume or extent, a realistic atmospheric circulation, or a realistic increase of temperature in the twentieth century (Hourdin et al., 2017b). Climate model tuning can then be seen as the degree of freedom left to climate modelers to make the results of coupled GCMs with a given set of parameterizations sufficiently realistic to be useful.

Nevertheless, this last use of tuning is more controversial. Some modelers point out that too much tuning might introduce significant compensating errors in GCM results and potentially lead to overconfidence in the interpretation of these results. Neglecting the importance of tuning could then result in an underestimation of the uncertainties surrounding climate projections. On a more practical note, some modelers do not want to push the tuning process too far, because a fine tuning of a coupled GCM requires a significant human and computer investment. As documented in Hourdin et al. (2017a), there is therefore no consensus among climate modelers on how far to go in the tuning process.

4.2.3 Tuning strategies

Having clarified *why* climate model tuning is important for GCMs, the next question is *how* it can be done, i.e., what strategies can be used in the tuning process. Two recent investigations of tuning practices revealed that a wide variety of tuning strategies are used by climate modeling groups (Hourdin et al., 2017a; G. A. Schmidt et al., 2017).

First, it is possible to focus on specific subgrid-scale processes by using a Single Column Model (SCM), that is to say a simplified version of an atmospheric GCM consisting of only one atmospheric column run in a constraint large-scale environment. Historically, data from observational campaigns have been used to build some reference cases for these SCMs. Most of the time, the observations themselves are not exhaustive enough to provide a reference, but they can be used to initialize and force high-resolution models (with a resolution of 1 km or less) with more or less realistic conditions deemed representative of the field data. When these high-resolution models are relatively consistent with the observations, they are in turn taken as a proper reference for evaluating and calibrating SCMs, as illustrated in Chapter 1 (Fig. 1.4).

This protocol was initiated and supported by the GEWEX Cloud System Studies (GCSS) since the mid-1990s (Browning et al., 1993; D. A. Randall, Xu, Somerville, & Iacobellis, 1996). Since a SCM contains the same set of subgrid parameterizations as the full GCMs it represents, it is a useful framework to understand the effect of parameterizations and tune the parameters they involve without digging into the complexity of the interactions between the physics and the dynamics of a climate model.

While very useful, tuning at the process level is not sufficient to ensure the quality of the emergent properties of the full GCM, such as regional patterns of precipitation, sea ice extent or, more importantly, the global balance between incoming and outgoing energy at the top of the atmosphere – the global net top-of-atmosphere flux which is, as we have seen, crucial to prevent global temperatures from drifting too far away from our current climate. To guarantee these emergent properties, the interactions between the physics and the dynamics of a climate model can be taken into account first component by component (atmosphere, ocean, sea ice, etc.) by running for instance atmospheric or oceanic-only GCMs with appropriate boundary conditions.

Stand-alone atmospheric simulations are generally performed under pre-industrial or present-day conditions on relatively short time scales, between 2 and 10 years. Such short simulations have proven to be sufficient to tune parameters to satisfy a variety of targets such as radiative fluxes at top of atmosphere (TOA), shortwave and longwave cloud radiative effects, precipitation, near-surface land temperatures, etc. Note that a distinction is generally made between the metrics that are explicitly used as a tuning target in the tuning process, and those that are only monitored. For instance, the equilibrium climate sensitivity – the long term temperature rise associated with a doubling of CO₂ in the atmosphere – is typically monitored in the tuning process, but rarely taken as an explicit tuning target (Hourdin et al., 2017a). In complement to atmospheric simulations, stand-alone ocean and sea ice simulations can also be performed following the CORE protocole (Coordinated Ocean-ice Reference Experiments, Danabasoglu et al., 2014), targetting once again a variety of metrics: the meridional overturning circulation, SSTs, salinity, etc. (G. A. Schmidt et al., 2017)

Tuning at a component level can be conducted over relatively short time scales and is computationnally efficient. Nevertheless, some global equilibrium – in particular the Earth’s energy budget – also have to be constrained precisely to avoid a drift of coupled models towards a climate state that would be too far from the present climate. Therefore, it is necessary to tune the different components of a climate model when they are coupled together in a full GCM. The duration of a coupled simulation is typically on the order of a decade to over a century. The tuning of a full GCM therefore requires a significant human and computational effort and is a long process – three years for the CMIP6 version of the IPSL GCM (Mignot et al., 2021) – with many back and forth adjustments at a process level, at a component level, and at a full GCM level.

Having described climate model tuning and its main characteristics, we will now see in what way such tuning is different from a traditional calibration procedure.

4.3 Climate model tuning vs calibration

Tuning is not a specificity of climate modeling. If we understand tuning in a broad sense as calibration, it is shared by many different scientific or engineering fields. However, some

climate modelers criticized the use of the term calibration to refer to tuning. In particular, D. A. Randall and Wielicki (1997) argued that:

Modelers sometimes say that they need data to “calibrate” their models. Model calibration is the same as tuning, except that [...] tuning has certain negative connotations, while calibration has positive connotations. We all appreciate that instrument calibration is a good thing. Surely, then, model calibration is also a good thing, unlike, for example, tuning. (D. A. Randall & Wielicki, 1997, p. 403)

According to the authors, the term calibration to refer to climate model tuning is a misnomer. It tends to legitimate climate model tuning by analogy with “instrument calibration”, whereas the two activities are radically different. This led them to make, a bit further, the following strong statement:

Tuning is bad empiricism. Calibration is bad empiricism with a bag over its head. The problem with tuning is that it artificially prevents a model from producing a bad result. (D. A. Randall & Wielicki, 1997, p. 404)

Apparently, D. A. Randall and Wielicki (1997) have a certain picture in mind when referring to instrument calibration. Nevertheless, they do not elaborate on this picture, nor do they justify why it does not fit climate model tuning. This is the objective of the present section. We first try to define and characterize a standard calibration procedure (4.3.1) and then compare it to what is done in climate model tuning (4.3.2).

4.3.1 Calibration – definition and characteristics

Definition

A widely used definition of calibration is given by the epistemologist Allan Franklin (Franklin, 1989, 1990, 1999) as “the use of a surrogate signal to standardize an instrument” (Franklin, 1999, p. 237). Typically, the instrument to calibrate is a material device, ranging from a common instrument such as a clock, a thermometer or a balance, to a highly specialized and sophisticated instrument used in laboratories. The governing principles that make the instrument able to measure what we want to measure are generally **well understood**. For instance, there is no debate about the theoretical ability of a mercury thermometer to measure temperature – we understand why the instrument should work in theory. Moreover, in practice, the instrument works as expected: it is **not defective**. This second point is fundamental to distinguish the calibration of an instrument from its repair.

Then, how is the calibration itself conducted? We can distinguish two steps:

1. the comparison between measurements taken by the instrument and the reference – the “surrogate signal” as Franklin puts it. The discrepancy between what the instrument measures in practice and what it should measure in theory must be evaluated against the reference used.
2. the use of the results of this comparison to correct the measurements of the instrument. These corrections can be made before the measurement, through a material correction on the instrument itself, for example by using a tuning wheel. It is also possible to use the instrument as such, but to apply a mathematical correction to its results to ensure a better agreement with the reference.

Measurement standards

In Franklin’s definition, the surrogate signal is any phenomenon assumed to be sufficiently well-known to play the role of a taken-for-granted reference with respect to some purpose. A blank calibration test can provide a first surrogate signal. The principle of a blank calibration test is to perform a measurement under the same conditions as the final measurement, but without the object to be measured. For example, the blank calibration of a balance consists in evaluating the position of the balance pointer without any load on the balance.

Other surrogate signals can be provided using measurement standards. Calibration tests using a measurement standard consist in measuring quantities of already-known values. The standards used as references should be **reliable**, that is they should have well-characterized properties, known before and independently of the calibration test. In addition, the references used should be **adequate**, i.e. close enough to the measurement we want to make with our instrument – the so-called “end-measurement” (Soler et al., 2013). For instance, if we want to use a high-precision balance to measure a sample of approximately 5 g, measurement standards of similar weight should be used. Indeed, the calibration of an instrument might depend on the range measured with this instrument.

In this regard, a distinction is sometimes made between a full calibration procedure and routine calibration tests. A **full calibration procedure** is designed to assess the response of the instrument under a range of conditions and calculate the associated measurement uncertainty. It includes many tests, and has to be carried out regularly (either at regular time periods, for example every year, or whenever a significant change in the operating conditions of the instrument occurs). It can be heavy, and is thus usually not performed too frequently. It is not specific, in the sense that it does not directly depend on one particular end-measurement. On the contrary, **routine calibration tests** are performed just before a planned sequence of end-measurements. They ensure that the instrument is well calibrated for the measurement at hand. The reference used in these tests should therefore be chosen specifically depending on the end-measurement. For complex instruments, containing many interacting sub-modules, both full and routine calibration procedures may include separate calibration tests for the different modules of the instrument.

Standardization and reproducibility

Standardization is an important characteristic of traditional calibration procedures. In particular, there is most of the time a collective agreement among a community of specialists about the standards, or *etalons*, that can serve as well-known references to calibrate a given instrument. The protocol used for calibration is also generally well documented. At an international level, the International Bureau of Weights and Measure has the objective of ensuring the coherence of measurements, by maintaining a database of standardized calibration practices and references around the world.

One of the objectives of the standardization of calibration is to guarantee the **reproducibility** of any measurement. By reproducibility, we mean that it is possible to reproduce any experiment in a laboratory and obtain the same results with a good degree of agreement. Since the measurements of an instrument depend on its calibration, the calibration used must be standardized to avoid introducing differences in the results of the experiment. Reproducibility is a necessary condition to compare measurements conducted under different conditions.

4.3.2 Similarities and differences between climate model tuning and calibration

At first sight, climate model tuning seems to share many characteristics with the picture of calibration just given. Indeed, when a climate model is tuned, it is first compared to well-chosen references, and then corrected to meet certain criteria defined depending on the intended use of the model. The correction is done on the model itself, through a modification of some of its parameters. The global net top-of-atmosphere flux mentioned previously, which is considered as an overall tuning objective by most modeling centers (Hourdin et al., 2017b), can be seen as a blank test for climate models: it consists in setting the pointer of the energy balance of climate models to zero under current conditions, before using them under slightly different conditions, for example to investigate the effect of an increase in the concentration of CO₂ in the atmosphere. As in a calibration procedure, measurement standards can also be used to complement this blank test. For instance, regional surface temperature of precipitation patterns given by observations are often used in the tuning process (Hourdin et al., 2017b). Since a GCM contains many interacting sub-modules, part of the tuning process might be more precise and target specifically one or several sub-modules. We have already mentioned SCMs, which can be used to tune one or several parameterizations in an idealized framework, using high-resolution simulations as references. In complement with SCMs, so-called “process-oriented metrics” are sometimes used to relate large-scale biases to the misrepresentation of specific subgrid-scale processes. An example of process-oriented metrics is given by Suzuki, Golaz, and Stephens (2013), who evaluate the tuning of cloud microphysical properties of a climate model using satellite observations.

Despite these similarities, climate model tuning is much more complex than the calibration procedures described previously. A first source of complexity comes from the references used. As in any calibration procedure, both the precision and the adequacy of the references used for climate model tuning are crucial. Some observations used as references, SSTs for instance, are relatively well known, while others, for example observations of precipitation, suffer from large uncertainties (Stephens et al., 2012). High-resolution simulations used as references in SCM tuning also exhibit biases of various kinds (Stevens & Lenschow, 2001; Vial, Bony, Stevens, & Vogel, 2017). The adequacy of the references raises even more questions. Climate models are mainly used to simulate future climate, potentially slightly different from the current one. By definition, there are no available observations of (hypothetical) future climate that could be used as references in the tuning process. The only climate change observations available are those of the twentieth century warming. Interestingly, there is however a debate in the climate modeling community on whether it is legitimate or not to use the twentieth-century warming as a tuning target (Hourdin et al., 2017a; G. A. Schmidt et al., 2017). From a calibration perspective, temperature records from the past century are one of the most adequate standards available – a transient climate is indeed more relevant for the study of future climate change than an equilibrium climate in preindustrial or present-day conditions. Nevertheless, many climate scientists consider that the twentieth-century warming should rather be used a posteriori to assess the quality of model results. There is indeed a consensus that different metrics should be used to tune a model and to assess the quality of its results. A choice therefore has to be made between the metrics used for tuning and those kept for the evaluation, and this choice varies from one modeling center to another (G. A. Schmidt et al., 2017).

The second source of complexity comes from climate models themselves. Climate models all

have structural limitations: even finely tuned climate models fail to reproduce some important characteristics of climate, for instance the propagation of Madden-Julian Oscillation (Ahn et al., 2020; Jiang et al., 2015; Lin et al., 2006), or the structure of the Intertropical Convergence Zone (Lin, 2007; Oueslati & Bellon, 2015). Because of these limitations, and because the references used are neither very precise nor adequate, tuning does not aim to guarantee a perfect match between climate model results and references used in the tuning process. This is another important difference with traditional calibration procedures. Since GCMs suffer from biases of various kinds, and since the references used for tuning are not perfect, a disagreement sometimes important between model results and references can be tolerated.

In practice, climate model tuning is often comparable to the search for an optimum within the parameter space. This search is complicated because of the large number of parameters and the various interacting sub-modules (parameterizations) contained in a climate model. In principle, the modularity of an instrument makes its calibration simpler. Indeed, the calibration of a complex modular instrument can generally be reduced to a series of calibrations of its sub-modules. This is not the case in climate model tuning because of what Lenhard and Winsberg (2010) calls the “fuzzy modularity” of climate models: parameterizations are too interdependent to be tuned separately. The various interactions between parameterizations have to be taken into account in the tuning process and parameters cannot be adjusted one by one: the full space of parameters has to be considered. Due to the number of parameters involved in climate models and to the cost of numerical simulations, it is however impossible to explore the full space of parameters. The search for an optimum is thus path dependent and there is a risk of being trapped in a local optimum and miss the global optimum.

These difficulties help to understand why, contrary to the calibration of laboratory instruments, climate model tuning has not been standardized yet. There is no consensus on the best approach to tune a climate model, and the tuning strategies vary greatly from one modeling center to another. These tuning strategies are moreover rarely well documented. In fact, despite the importance of tuning, few climate modelers explicitly discussed the tuning of their GCMs in the scientific literature (among the few exceptions, see Golaz, Horowitz, & Levy, 2013; Hourdin et al., 2013; G. A. Schmidt et al., 2014). This can be understood by the fact that tuning has long been considered as the dark side of climate modeling (Hourdin et al., 2017a): a necessary evil for GCMs to “work”, but without any solid scientific basis. Climate model tuning is sometime considered as a way to introduce some *error compensations* in the model, and thus hide some of its failures. There is a risk to *overtune* the model, that is to make it work for bad reasons. These terms are however rarely well defined and can be confusing. We will come back to this ambiguous terminology in the next section and try to clarify it.

4.4 Some terminology – Compensating errors, structural error and overtuning

4.4.1 Non-accuracy of parameterizations and consequences for tuning

Climate modelers and philosophers of science often discussed the risk of going “too far” in the tuning process. D. A. Randall and Wielicki (1997) for instance argued that tuning might “artificially prevents a model from producing a bad result”. Though they recognized that in some

situations, tuning is a necessary evil, they argue that it should be limited to processes that are both important for model results and poorly understood. As a corollary, parameters that have a direct empirical equivalent and that can be measured should be set once and for all before the model is run. W. S. Parker (2011) nuanced this view because she argued that it applies only to models which are thought “to provide a very accurate representation of (relevant) aspects of a target system, with parameters that have clear physical correlates”, which is not the case for climate modeling. On the contrary, according to her, parameter values that give the best model performance might be noticeably different from measured values, because of significant errors elsewhere in the model. Golaz et al. (2013) gave an example of such discrepancy for the volume-mean radius at which cloud droplets start to precipitate, a critical parameter in the atmospheric GCM of the Geophysical Fluid Dynamics Laboratory (GFDL). They found that simulations of twentieth-century warming were more realistic using a value of this parameter smaller than observed (Pawłowska & Brenguier, 2003; Suzuki, Stephens, & Lebsock, 2013).

The choice of some parameter values different from those found in observations can be justified because of the non-accuracy of representations used in climate models. When developing a new parameterization, climate modelers certainly have a physical image in mind: updrafts transporting heat and moisture from the surface to the cloud layer, cold pools spreading at the surface, etc. These physical images are inspirations to develop parameterizations. They provide a conceptual framework to write equations to represent relationships between the different variables in the model. However, it is important to make a distinction in a parameterization between what we are trying to reproduce and what we are trying to represent. The goal of subgrid-scale parameterizations is to *reproduce* some *target variables* at the scale of the grid – temperature and humidity tendencies, average cloud fraction, surface precipitation, etc. – thanks to a *representation* of subgrid-scale processes. As representations, intermediate variables and parameters are not expected to match perfectly the real world.

Let’s illustrate what it means with a simple example. In observations, let’s imagine that we have an ellipsoidal object (for instance a cold pool, see 2.5). We are trying to predict, thanks to a simple model, the mean area of this object and see if it matches the one found in observations. In addition with this target variable, we can diagnose other quantities in observations, for instance the radius of the object, calculated as the mean distance from its edges to its center. In the model, for simplicity, we assume that our object is circular. The radius is one parameter of the model, and we calculate the area of the object as the square of its radius. If we want to reproduce correctly the mean area, we have to represent our object with a radius slightly different from the one diagnosed in observations. Just because in the real, non-idealized world, the area of an ellipse is not exactly the square of its radius as we have defined it. This simple example shows that parameters (and more generally subgrid-scale parameterizations) do not have to match observations perfectly, even if they are inspired by observations, because there is always some differences between what they mean in the model and what they mean in the real world.

There is another fundamental reason why it is not possible to constrain the parameters of a GCM from observations, even when they have an empirical equivalent. Parameterizations attempt to calculate the mean effect of the phenomena they represent. However, atmospheric phenomena are very variable: in nature, there is for instance a multiplicity of updrafts with different properties, higher or lower vertical speeds, larger or smaller surface fractions, etc. Since fluids interact in a non-linear way, the gross arithmetic mean may not be representative

of the collective effect of convective phenomena on the model state variables – what we want to reproduce with convection parameterizations. An effective mean must there be used, which may be significantly different from the gross arithmetic mean. But it is often unclear how to calculate such an effective mean, so parameters are usually weakly constrained by observations.

4.4.2 Compensating errors

The term “compensating errors” is often used to express the fact that parameterizations or more generally climate models can give good results for wrong reasons. For instance, in discussing the different reasons why climate model results can be considered robust or not, Winsberg (2018b) warned that to have confidence in climate models, climate modelers “would have to systematically rule out, as possibilities, the various ways in which the models that are getting that right are doing so as a result of compensating errors” (Winsberg, 2018a, p. 197). This idea of “compensating errors” also appeared in Hourdin et al. (2017a), which explained that tuning a climate model may compensate certain deficiencies in the formulation of the model itself. The previous discussion shows that compensating errors may exist in a climate model, but with a precise meaning. We argue that the term “compensating errors” is a misnomer when we use it to refer to the intermediate variables used in parameterizations. For instance, in 4.2.2, we explained that the small value of entrainment and detrainment rates in mass-flux convection parameterizations compared to the ones calculated locally by Roms (2010) in LES is not an error: it simply means that entrainment and detrainment rates in the parameterization have a different meaning than in Roms (2010). Similarly, in our previous example, the fact that we chose a slightly smaller radius than the one observed would not make our model wrong. Parameterizations may contain bugs and inconsistencies, but the word error is not adequate to speak about the representations used in parameterizations.

It does not mean that a climate model as a whole is not testable from observations or high-resolution simulations. We might find errors – and even “compensating errors” – in climate model results. For instance, a climate model can get the correct radiative effect of low clouds by underestimating the low cloud cover and overestimating their radiative effects – a behavior that has been noted in many models and referred to as the “too few, too bright” cloud problem (Nam, Bony, Dufresne, & Chepfer, 2012; Webb, Senior, Bony, & Morcrette, 2001). Here, we may speak of “compensating errors”, because these errors concern target variables that we try to reproduce, rather than represent. In our previous example, the area of the ellipsoidal object was the target variable of our model. If it were inconsistent with observations, there would be indeed an error in the model. Similarly, the target variables of a climate model can be proven to be correct or erroneous, but not the intermediate variables used in the parameterizations.

4.4.3 Structural error vs structural limitations

The non-accuracy of parameterizations legitimises the tuning process as it is difficult to know in advance which parameter values will give the best agreement between model key outputs and observations – even parameters that have an equivalent in the observations may (and sometimes should) diverge from their observed value. When there is a mismatch between certain model output variables and observations, it is first necessary to tune the model in order to reduce as much as possible this mismatch. Some errors are nevertheless resistant to the tuning process. In this case, some use the term “structural error” of a climate model – sometimes also called “structural uncertainty or model divergence” – to refer to the error associated with

the structure of the model itself, regardless of the choice of parameters (Hourdin, Williamson, et al., 2020; Knutti, 2010; Lenhard & Winsberg, 2010). With what we have seen previously, parameterizations, as idealizations, are neither right nor wrong. The term “structural error” is therefore misleading.

There is in fact, no real errors in the model formulation, only things a model can or cannot do. The term “structural limitations” seems therefore more appropriate than “structural error”. Speaking about the structural limitations of a model actually raises the question: what worlds can this model explore when we go through its whole parameter space? The structural limitations of GCM can be evaluated by testing it with different sets of parameters: they correspond to the results the GCM never obtains regardless of the parameter set chosen.

Furthermore, we propose a distinction between individual and collective structural limitations. Individual limitations happen when it is not possible to predict accurately a target variable, for instance the cloud fraction in the tropics, whatever the accuracy desired on the other variables. Collective limitations refers to the impossibility of predicting simultaneously with a sufficient degree of accuracy two or more output variables, for instance the cloud cover in the tropics *and* the ground precipitation in this region. We argue that it is much more appropriate to speak of the individual and collective structural limitations of a GCM than of an ill-defined structural error. To clarify discussions about the tuning of GCM, we therefore recommend using these terms.

4.4.4 Overtuning

The concept of structural limitations helps to better define what is meant by *overtuning*, a term generally employed in climate science when the parameters of a climate model are adjusted to fit one data set too closely and fail to fit other data reliably (Hourdin et al., 2017a; D. Williamson, Blaker, Hampton, & Salter, 2015). Even if overtuning was originally a statistical concept, its meaning in climate science is very different from its meaning in statistics. In statistics, an over-tuned (or overfitted) model is one that includes more terms or parameters than necessary, or uses more complicated approaches than necessary (e.g. Hawkins, 2004). An extreme example is a model which would contain as many parameters as the number of observations. Such a model could predict the training data simply by memorizing the data entirely, but would fail when confronted with new data. On the contrary, the term underfitting is used when a statistical model cannot adequately capture the underlying structure of the data (Everitt & Skron dal, 2002). It generally occurs when certain terms or parameters relevant to characterizing the data are missing. For example, a linear model applied to non-linear data would underfit these data. In statistics, the concepts of underfitting and overfitting therefore concerns the structure of the underlying model - mainly its shape and the number of parameters it contains. By contrast, in climate modeling, overtuning concern the values of model parameters, rather than the model structure itself.

The definition of overtuning for climate models suggests a causal link between the success of a climate model with respect to a certain reference and its failure when compared to other sources of data. It would be because a climate model works very well in one situation that it would be broken in other cases. On the contrary, we argue that if we have only one reference at our disposal, and if this reference has no systematic bias, it is legitimate to look for the best possible match between our model and the reference. In the absence of other information,

the optimum found in the parameter space with respect to a specific reference is indeed the best possible estimate of the global optimum of the model. In this particular situation, the search for an optimum is done within an *individual structural limitation* of a climate model, that is for one target variable only. Nevertheless, if more references are available, the search for an optimum must be started again by considering all the references, in order to find the best possible balance given the *collective structural limitations* of a climate model.

This search has to take into account the personal preferences of climate modelers, in particular the output variables they target, the associated references, and how they compare these output variables to the chosen references, i.e. the metrics they use. An overtuned model is one for which a better balance in model results given the personal preference of climate modelers can be found. There is thus a subjective component in the definition of overtuning: a model that seems well-balanced for some climate modelers given its structural limitations may seem overtuned for others. Different climate modelers might indeed have different priorities, and depending on these priorities, tolerate more or less different kinds of errors. For example, some climate modelers might place more emphasis on predictions of mid-latitude heat waves, while others might prioritize predictions of precipitation in the tropics, or modes of intraseasonal variability such as the Madden-Julian oscillation.

4.5 Automatic tuning methods: a shift in tuning paradigm?

4.5.1 Limitations of climate model tuning by hand

In the previous subsections, we defined the objective of tuning as the search for the best possible balance within the structural limitations of a climate model. We have seen that to find this balance, it is necessary to compare target outputs of climate model simulations (either in SCM or in a full GCM mode) to well-chosen references. For each set of parameters, the differences between the references chosen and the target variables of the simulations can be evaluated using metrics. The values of the parameters finally kept are those that seem to perform best according to these metrics, with potentially different weights for different metrics. The choice of target variables and metrics reflect the priorities of the climate modeler, there is thus a subjective component in the tuning process.

As the parameter space – the set of all combinations between parameter values – is extremely large, not all values of parameters can be tested. It is not possible to know comprehensively the structural limitations of a climate model. The choice of appropriate parameters is therefore a delicate task, involving not only scientific skills, but also a great deal of intuition and expert judgment (Hourdin et al., 2017a). There again, the subjectivity of climate modelers might play a role. By experience and by following a trial and error method, climate modelers need to intuit which parameters are the most useful for tuning the model and which values of these parameters need to be chosen to make the model perform best according to the various metrics. The full tuning of a GCM can take several months to several years, and contrary to a traditional calibration procedure, is in part subjective and difficult to reproduce, making it challenging to compare different climate models.

4.5.2 Automatic tuning methods could make climate model tuning more similar to a traditional calibration procedure,

Recent developments in automatic tuning methods could make the tuning process both faster and more reproducible. In chapter 5, we will see in practice how to automate certain steps of the tuning process with the HIGH-TUNE (HIGH-resolution simulations to improve and TUNE boundary layer cloud parameterizations) explorer *htexplo* a tool developed recently by two French climate modeling groups (IPSL and CNRM) in collaboration with Exeter University (Couvreur et al., 2021; Hourdin, Williamson, et al., 2020). To use this tool, climate modelers have to choose explicitly first the tuning parameters and their range of acceptable values, then the references, the metrics and the uncertainty around them, and finally the degree of agreement between the reference and the model they are ready to accept. Previously, all of these choices were implicit, and sometimes even unconscious for climate modelers. With *htexplo*, the tuning of climate models remains a complex issue, but better posed, in the sense that it can be broken down into a series of simpler problems, such as: which tuning parameters should be chosen? what is the range of acceptable values around them? which references and metrics are the most relevant? how to take into account the uncertainty around the references and around the model in the tuning process?

Once these different inputs have been given, *htexplo* searches for optima globally in the parameter space. This is a major advantage over traditional tuning methods. When tuning is done by hand, moving the parameters one after the other, there is a risk of being blocked in a local optimum in the parameter space. On the contrary, as *htexplo* explores a much larger sample of parameter space, it helps climate modelers to understand what a climate model can and cannot do. Expert judgment remains crucial to explore and understand the structural limitations of a climate model, but *htexplo* guides this expert judgment rather than replacing it. It also facilitates the comparison between model outputs and target metrics, thus preventing the risks of introducing compensating errors in the model. Indeed, if many metrics are used, model errors will be more difficult to compensate by tuning for all the metrics. Therefore, the more diverse the references used, the lower the risks of introducing compensating errors in the model. Finally, *htexplo* reduces the risks of overtuning, as defined previously, by facilitating the search for the best possible balance in the whole parameter space with respect to well-chosen references and given the structural limitations of a climate model. As a consequence, *htexplo* makes the tuning process more *robust*.

It also makes it more standardized. One aspect of this standardization concerns the references used in the tuning process. In a standard calibration procedure, there is usually a collective agreement among a community of specialists on the standards used as references to calibrate an instrument. In contrast, there is no consensus in the climate modeling community on the most appropriate references to tune a climate model (Hourdin et al., 2017b). In support of *htexplo*, reference banks are gradually being developed by climate modeling centres: *htexplo* relies in particular on the development of a bank of standardized 1D case studies, which allow climate models to be tuned against LES references under simplified conditions. This standardization of the references used in the tuning process makes it more similar to a traditional calibration procedure.

4.5.3 facilitate the interpretation of GCM results,

More generally, automatic tuning methods such as *htexplo* bring an important building block in the epistemological foundations of climate modeling. Historically, the tuning of climate models made the interpretation of model results difficult. In particular, in multi-model ensembles produced by CMIP (see Chapter 6), it was not clear how much of the difference between models was due to their structural differences and how much to their tuning. *Htexplo* makes it possible to separate these differences by playing on two tables. First, it allows to assess systematically the parametric uncertainty model by model. With *htexplo*, the range of plausible worlds allowed by a given model can be known. For instance, it is possible to use *htexplo* to explore the spread in the model equilibrium climate sensitivity for different parameter sets which give reasonable results with respect to the tuning references chosen. Second, *htexplo* opens the door to a tuning of different climate models using a similar standardized protocol. This would make the results of climate models comparable independently of their tuning and allow the remaining model spread – due only to model structural differences – to be evaluated.

The validation of climate models would be also facilitated by automatic tuning methods. In practice, there is a consensus that references used to validate climate model results should not be used for tuning. However, as the tuning of GCMs is rarely documented, it is difficult to check that different references are used to validate and tune climate models. As a consequence, when models of CMIP multi-model ensembles are compared and evaluated using the same set of observations, some models may have been tuned using the observations taken as references. Automatic tuning methods could address this problem by offering the potential to separate and standardize the references used in the tuning process and those left for the validation of climate models.

4.5.4 and guide the improvement of parameterizations

Finally, a major motivation behind *htexplo* is to guide the development of new parameterizations. When a new parameterization is introduced into a climate model, a question arises: has this new parameterization improved the model or not? Answering this question seems a prerequisite for scientific progress in climate science. Otherwise, there is a risk for any climate model to take “one step forward and two steps back” when a new parameterization is added. However, in most cases, the answer is far from obvious. In fact, when a new parameterization is introduced in a climate model, the quality of its results is often initially degraded as the changes brought to the model can break the delicate balance of its tuning (Jakob, 2010). A long tuning phase has to be carried out before potential improvements can be seen. Moreover, whatever the results of this tuning phase, there will always be some doubt as to whether the observed improvements or deteriorations are due to the new tuning or to the new parameterization itself. It is therefore rare to see in the literature a systematic comparison of the results given by a climate model before and after the addition of a new parameterization.

These difficulties and the effort tuning requires slow down the development of new parameterizations considerably and discourage climate scientists from engaging in the development of new parameterizations. This could explain the often-quoted lack of progress in model development (Jakob, 2010; D. Randall et al., 2003). By making the tuning process both shorter and more reproducible, *htexplo* is a game changer. It could guide the development of new parameterizations and enable a quick assessment of the improvements brought by a new parameterization. We will see this in practice in the following chapter, where we will use *htexplo* to

guide the development of a cloud and precipitation overlap parameterization in the atmospheric model LMDZ.

4.6 Conclusion

Climate model tuning challenges existing epistemological frameworks and requires the construction of new ones. So far, epistemological studies on tuning have been limited to the difficulties posed by tuning on the interpretation of GCM results. In particular, many philosophers focused on the validation of GCM results and noted that because of tuning, coupled GCMs are not testable from observations in the same sense that a theory is testable from observations. On the contrary, few epistemological studies investigated tuning process itself, as a distinct scientific activity. Based on the recent effort of some climate modelers to document the tuning strategies they use (Mauritsen & Roeckner, 2020; Mignot et al., 2021), and our own experience of tuning with the atmospheric model LMDZ, we attempted in this chapter to define climate model tuning as a specific scientific practice. We gave its scope and purpose, the needs it meets and some of the epistemological questions it raises. We also compared climate model tuning to calibration procedures used in other scientific fields. Noting that standardization and reproducibility were two important characteristics of traditional calibration procedure, we explained why automatic tuning methods could make the tuning process more robust and more similar to a traditional calibration procedure. Hopefully, these epistemological reflections will inspire future fruitful collaborations between philosophers of science and climate modelers to further build the epistemological foundations of climate model tuning.

Chapter 5

Parameterization development and tuning in practice – Cloud and precipitation overlap in LMDZ

Contents

| | | |
|------------|---|------------|
| 5.1 | Introduction | 113 |
| 5.2 | Definition of progress for parameterizations | 115 |
| 5.3 | Presentation of the cloud and precipitation overlap parameterization | 117 |
| 5.3.1 | Motivations | 117 |
| 5.3.2 | Description of the new parameterization | 120 |
| 5.4 | Results | 124 |
| 5.4.1 | First results in SCM without tuning | 125 |
| 5.4.2 | SCM tuning | 129 |
| 5.4.3 | GCM tuning | 136 |
| 5.5 | Conclusion | 144 |

5.1 Introduction

In the previous chapters, my reflection on convection parameterizations and tuning was based partly on a critical review of the philosophy of science and climate modeling literature on the subject, partly on the many discussions with climate modelers that I had during my three years at the Laboratoire de Météorologie Dynamique, and partly on my own experience in developing parameterizations and tuning with the LMDZ atmospheric model. A central experience of my Ph.D. thesis was indeed to introduce a new parameterization to the LMDZ atmospheric GCM and to conduct a substantial work of model re-tuning, using the HIGH-TUNE explorer *htexplo*, following the addition of this new parameterization. The new parameterization implemented led to significant improvements in both the Single Column Model (SCM) and General Circulation Model (GCM) version of LMDZ and will therefore most likely be used in the reference version of LMDZ for the next phase of CMIP. Just as my work on cold pools nourished my reflection on objects in Chapter 2, the work done with this parameterization nourished my

epistemological reflection on parameterizations and tuning.

I present this work here as a case study to show how some of the epistemological reflections from the previous chapters arise in practice when developing a new parameterization or tuning a GCM. In particular, a question that has been on my mind throughout this work is: how do we know whether a given parameterization improves a model or not? This question is crucial for any climate modeler. Anyone who adds a new parameterization to a GCM wonders whether the new version of the model is better or not than the previous one. When a new parameterization improves a GCM, it should be kept, otherwise it should be abandoned. If we are not able to distinguish between parameterizations that improve a given GCM and those that deteriorate it, there is a risk of moving backwards when adding new parameterizations to a GCM.

Yet, we will see that the meaning of progress for a GCM is not easy to define. Let's take a step back and ask ourselves what we mean by scientific progress in general. In his book *The Structure of Scientific Revolutions*, Kuhn explains that the progress of science happens not only through the accumulation of knowledge, but also through the replacement of some ideas by others. When two ideas are in competition, their puzzle-solving power is crucial to determine which idea will be preferred. A new idea will be preferred if it preserves much of the ability of its predecessor to solve problems, but also addresses important and previously unexplained issues (Kuhn, 1962, p. 129).

If we try to apply Kuhn's theory of scientific progress to parameterizations, we could say that a new parameterization improves a climate model if the results between the new version and the original version are comparable, except for some aspects of model results for which the new version shows significant improvements. Nevertheless, this account of scientific progress for parameterizations seems overly simplistic. First, the ability of a new parameterization to solve outstanding anomalies and to preserve previous model abilities depends on the tuning of model parameters. Without an appropriate tuning phase, the introduction of a new parameterization may break the delicate balance between model parameters and lead to unrealistic results. Moreover, even after an adequate tuning phase, a new parameterization often deteriorates some aspects of the model. Kuhn's criteria for evaluating scientific progress must therefore be relaxed – in practice, it is very rare that a new parameterization preserves most of the GCM abilities, and in addition improves some of its major deficiencies. Finally, the conceptual progress brought by a parameterization must be taken into account. Indeed, the results of a model must not only be accurate, but also interpretable. A new parameterization may facilitate or not this interpretation.

New criteria must therefore be found to define the specific meaning of the word *progress* for parameterizations. This chapter has two main objectives: first, it aims to present the scientific results we obtained by adding a cloud and precipitation overlap parameterization in LMDZ and re-tuning the model using the HIGH-TUNE explorer *htexplo* (see 4.5.2). This is one of the first times that *htexplo* is used to re-tune a GCM following the addition of a new parameterization, and our results provide a test bed for this automatic tuning method. By documenting the various tests we made and the results we got, we hope to facilitate and encourage the future use of this tool. The second objective of this chapter is to conduct an epistemological reflection on the notion of scientific progress for parameterizations. This reflection is initiated in section 5.2 where we distinguish two types of progress for parameterizations and propose some criteria to evaluate them. We continue this reflection in the following sections, where we present

respectively the cloud and precipitation overlap parameterization (Section 5.3) and the SCM and GCM results (Section 5.4). For clarity, in these sections, we distinguish scientific results from the epistemological reflections that accompany them by placing these epistemological reflections in blue boxes. These boxes allow the reader to focus, depending on his or her interest, either on the technical content of the chapter, on its epistemological content, or on both.

5.2 Definition of progress for parameterizations

It is not possible to compare two parameterizations per se, but only two parameterizations within a given GCM. Indeed, a GCM is not a modular structure where any parameterization could be easily plugged in or out (Jebeile & Crucifix, 2020). The different parameterizations of a model are largely interdependent on each other – this is what Lenhard and Winsberg (2010) call the “fuzzy modularity” of climate models. When a new parameterization is added to a model, the quality of the results obtained depends as much on the parameterization itself as on its interactions with the other parameterizations of the model. It explains why adding a new parameterization to a model sometimes does not produce the expected effects, even after a tuning phase. The improvements or degradations brought by a new parameterization are largely dependent on the parameterizations already present in the model. A same parameterization implemented in two different GCMs might improve one model and degrade the other.

Consequently, in the following, we do not give criteria to compare two parameterizations, but to compare two versions of the same GCM containing different parameterizations. We distinguish two types of progress brought by a parameterization in a given GCM: **conceptual progress** and **progress in terms of model results**. We define the conceptual progress brought by a parameterization to a GCM based on the following four criteria:

- its **Coherence**. We can distinguish between the coherence of a parameterization and the coherence of different parameterizations of a GCM taken together. A parameterization is coherent if it can be deduced logically from a few basic principles. The more a parameterization is based on well-understood and justified concepts, the more it is coherent. The parameterizations of a GCM taken together are coherent if they are logically deduced from the same assumptions, or at least from assumptions that are not contradictory. In particular, if a new parameterization combines processes that were previously described by different parameterizations, it is likely to increase the coherence of the model.
- its **Interpretability**. The parameterizations of a GCM are interpretable if their parameters and variables can be compared to real-world phenomena. For example, mass-flux parameterizations make GCMs more interpretable than ad hoc moist convective adjustments, because their formalism is easier to interpret in terms of real physical processes (see 3.3.1). The more realistic a parameterization is in its formulation, the more interpretable it is since, by design, a realistic parameterization is based on objects that exist in the real world. To increase the realism of a parameterization, and thus its interpretability, it is often necessary to use ad hoc assumptions. The latter, by definition, are not well justified and therefore decrease the coherence of a parameterization. Hence, a balance has sometimes to be found between the coherence of a parameterization and its interpretability. A perfectly coherent parameterization is often very idealized, therefore difficult to compare to real phenomena. Conversely, most parameterizations that are easily interpretable contain ad hoc assumptions in their formulation, and are therefore not fully justified.

- its **Simplicity**. A new parameterization introduces new assumptions and formulas, which necessarily complicate the model. The degree of complexity brought by a new parameterization can be assessed in part by the number of new variables and new tuning parameters it introduces. The more variables and tuning parameters a parameterization contains, the more it tends to complicate a GCM. The principles used to build a parameterization can also be more or less complex. If a new parameterization comes in place of a more complex one, it can, on the contrary, simplify the model.
- its **Comprehensiveness**. The ultimate goal of GCMs is to represent all processes relevant for climate. If a parameterization adds an important missing process to an existing GCM, it improves its comprehensiveness.

We will refer to these criteria as the CISC criteria (Coherence, Interpretability, Simplicity, Comprehensiveness). The CISC criteria provide a framework for evaluating the improvements brought by a parameterization in a given climate model. The assessment of each criterion depends on the parameterizations already present in the model: a parameterization that would bring a conceptual progress to a given climate model may, on the contrary, unnecessarily complicate another model with a different structure.

What about progress in terms of model results? To compare the results of a GCM with a new parameterization to those given by the original GCM, different choices have to be made. First, **the conditions under which the GCM simulation is carried out have to be defined** (e.g. SCM case studies, stand-alone atmospheric simulations forced with appropriate boundary conditions, coupled simulations, etc.). Depending on these conditions, different types of progress associated with the addition of a new parameterization can be observed: in particular, progress in SCM case studies does not systematically translate into progress in atmospheric or coupled simulations.

Second, **a tuning protocol as standardized as possible should be used to tune both the standard and the new version of the model**. Without such a protocol, this will be difficult to diagnose whether the differences observed are due to the parameterization itself or to a different tuning protocol. Historically, this difficulty has considerably hindered the development of parameterizations.

Finally, **we have to choose how the comparison between the original and the new version is conducted**. This comparison consists of three steps: 1. First, it is necessary to choose **which variables to focus on**. Not all output variables of the model are equally important, and the subset of output variables to evaluate has to be defined. 2. Then, we need to precise **which references we use** to evaluate these output variables. Various references can be used – satellite observations, reanalyses, high-resolution simulations – to target the different output variables chosen. 3. Finally, we have to choose **how the comparison between the output variables and the references is calculated** by defining some “metrics”. The metrics give the formula used to compare the proximity of model output variables to the references chosen. Here again many choices can be made: it is possible to use various statistical indicators, such as the mean, the variance, the distribution in space and time, etc. These three steps objectively define which aspects of model results are most important for the climate modelers, thus what the comparison is about.

In simplified simulations such as SCM case studies, the improvements provided by a new

parameterization sometimes jump off the page. It happens when most of the model chosen output variables are equally close to the references, except for one or a few output variables for which a significant improvement is observed. In this case, Kuhn’s criteria does apply: successful parameterizations “preserve a great deal of the most concrete parts of past achievements and [...] permit additional concrete problem-solutions besides” (Kuhn, 1962, p. 169). On the contrary, in more complex simulations such as global simulations (atmospheric-only or coupled simulations), the progress brought by a new parameterization is often less obvious, as an improvement of some model output variables is virtually always associated with a degradation of others. In this case, whether a new parameterization improves or not a given climate model is not clear. In the following, we will look at these issues in practice by introducing a cloud and precipitation overlap parameterization in LMDZ and examining whether it improves the model or not.

5.3 Presentation of the cloud and precipitation overlap parameterization

5.3.1 Motivations

Variety of motivations for developing new parameterizations

In parameterization development, innovation is no guarantee of success: a new parameterization may make a model unnecessarily complex, or deteriorate rather than improve its results. There must therefore be good reasons for embarking on the development of a new parameterization. Improving the model results is often an important motivation and typically, the development of a new parameterization aims at **reducing major biases identified in a model**. The first moist convective adjustment schemes were for instance developed with the objective of correcting the discrepancy between the atmospheric profiles given by GCMs in the tropics and the observations (see Chapter 3).

The development of a new parameterization can also be motivated by the conceptual progress it brings to the model. The objective is then either **to simplify or unify the model, to correct some errors or inconsistencies, to represent processes deemed important that are absent from the model, or to make the parameterizations easier to interpret in terms of physical processes**. Interestingly, some parameterizations that were initially motivated by conceptual progress eventually led to progress in terms of model results. For example, the cold pool parameterization of Grandpeix and Lafore (2010) led to a correction of the diurnal cycle of convection over land in LMDZ (Rio, Hourdin, Grandpeix, & Lafore, 2009), although it was initially developed to improve LMDZ conceptually, by representing in the model a process considered important – cold pools – but ignored until then.

The parameterization discussed in this chapter is intended to improve the model both conceptually and in terms of model results. It was initially motivated by an anomaly found when comparing the SCM version of LMDZ to large-eddy simulations (LES) in the RICO case (see Section 1.3.2 for more details on this methodology). This case is based on the data from the “Rain in Cumulus over Ocean” field campaign, which was dedicated to the study of rain under shallow cumuli in trade wind regions (Rauber et al., 2007). It is a composite case based on a

three-week period with typical trade wind cumuli and a fair amount of rain, about 0.3 mm/day at the surface. In the standard version of LMDZ, however, rain is evaporated before reaching the ground. Figure 5.1 shows that most rain is evaporated in the cloud layer or just below cloud base. As the cloudy air is close to saturation, we would expect instead that a significant amount of rain falls in the cloud layer without being evaporated and reaches the surface, a result confirmed by LES and observations (VanZanten et al., 2011).

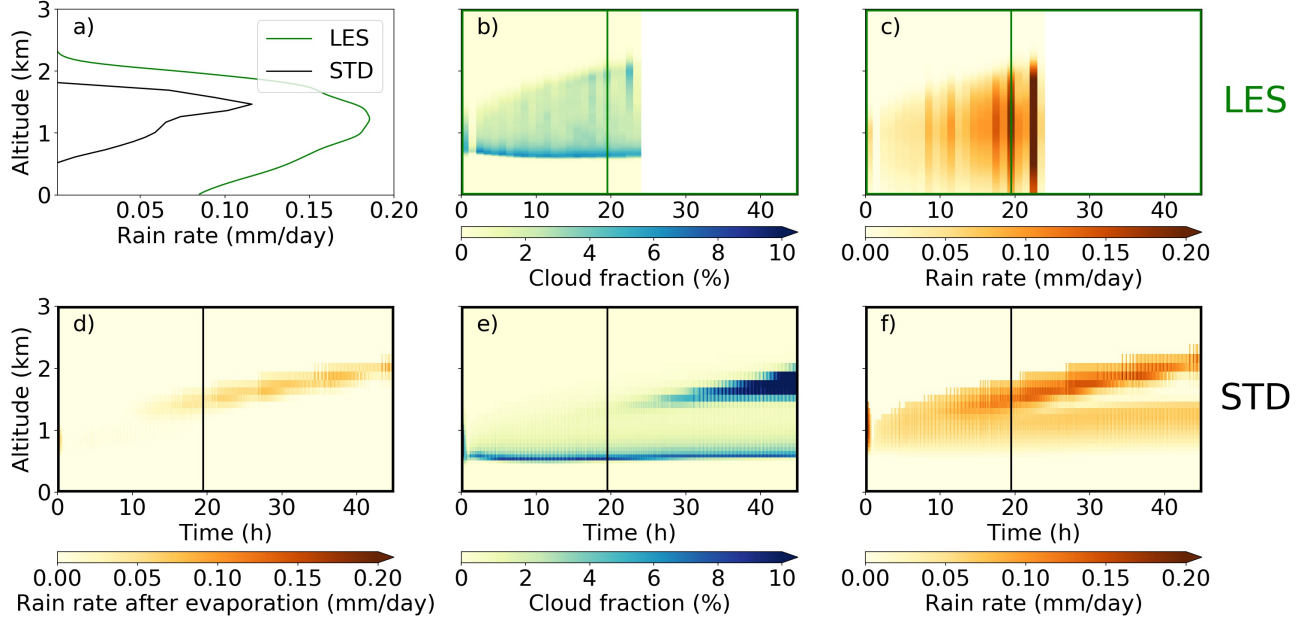


Figure 5.1: (a) Rain rate profile after 20 hours of simulation in a reference LES and in the standard version of LMDZ in the RICO case. (b) Cloud fraction and (c) rain rate given by the LES. (d) Rain rate after evaporation (e) cloud fraction and (f) rain rate after evaporation given by the standard version of LMDZ with 95 vertical levels. The LES used as reference was conducted with the MESO-NH model and is documented in Lac et al. (2018).

To solve this bias, we looked at the representation of precipitation evaporation in LMDZ large-scale condensation parameterization *fisrtilp*. In this module, the equations are implemented in a vertical loop, from top to bottom. Each vertical level k receives a precipitation flux from the level immediately above, with a precipitation flux initialized to zero for the top level. Then: 1) part of the precipitation flux from level $k + 1$ is evaporated 2) the cloud fraction and cloud water content at level k are calculated 3) part of the newly created cloud is converted into solid or liquid precipitation, thus increasing the corresponding precipitation flux. Here are the formulas used in each step.

1) Precipitation evaporation In the first step, the evaporation of precipitation (solid and liquid) is given by the formula introduced in Sundqvist (1988):

$$\frac{\partial P_{l,s}}{\partial z} = \mathbf{EVAP} \left(1 - \frac{q_t}{q_{sat}} \right) \sqrt{P_{l,s}} \quad (5.1)$$

with $P_{l,s}$ the liquid or solid precipitation flux density, in $\text{kg m}^{-2} \text{s}^{-1}$, q_t/q_{sat} the relative humidity at level k and \mathbf{EVAP} a tuning parameter.

2) Cloud formation The cloud fraction α and the solid and liquid water mixing ratios $q_{l,s}$ are then calculated from statistical schemes that describe the water distribution in the grid box (see Jam et al., 2013; Madeleine et al., 2020, for more details).

3) Precipitation formation Finally, a certain fraction of cloud liquid water is converted into precipitation. For liquid clouds, the equation governing this conversion is also taken from Sundqvist (1988):

$$\frac{dq_l}{dt} = -\frac{q_l}{\mathbf{CLTAU}}(1 - e^{-(\frac{q_l}{\mathbf{CLC}})}) \quad (5.2)$$

where \mathbf{CLTAU} is the characteristic time constant of precipitation formation and \mathbf{CLC} a threshold for liquid cloud water above which the formation of precipitation increases sharply. For ice clouds, the formation of precipitation is written:

$$\frac{dq_s}{dt} = \frac{1}{\rho} \frac{\partial}{\partial z}(\rho w_{iw} q_s) \quad (5.3)$$

with ρ the density of the air and w_{iw} , the fall speed of ice crystals.

Taking a closer look at these equations, we realized that the formula used for the evaporation of precipitation in LMDZ was different from that given in Sundqvist (1988). Indeed, Sundqvist considers that the cloudy fraction of the grid is saturated, so that precipitation evaporation only occurs in the clear-sky fraction of the grid. Consequently, he applies the formula 5.1 using the clear-sky relative humidity $q_{clr}/q_{sat,clr}$ and multiplies the result by the clear-sky fraction to deduce the total evaporation in the grid box. In contrast, in the standard version of LMDZ, the formula 5.1 is applied using the average relative humidity in the grid q_t/q_{sat} . In LMDZ, evaporation occurs in fact before the formation of clouds, and is calculated without taking their presence into account. Precipitation evaporation is calculated over the whole grid box as if there were no clouds, whereas in the real world, one would expect only the non-cloudy precipitation fraction to fall in unsaturated air, and thus to evaporate.

In addition to this difference, an assumption is added in LMDZ compared to Sundqvist (1988): precipitation evaporation cannot saturate a fraction of the grid box greater than the area of clouds above. In other words, the quantity of water vapor added by evaporation at level k is limited to $\alpha_c^{ev}(q_{sat} - q_t)$, with α_c^{ev} the maximum cloud fraction in the overlying layers (reset to zero whenever precipitation stops at a certain level) (Madeleine et al., 2020). This somewhat arbitrary assumption was added to the model to limit the overestimation of precipitation evaporation in formula 5.1.

To summarize, in the standard version of LMDZ, precipitation is on the one hand overestimated, especially in the cloud layer, because formula 5.1 is applied to the whole grid box rather than to the clear-sky fraction of the grid only. On the other hand, to partially compensate for this overestimation, an ad hoc assumption has been added to limit precipitation evaporation to the maximum cloud fraction in the overlying layers. This treatment of precipitation evaporation in LMDZ did not seem very adequate and we suspected it to be responsible for the unrealistic behavior of LMDZ in the RICO case. Panel d in Fig. 5.1 supports this claim by showing that the rain is almost entirely evaporated at each level in LMDZ. The comparison of panel d and f shows that most of the rain formed at level $k + 1$ is evaporated at level k . As a consequence, all the rain is evaporated above cloud base in LMDZ in RICO and does not reach

the lowest subcloud layers.

To correct this unrealistic behavior, we looked for a parameterization taking into account the fact that part of the precipitation flux falls in cloudy air and is thus not evaporated. This led us to adapt to LMDZ the parameterization of Jakob and Klein (2000), which separates the clear-sky and the cloudy fraction of the precipitation flux.

5.3.2 Description of the new parameterization

Adapting a parameterization to a new model

Adapting a parameterization developed for a certain GCM to a new model is not straightforward. The language of the new model has to be taken into account, that is both its vocabulary – the variables it contains – and its syntax – how and in which order processes are represented in the model. Two parameterizations based on the same ideas and equations can therefore be implemented quite differently in two different GCMs. The order in which processes are represented is particularly important because many modules of the code interact with each other. By choosing to represent a process A before a process B , one represents, in the best case, only the influence of $A \rightarrow B$. If B is calculated after A , the reciprocal influence of $B \rightarrow A$ cannot be taken into account, at least not directly. B can at best exert an indirect effect on A in the next time step through its effect on the large-scale variables of the model.

In our example, there is a reciprocal interaction between the evaporation of precipitation and the formation of clouds at each vertical level. The evaporation of precipitation depends on the presence or absence of saturated air, i.e. clouds, in the grid. Conversely, the evaporation of precipitation increases the relative humidity at each vertical level, and may thus also increase the cloud fraction at each level. In Jakob and Klein (2000), precipitation evaporation follows cloud formation, whereas it precedes it in LMDZ. It was necessary to take into account this difference to use nevertheless Jakob and Klein (2000) in LMDZ – the adaptations that had to be made are detailed in the text.

In addition with these adaptations, some minor changes are often made to the original parameterization when it is implemented in a new model. They generally aim to correct certain problems or deficiencies identified in the model following the addition of the new parameterization. For instance, when the parameterization of K. A. Emanuel (1993) was implemented in LMDZ (see 3.4.1), an improved mixing representation was introduced and detailed in Grandpeix et al. (2004). In our case, to guarantee a linear behavior of the model, we choose to add to Jakob and Klein (2000) a linear decrease of the clear-sky and cloudy precipitation fractions when the corresponding precipitation intensities fall below a certain threshold (see below).

The parameterization given by Jakob and Klein (2000) relies on a separation of the total precipitation flux into a clear-sky precipitation flux $P_{l,s}^{clr}$ and a cloudy precipitation flux $P_{l,s}^{cld}$. For simplicity, we omit the indices l and s in the following, but one should keep in mind that the proposed separation is performed at each level for both liquid and solid precipitation fluxes.

The parameterization of Jakob and Klein (2000) is based on the following equations:

$$\begin{aligned} P &= P^{clr} + P^{cld} \\ \alpha_P &= \alpha_P^{clr} + \alpha_P^{cld} \end{aligned} \quad (5.4)$$

with α_P the total fraction occupied by precipitation, α_P^{clr} the fraction occupied by precipitation falling in clear-sky, and α_P^{cld} the fraction occupied by precipitation falling in clouds. There are therefore four independent variables – P^{clr} , P^{cld} , α_P^{cld} and α_P^{clr} – that needs to be calculated at each vertical level k .

To adapt Jakob and Klein (2000) to LMDZ, we make the following modifications to the representation of precipitation in *fisrtlp*. First, when calculating the evaporation of precipitation (step 1), we assume that only the clear-sky precipitation flux evaporates following a discrete version of equation 5.1:

$$\Delta P_k^{clr} = \min(\Delta z \times \mathbf{EVAP}(1 - \frac{q_{clr}}{q_{sat,clr}}) \sqrt{P_{k+1}^{clr}}, P_{k+1}^{clr}) \quad (5.5)$$

$q_{clr}/q_{sat,clr}$ being, in principle, the relative humidity in the clear-sky fraction of the grid box. A technical problem appears because in LMDZ, the evaporation of precipitation (step 1) precedes the formation of clouds (step 2). Thus, we do not know at this stage what the relative humidity is in the clear-sky fraction of the grid box. Two options are possible:

1. take the relative humidity in the clear-sky fraction at level $k + 1$ in the equation 5.5. Physically, this choice can be justified because air cooled by evaporation tends to sink slightly, which results in a difference between the altitude where evaporation is calculated and the one where cooling is applied. The effect of this choice, however, depends on the thickness of the layer.
2. keep the average relative humidity in the entire q_t/q_{sat} grid box considering the following two limiting cases:
 - if $\alpha_{c,k} \ll 1$, $q_{clr}/q_{sat,clr} \approx q_t/q_{sat}$
 - if $\alpha_{c,k}$ is higher, under a maximum-random assumption for cloud overlap, the clear-sky precipitation flux will be relatively small, so the choice of $q_{clr}/q_{sat,clr}$ or q_t/q_{sat} for the calculation of evaporation will not affect strongly moisture tendencies over the grid box.

In the current version of the parameterization, we chose this second option.

Then, the clear-sky precipitation flux at level k is given by :

$$P_k^{clr} = P_{k+1}^{clr} - \Delta P_k^{clr} \quad (5.6)$$

If all the precipitation flux evaporates at level k , i.e. if $P_k^{clr} = 0$, $\alpha_{P,k}^{clr} = 0$. Otherwise, we assume that the evaporation of precipitation is spatially homogeneous and $\alpha_{P,k}^{clr} = \alpha_{P,k+1}^{clr}$. There is no evaporation in the cloud fraction of the grid, so $\alpha_{P,k}^{cld} = \alpha_{P,k+1}^{cld}$ and $P_k^{cld} = P_{k+1}^{cld}$.

The cloud formation itself (step 2) does not change compared to 5.3.1 and gives access to the cloud fraction $\alpha_{c,k}$. It defines a new partition between cloudy and clear air at level k and we need to update $\alpha_{P,k}^{cld}$, $\alpha_{P,k}^{clr}$, P_k^{cld} and P_k^{clr} accordingly. Indeed, a fraction of the cloudy

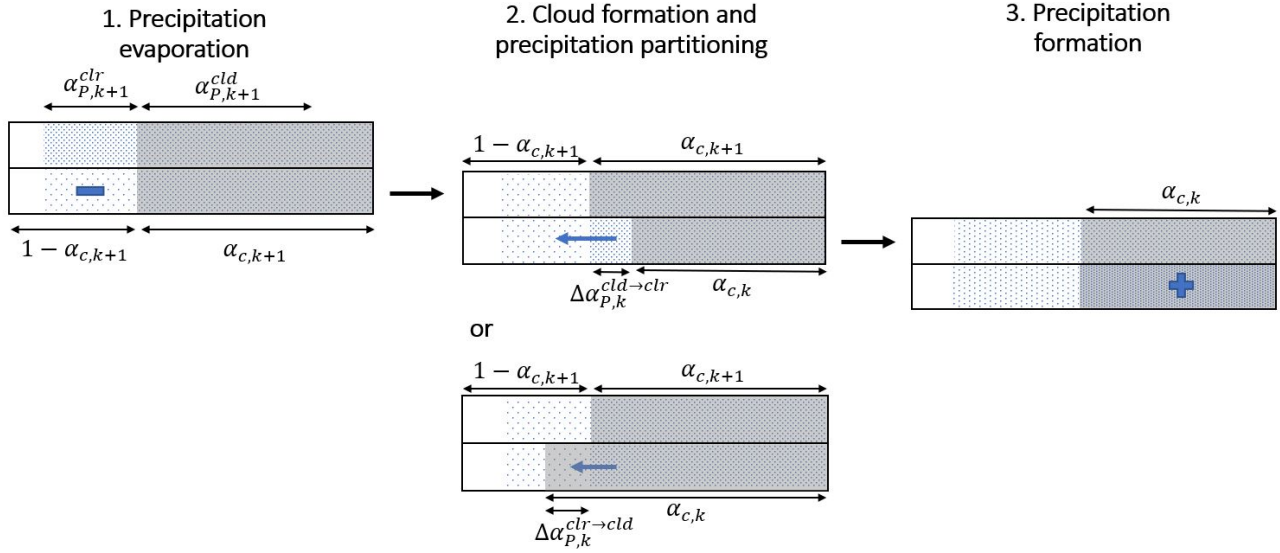


Figure 5.2: Diagrams describing the three main steps in the treatment of precipitation in the new parameterization: (1) precipitation evaporation, (2) cloud formation and precipitation partitioning, and (3) precipitation formation. $\alpha_{P,k+1}^{clr}$ and $\alpha_{P,k+1}^{cld}$ refer respectively to the cloudy and clear-sky precipitation fractions, $\alpha_{c,k}$ and $\alpha_{c,k+1}$ to the cloud fractions at levels k and $k+1$, and $\Delta\alpha_{P,k}^{cld \rightarrow clr}$ ($\Delta\alpha_{P,k}^{clr \rightarrow cld}$) the cloudy precipitation fraction that passes into clear sky (respectively the clear-sky precipitation fraction that passes into a cloud). The blue arrows indicate a transfer of precipitation due to a homogenization of the cloudy or clear-sky precipitation fraction. The \pm refer to an increase/decrease in precipitation flux due to precipitation formation/evaporation. At the top level, before precipitation formation (step 3), both clear-sky and cloudy precipitation fluxes (and associated precipitation fractions) are zero.

precipitation flux might fall into clear sky and vice versa (see Fig. 5.2). To calculate the new partitioning of precipitation between clear and cloudy air, we adopt the following method, given in Jakob and Klein (2000). Noting C_k the total area covered by clouds from k to k_{top} , we first consider that:

$$(1 - C_k) = (1 - C_{k+1}) \times \frac{1 - \max(\alpha_{c,k}, \alpha_{c,k+1})}{1 - \min(\alpha_{c,k+1}, 1 - \delta)} \quad (5.7)$$

where $\delta = 10^{-6}$ to prevent division by zero. The equation expresses a maximum overlap for clouds in adjacent levels and a random overlap for clouds separated by clear levels (see Jakob & Klein, 2000, for more details). With this assumption, the fraction of the precipitation flux that passes from clouds to clear sky (see Fig. 5.2) is given by the following formula:

$$\Delta\alpha_{P,k}^{cld \rightarrow clr} = \alpha_{P,k+1}^{cld} - \min(\alpha_{c,k}, \alpha_{P,k+1}^{cld}) \quad (5.8)$$

and we deduce the corresponding transfer of precipitation flux as :

$$\Delta P_k^{cld \rightarrow clr} = \frac{\Delta\alpha_{P,k}^{cld \rightarrow clr}}{\alpha_{P,k+1}^{cld}} \times P_{k+1}^{cld} \quad (5.9)$$

For the flux from clear to cloudy air, we notice that the portion $\Delta C = C_k - C_{k+1}$ corresponds to the portion of clouds at level k not overlapped by clouds at any higher level. Thus, no precipitation can fall into it. Following Jakob and Klein (2000), we deduce the formula:

$$\Delta\alpha_{P,k}^{clr \rightarrow cld} = \max(0, \min(\alpha_{P,k+1}^{clr}, \alpha_{c,k} - \Delta C - \alpha_{c,k+1})) \quad (5.10)$$

and the corresponding transfer of precipitation flux:

$$\Delta P_k^{clr \rightarrow cld} = \frac{\Delta \alpha_{P,k}^{clr \rightarrow cld}}{\alpha_{P,k+1}^{clr}} \times P_{k+1}^{clr} \quad (5.11)$$

After the formation of clouds, we can update all variables describing the partitioning of precipitation:

$$\begin{aligned} \tilde{\alpha}_{P,k}^{cld} &= \alpha_{P,k+1}^{cld} + \Delta \alpha_{P,k}^{clr \rightarrow cld} - \Delta \alpha_{P,k}^{cld \rightarrow clr} \\ \tilde{\alpha}_{P,k}^{clr} &= \alpha_{P,k+1}^{clr} - \Delta \alpha_{P,k}^{clr \rightarrow cld} + \Delta \alpha_{P,k}^{cld \rightarrow clr} \\ \tilde{P}_k^{cld} &= P_{k+1}^{cld} + \Delta P_k^{clr \rightarrow cld} - \Delta P_k^{cld \rightarrow clr} \\ \tilde{P}_k^{clr} &= P_{k+1}^{clr} - \Delta P_k^{clr \rightarrow cld} + \Delta P_k^{cld \rightarrow clr} \end{aligned} \quad (5.12)$$

the tilde referring to the updated values in the current step.

Finally, in the last step (step 3), part of cloud condensates is converted into precipitation according to the equations 5.2 and 5.3. At level k , it generates a cloudy precipitation flux ΔP_k^{cld} . We assume that this flux is uniform below the cloud, therefore, at the end of the three steps at level k , we have:

$$\begin{aligned} \tilde{\alpha}_{P,k}^{cld} &= \alpha_{c,k} \\ \tilde{P}_k^{cld} &= P_k^{cld} + \Delta P_k^{cld} \end{aligned} \quad (5.13)$$

with the same convention for the tilde as in 5.12.

These three steps are essentially the adaptation of Jakob and Klein (2000) parameterization to LMDZ. We choose to add a fourth step to the original formulation in order to guarantee a linear decrease of the area covered by the precipitation flux when the precipitation intensity $I_k^{clr,cld} = P_k^{clr,cld} / \alpha_{P,k}^{clr,cld}$ is lower than a threshold value **RI**. In the original formulation, when the flux of precipitation (either P_k^{cld} or P_k^{clr}) goes to zero, the area it covers ($\alpha_{P,k}^{cld}$ or $\alpha_{P,k}^{clr}$) may drop abruptly from a high value to zero. To avoid such a sharp transition, we add the following last step at each vertical level k (see Fig. 5.3):

$$\begin{aligned} \alpha_{P,k,new}^{cld} &= \min(\alpha_{P,k}^{cld}, \frac{1}{\mathbf{RI}} \times I_k^{cld}) \\ \alpha_{P,k,new}^{clr} &= \min(\alpha_{P,k}^{clr}, \frac{1}{\mathbf{RI}} \times I_k^{clr}) \end{aligned}$$

Does the parameterization improve the model conceptually?

To determine whether this new parameterization represents a conceptual progress, we apply the CISC criteria to the new version of LMDZ:

- **Coherence:** the new parameterization helps to solve an incoherence in formula 5.1, used to calculate precipitation evaporation. This formula was supposed to be applied in clear sky only, but was actually used over the whole grid box even when clouds were present.
- **Interpretability:** in the standard version of LMDZ, an arbitrary assumption was used to limit the amount of precipitation evaporated to the maximum cloud fraction in the overlying layers (see 5.3.1 and Madeleine et al., 2020). The new parameter-

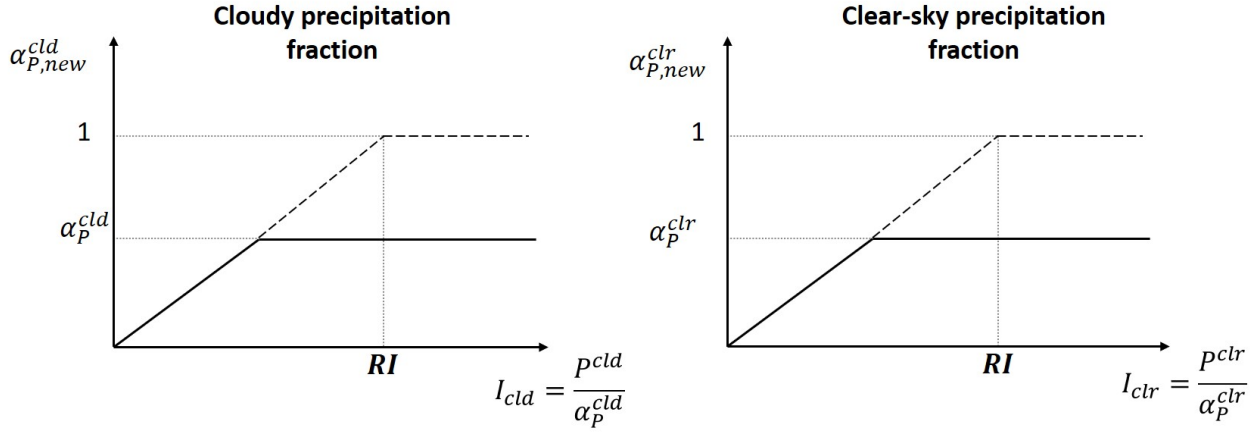


Figure 5.3: Diagram describing how the limitation of the cloudy (left) and clear-sky (right) precipitation fractions are calculated when the corresponding precipitation intensities fall below a certain threshold **RI**. In each case, the solid line determines the value effectively taken by the new cloudy/clear-sky precipitation fraction for a given rain intensity $I_{cld/clr}$ and the dotted line the maximum value permitted.

ization replaces this assumption by variables $(P^{clr}, P^{cld}, \alpha_P^{clr}, \alpha_P^{cld})$ and equations that are easy to interpret physically.

- **Simplicity:** the parameterization involves only 4 new variables and 1 tuning parameter (RI) and is based on simple geometrical considerations. Therefore, it does not complicate the model much.
- **Comprehensiveness:** the cloud and precipitation overlap parameterization takes into account an important process for precipitation evaporation absent from the standard version of LMDZ: the fact that part of the precipitation flux falls in cloudy air and is thus not evaporated.

The parameterization exposed here represents improvements in terms of Coherence, Interpretability and Comprehensiveness and shows only minor degradations in terms of Simplicity. Consequently, according to the CISC criteria, it represents a conceptual progress compared to the standard version of LMDZ.

5.4 Results

Ten years after Jakob and Klein (2000) was implemented and tested in the ECMWF model, Jakob (2010) looks back on this attempt and writes: “Although they clearly showed the need for such a parameterization based on a process study (Jakob & Klein, 1999), the implementation of the new parameterization into the full GCM had very little effect on the model results.”. In fact, improving a model conceptually does not imply improving its results. On the contrary, there may be compensating errors between the different parameterizations of a GCM (as defined in 4.4.2) and the improvement of a certain parameterization may remove some of these compensations, thus making new biases appear in the model.

Nevertheless, we will see in the following that the adaptation of Jakob and Klein (2000) to LMDZ gives very encouraging results in SCM case studies and in the global atmospheric GCM. This supports both the relevance of this parameterization for the LMDZ model, and the usefulness of the HIGH-TUNE explorer *htexplo* to re-tune a GCM following the addition of a new parameterization.

5.4.1 First results in SCM without tuning

In the RICO case, no precipitation reaches the ground in the standard version of LMDZ (see Fig. 5.1). The new parameterization is intended to correct this deficiency and to take into account the fact that part of the precipitation flux falls in a cloud, thus is not evaporated. Figure 5.4 gives the cloudy, clear-sky and total precipitation fluxes and the corresponding fractions in the new version of LMDZ in the RICO case. We fix the threshold value chosen for the limitation of precipitation fraction **RI** at $0.5 \text{ kg m}^{-2} \text{ s}^{-1}$. In each panel, the cloud layer predicted by the model is circled in gray. As expected, when a cloud is present, most of the precipitation is cloudy. On the contrary, below the cloud layer, all the precipitation flux falls into clear sky. It is partly evaporated, resulting in the progressive decrease of P^{clr} observed in panel b. The corresponding precipitation fraction α_P^{clr} (panel e) decreases as well in the lowest layers, suggesting that the local intensity of the precipitation flux P^{clr}/α_P^{clr} is lower than **RI**.

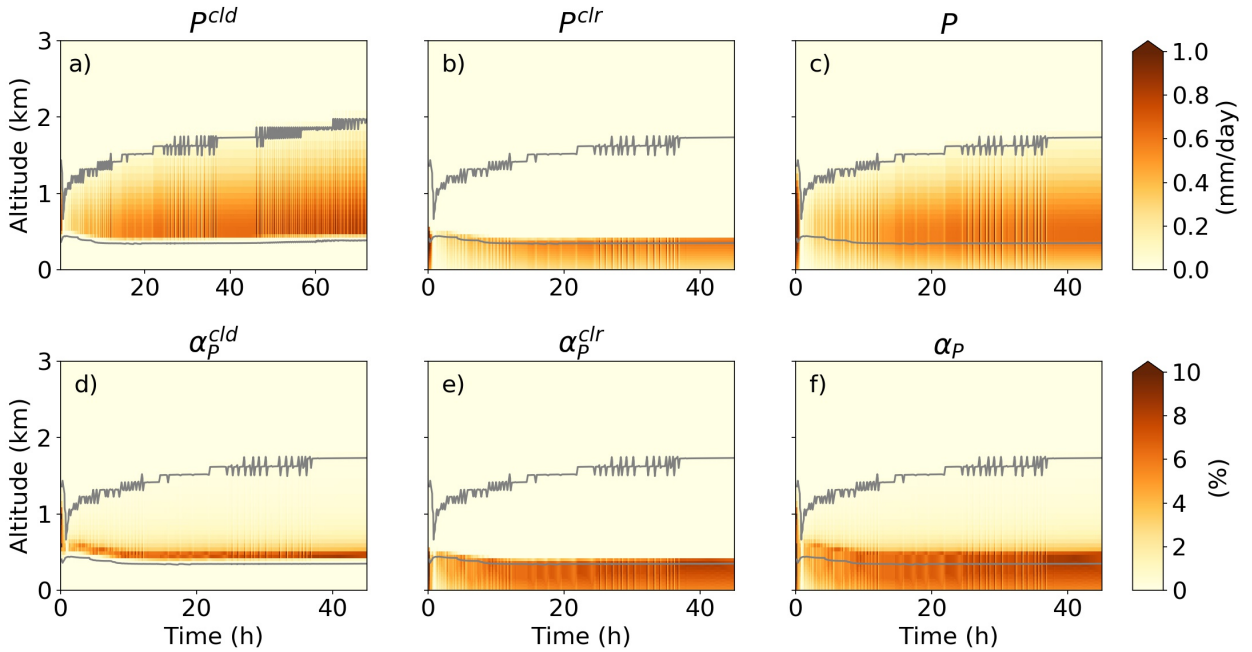


Figure 5.4: Cloudy (a), clear-sky (b) and total (c) precipitation flux densities and associated precipitation fractions (d-e-f) in the RICO case using the 95-level version of LMDZ containing the new parameterization. A value of $0.5 \text{ kg.m}^{-2}.\text{s}^{-1}$ is chosen for **RI**. Other parameter values are left unchanged with respect to the standard version of the model. In each panel, the cloud layer is contoured in gray.

Target variables vs intermediate variables

In chapter 4, we made a distinction between **target variables**, which a climate model seeks to predict as accurately as possible, and **intermediate variables**, which play a role in parameterizations but can be legitimately approximated. The justification we gave for the approximation of intermediate variables is based on the non-accuracy of the representations used in parameterizations. Here, the clear-sky and cloudy precipitation fluxes and the associated fractions could in principle be diagnosed in high-resolution simulations or observations. However, even with such diagnostics, the purpose of our scheme would not be to reproduce as accurately as possible these intermediate variables, but rather to predict realistic values for the model target variables. Therefore, when assessing the quality of model results, we do not focus on the clear-sky and cloudy precipitation fluxes and associated fractions, but on two target variables of the model: the cloud fraction and the surface rain rate.

In addition to the RICO case, we consider two other cases: ARMCU and SANDU. The first is a case of a diurnal cycle of continental shallow convection developed from observations collected on 21 June 1997 at the Atmospheric Radiation Measurement (ARM) site in Oklahoma, USA (Brown et al., 2002). The second is a transition case from stratocumulus to cumulus over subtropical oceans. As Sandu and Stevens (2011) explain, this second case was built by compositing the large-scale conditions sampled along a set of individual Lagrangian 3-day trajectories monitored in the northeastern Pacific during the summer months of 2006 and 2007. In addition to the REF case shown here, Sandu and Stevens (2011) provide a FAST case and a SLOW case, corresponding respectively to a faster and slower stratocumulus to cumulus transition, that we will use later in the tuning process.

To assess the quality of the results produced by the new version of LMDZ, we use a reference LES for each case. The ARMCU and RICO simulations are provided by the MESO-NH model and detailed in Lac et al. (2018), and the SANDU simulation was performed with the UCLA model (see Sandu & Stevens, 2011, for more details). The panels (a-b-c) in Fig. 5.5 give the cloud fractions simulated by the standard version of LMDZ (STD), by the new version (NEW) and by the reference LES. We choose in each case a time interval characteristic of the simulation: between the 7th and the 9th hour of simulation for the ARMCU case, between the 19th and the 25th hour of simulation for RICO and between the 60th and the 65th hour of simulation for SANDU. The panels (d-e-f) show the evolution of the cloud fraction during the whole simulation in the new version of LMDZ and enable us to check the representativity of the time interval chosen (highlighted in gray).

In **ARMCU**, the cloud fraction is comparable between the STD and the NEW version of the model. Both predict a cloud base altitude consistent with that given by the LES, but with a slightly lower cloud fraction near cloud base. Differences are observed with respect to the surface rain rate, which is zero in STD and positive in NEW. The surface rain rate in NEW is more or less in phase with that of the LES, but overestimated by a factor of 2 to 3.

In **RICO**, the surface rain rate is also null in STD, but positive in LES and NEW. As in ARMCU, surface rain rate is overestimated in NEW compared to LES. In addition, cloud base in NEW is lower than in STD and LES, leading to an overall cloud profile significantly further

from LES than that given by STD. This represents a major degradation in the NEW version compared to the STD version of LMDZ in RICO.

In **SANDU**, the cloud fraction and the surface rain rate are comparable between NEW and STD but largely overestimated compared to the LES. This can be explained by the presence of stratocumulus clouds in this case study, revealed by the relatively large cloud fraction in panels c and f. Since stratocumulus clouds are thin, the precipitation formed falls almost immediately into clear sky. Moreover, as their surface fraction is close to 1, it is virtually equivalent to evaporate over the whole cell (as in STD) or only over the clear-sky precipitation fraction (as in NEW). The precipitation and the cloud fraction in the SANDU case are therefore hardly affected by the new parameterization. The main difference between the STD and NEW simulations comes from the limitation of the clear-sky precipitation fraction when precipitation intensity falls below **RI**. This limitation results in a more gradual decrease of the rain rate towards the end of the simulation (after 40 hours) in NEW with respect to STD. In Fig. 5.5 panel l, it can be seen by the presence of more spikes on the black curve than on the red one in the last hours of the simulation.

Obviously, the LES used is not an exact reference and may contain biases. In particular, studies have shown that there are many uncertainties in the representation of microphysical processes in LES (e.g. Chlond & Wolkau, 2000; Grabowski, 2014). In the RICO case, VanZanten et al. (2011) documented for instance marked differences in surface rain rates in twelve LES using different microphysical parameterizations. Nevertheless, even if the surface rain rates predicted by these LES differed quantitatively, all but two of the simulations predicted positive surface rain rates at the end of the simulation, consistently with what was expected from observational data. Similarly, Van der Dussen et al. (2013) compared the surface rain rates predicted by six LES in SANDU – once again, quantitative differences were observed, but the surface rain rate stayed consistently below 1 mm/day in all simulations and, unlike the standard version of LMDZ, it did not show any spikes: it was instead continuous from the beginning to the end of each simulation. In terms of surface rain rate, LES should certainly not be taken as the truth, and we do not intend that the surface rain rate predicted by LMDZ matches perfectly with what LES predicts. Nonetheless, a LES gives qualitatively the rain rate that is expected at the surface and is thus a relevant reference to evaluate the impact of the new parameterization on SCM results.

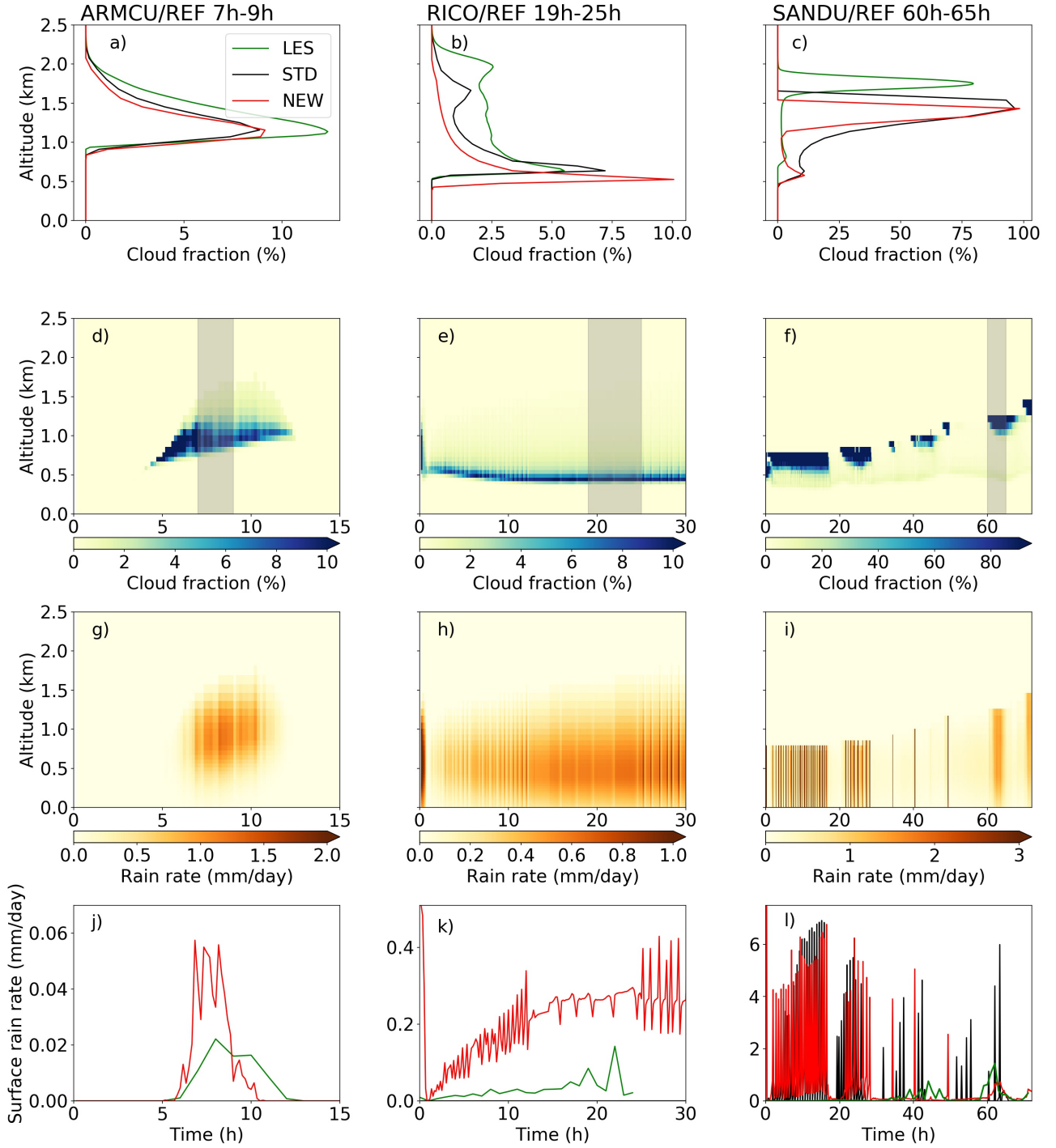


Figure 5.5: Cloud fraction in the standard version (STD, black), the modified version (NEW, red) of LMDZ at 95 vertical levels, and in the LES (green) in the (a) ARM, (b) RICO and (c) SANDU cases. The cloud fraction is averaged over the time interval specified for each case and highlighted in gray in panels (d-e-f), which show the evolution of the cloud fraction in ARM, RICO and SANDU in the modified version of LMDZ. (g-h-i) Same panels as (d-e-f) but for the rain rate in each case. (j-k-l) Surface rain rate in the standard version of LMDZ, the modified version, and in the LES simulation with the same colors as (a-b-c). For ARM and RICO, there is no rain at the surface in STD.

5.4.2 SCM tuning

Difficulties of SCM tuning by hand

In terms of cloud fraction, the results with the new parameterization are further from LES than with the standard version of LMDZ. In the RICO case, in particular, panel b in Fig. 5.5 shows an unrealistic decrease of the cloud base height in the new version compared to the standard version and LES. The new parameterization seems to break a certain balance between the different LMDZ parameterizations that ensure a realistic cloud base height in RICO. To restore this balance, the parameters of the model have to be tuned.

We initially tried to tune LMDZ parameters by hand. We used in particular the parameters **EVAP** and **CLTAU**, which are respectively involved in the formulas governing the formation and evaporation of precipitation in LMDZ. By a trial and error method, changing the value of one parameter at a time, we managed to obtain a good agreement with the LES in the RICO case in terms of cloud fraction, but the cloud fraction in SANDU or surface precipitation in one of the three cases then became unrealistic. If we would have spent more time on this manual tuning, we would probably have obtained a good agreement with LES in terms of surface precipitation *and* cloud fraction in the three cases considered, but this small experience was enough to convince us of the relevance of the HIGH-TUNE explorer *htexplo* to tune the parameters of LMDZ. More precisely, we noted in practice two difficulties specific to manual tuning:

- **its lack of reproducibility.** The values finally adopted for the parameters depend on the path chosen. In our example, we searched by trial and error for the best set of parameters, changing one parameter after another. We first adjusted the **EVAP** parameter. If we had started with another parameter, for example **CLTAU** or **CLC**, perhaps we would have obtained different results. The tuning of a full GCM is far from being limited to two or three parameters and depends even more on the path chosen.
- **the expertise it requires**, which relies on the one hand on a theoretical understanding of the role and structure of the different parameterizations of the model, and on the other hand, on a practical knowledge of the results produced by the model under different conditions and with different sets of parameters. The training needed to obtain this expertise requires a significant investment and is specific to a given climate model. Indeed, each model having its own language and behaviors, the practical knowledge acquired on a model cannot be easily transposed from one model to another. In our example, as we did not have a global expertise on the model, we limited ourselves to the parameters involved in the module *fisrtilp*.

These difficulties are linked to the lack of standardization of the tuning process by hand. In a classical calibration process (and in *htexplo*), standardization is indeed what makes the calibration reproducible, and also, what allows a non-expert user of an instrument to calibrate it.

Due to the difficulties of re-tuning the model by hand following the addition of a new parameterization, we used the HIGH-TUNE explorer *htexplo* described in Couvreux et al. (2021);

Hourdin, Williamson, et al. (2020). This tool uses an approach developed by D. Williamson et al. (2013), whose principle is to remove certain regions of the parameter space iteratively, refocusing the search on acceptable range of parameters at each step. The acceptability (or non-acceptability) of parameters is first defined by comparing the results of SCMs for different sets of parameters to LES. Once the parameter space has been significantly reduced using SCM simulations, a 3D tuning phase is performed with atmospheric GCM simulations to further reduce the parameter space using global metrics.

The full protocol used in the SCM tuning phase is documented in Couvreur et al. (2021); Hourdin, Williamson, et al. (2020). We present here only a sketch of its main steps:

1. One has to choose some tuning parameters and acceptable range of values for each of these parameters.
2. From this range of parameter values, the algorithm selects sets of parameters designed to optimally sample the parameter space.
3. SCMs are run with the selected parameter sets in different cases and compared to LES. This comparison is based on metrics, calculated from model target variables (such as temperature, specific humidity or cloud fraction).
4. A meta-model called the “emulator” extrapolates the value of the different metrics calculated for each set of parameters to the whole parameter space (see D. Williamson et al., 2013, for more details on how this emulator works).
5. For each point in the parameter space and each metric, a distance from the reference, called implausibility, is computed. This distance distinguishes between acceptable and non-acceptable values of the parameter space. It takes into account the difference with the LES simulations used as references, as well as three main sources of uncertainties: 1) The uncertainty around the reference, here the LES simulations used. 2) The uncertainty of the emulator, as we do not sample the entire parameter space, but a reduced (well-chosen) sampling of the parameter space. D. Williamson et al. (2015, 2013); D. B. Williamson, Blaker, and Sinha (2017) explain how to take into account this uncertainty in the calculation of implausibility. 3) The uncertainty associated with the tolerance to error for the metric considered. This tolerance to error is chosen according to the expertise of the climate modeler about the model structural limitations, i.e. what the model can and cannot do (as we defined them in 4.4.3). Depending on these structural limitations, the climate modeler might ask more or less to the model: indeed, if a model is asked to be more accurate than permitted within its structural limitations, it will empty the parameter space. It is therefore crucial for the climate modeler to have a good expertise on the structural limitations of the model to choose a relevant tolerance to error. For a given metric, the tolerance to error chosen reflects how close to the reference the climate modeler thinks his model can be.
6. Based on the implausibility, regions of the parameter space that are too far from the references used are ruled out. The remaining values form the NROY space (Not Ruled Out Yet) over which a new iteration – so-called “wave” – can be performed. The parameter space is progressively reduced by successive waves.

The first step – choosing some tuning parameters and a range of possible values for each of these parameters – is similar to what a climate modeler has to do, consciously or unconsciously,

when tuning a climate model manually. There is however a major difference: with *htexplo*, the choice of parameters and parameter values becomes explicit, whereas it was implicit before. Similarly, the choice of metrics and tolerance to errors has to be prescribed explicitly. As we have seen in chapter 4, this represents a major step forward in terms of standardization and reproducibility.

It was evident in our case as we used a set of parameters, a set of metrics and associated tolerances to error similar to the ones that proved successful in previous studies (Hourdin, Williamson, et al., 2020, in particular). More precisely, three types of parameters have been chosen: 1) parameters directly involved in the new parameterization or in the formation or evaporation of rain (EVAP, CLC, CLTAU, RI) 2) parameters involved in convection or in the formation of shallow clouds (A1, A2, B1, BG1, DZ) 3) parameters involved in the tuning of deep clouds, which do not have any influence on the shallow cumulus cases considered for the SCM tuning, but are kept for the GCM tuning phase (FALLV, OMEPMX, REI, RSQPO, RQSDP, RQSH). Table 5.1 gives, for each parameter, the minimum and maximum values defining the range explored, the reference value used in the standard version of LMDZ, and what the parameter controls in the model.

Figure 5.2 lists the metrics used for each case study. The metrics are averaged over a period of a few hours to smooth out possible numerical oscillations. The first two metrics $\theta_{400-600hPa}$ and $q_{v,400-600hPa}$ are vertical averages of potential temperature and specific humidity between 400 and 600 m. They assess the thermodynamical behavior of the subcloud layer. For cloudy cases, the maximum cloud cover $f_{cld,max}$ or the cloud height can also serve as metrics. The cloud height is computed using either an average height $z_{cld,ave} = \int_0^\infty f_{cld} z dz / \int_0^\infty f_{cld} dz$ or a weighted height $z_{cld,max} = \int_0^\infty f_{cld} z^4 dz / \int_0^\infty f_{cld} dz$, with a power 4 added to concentrate the integral on the highest cloud fractions (see Hourdin, Williamson, et al., 2020, for more details).

In addition to the metrics, a tolerance to error has to be prescribed for each metric. For the potential temperature and specific humidity, we prescribe the tolerance in terms of an absolute tolerance $\Sigma_T = 0.5$ K and $\Sigma_q = 0.0005$ kg/kg. We also prescribe a relative error on the height of clouds $\Gamma_z = \Sigma_z/z = 0.12$ and cloud fraction $\Gamma_f = \Sigma_f/f = 0.25$. These tolerances to error allow to compute the implausibility, a distance between the metric prediction by the emulator and the reference metric value (Hourdin, Williamson, et al., 2020; D. Williamson et al., 2013; D. B. Williamson et al., 2017). The space of parameters NROY is progressively reduced thanks to these implausibility calculations. A point in the parameter space is kept if its implausibility is lower than a threshold, progressively reduced from 3 in the first four waves, to 2.5 in the following three, and finally to 2 at wave 8 and the following ones.

At each wave, 90 simulations are performed. Figure 5.6 shows the error with respect to LES normalized by the tolerance to error metric by metric for each simulation performed with the new version (NEW+TUNING, blue dots) and standard version (STD+TUNING, red dots) of LMDZ at waves 1, 5, 15 and 30. In addition with the metrics defined previously, we use the average of the different metrics AVE. The dashed line represents the normalized error given by the standard version of LMDZ (STD) without tuning, i.e. with the original set of parameters. In the top panels, we see that the normalized errors in the ARMCU metrics stay low throughout the waves, suggesting that the ARMCU metrics are not very restrictive for the parameter space. On the contrary, the mean error (blue and red squares) by metric decreases with the number of waves already performed for the RICO and SANDU metrics, as well as

| Name | Min | Max | Ref | Controls |
|---------------|--------|-------|--------|---|
| EVAP | 5e-5 | 5e-3 | 1e-4 | Reevaporation of precipitation |
| CLC | 5e-5 | 5e-3 | 6.5e-4 | Autoconversion of cloud liquid water to rainfall |
| CLTAU | 4000 | 15000 | 900 | Characteristic time for the formation of rainfall |
| RI | 1e-3 | 5 | NA | minimum local rain intensity |
| A1 | 0.5 | 1.2 | 0.667 | Contribution of buoyancy to the plume acceleration |
| A2 | 1.5e-3 | 4e-3 | 2e-3 | Drag term in the plume acceleration |
| B1 | 0 | 1 | 0.95 | Scaling factor for entrainment or detrainment |
| BG1 | 0.4 | 2 | 1.1 | Width of the environment subgrid-scale water distribution |
| DZ | 0.07 | 0.15 | 0.07 | environmental air altitude shift for buoyancy calculation |
| FALLV | 0.3 | 2 | 0.8 | speed of fall of ice crystals |
| OMEPMX | 0.005 | 0.01 | 0.001 | maximum efficiency of cloud water → precipitation conversion |
| REI | 0.5 | 1.3 | 1 | effective radius of cloud particles |
| RQSPO | 40000 | 60000 | 45000 | parameters used to define the standard deviation of the subgrid-scale water distribution |
| RQSDP | 7000 | 25000 | 10000 | |
| RQSPH | 0.05 | 0.6 | 0.4 | |

Table 5.1: Parameters used in the tuning process with *htexplo*. The minimum and maximum values explored are given, as well as the reference value used in the standard version of LMDZ and what each parameter controls in the model. Inspired from Hourdin, Williamson, et al. (2020).

for the average error AVE, and becomes progressively better than STD. The best simulations for each wave in NEW+TUNING and STD+TUNING are indicated in gold and give better results than STD for most of the metrics. These results show that *htexplo* find regions of the parameter space where the new version of LMDZ predicts results closer to LES than STD and comparable with the best simulations found in STD+TUNING.

To compare the results given by the best simulations in NEW+TUNING and STD+TUNING, we represent in Fig. 5.7 the surface rain rate and cloud fraction in ARMCU, RICO and SANDU for one best simulation of the new version of LMDZ (NEW+TUNING, in blue), ten best simulations of the standard version of LMDZ (STD+TUNING, in black) and a reference LES (in green). For these best simulations, the cloud fractions predicted by the new version and the standard version of LMDZ after tuning are comparable in the three case studies considered. Significant differences are on the contrary observed in terms of surface rain rate. In STD+TUNING, there is no surface precipitation in ARMCU and RICO, whereas in SANDU, the precipitation tends to be overestimated at the beginning of the simulation and becomes null after a few hours of simulations, in contradiction with what is predicted by the reference

| Case | ARMCU | RICO | SANDU | SANDU | SANDU |
|-----------------------|-------|------|-------|-------|-------|
| Subcase | REF | REF | REF | SLOW | FAST |
| time | 7-9 | 7-9 | 19-25 | 50-60 | 50-60 |
| $\theta_{400-600hPa}$ | X | | | | |
| $q_{v,400-600hPa}$ | X | | | | |
| $f_{cld,max}$ | X | X | | | |
| $z_{cld,ave}$ | | | X | | |
| $z_{cld,max}$ | | | X | X | X |

Table 5.2: Metrics retained for the SCM/LES tuning. Inspired from Hourdin, Williamson, et al. (2020)

LES. These results suggest that in the standard version of LMDZ, it is not possible to have simultaneously realistic cloud fractions and realistic surface rain rates in these three case studies. They thus highlight a potential structural limitation of the standard version of LMDZ. On the contrary, in the new version, some simulations give surface rain rates compatible with those predicted by the LES in the three cases, as well as realistic cloud fractions, as illustrated by the blue curve (NEW+TUNING).

In terms of model results, in these three case studies, Kuhn’s criterion does apply: a relatively large part of the previous abilities of the model are conserved, but in addition, the new parameterization seems to solve certain problems in model results unsolvable with tuning alone – or at least, with the tuning protocol used, as a different tuning protocol might lead to different results. The new version of LMDZ and the standard version are further compared in Hwong et al. (2021) in SCM simulations under idealized radiative-convective equilibrium conditions forced with anomalous temperature and moisture tendencies. In this study, the new version of LMDZ (after tuning) also performed significantly better than the standard one. As we will see in the following, whether or not the new parameterization improves the model is more questionable when analyzing results of the global GCM.

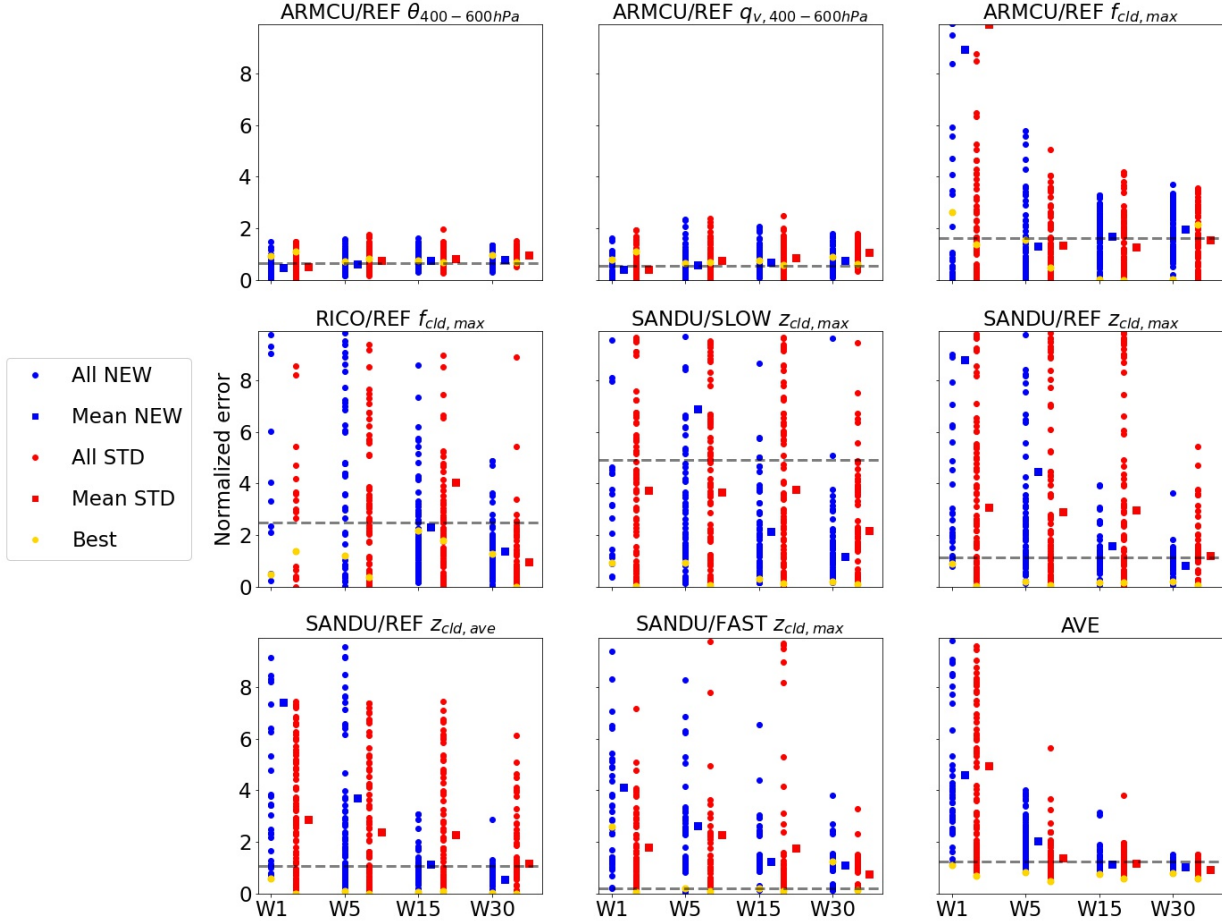


Figure 5.6: Normalized error after wave 1, 5, 15 and 30 in the tuning process with *htexplo* for the new version of LMDZ (in blue) and the standard version (in red). Each dot represents one simulation. The best simulation for each wave is the one with the lowest score averaged over all metrics (AVE) and is indicated with a gold dot. Blue and red squares give the mean error for all simulations of respectively the new and standard version of LMDZ for each metric and each wave. The dashed line represents the normalized error obtained with the standard version of LMDZ without tuning.

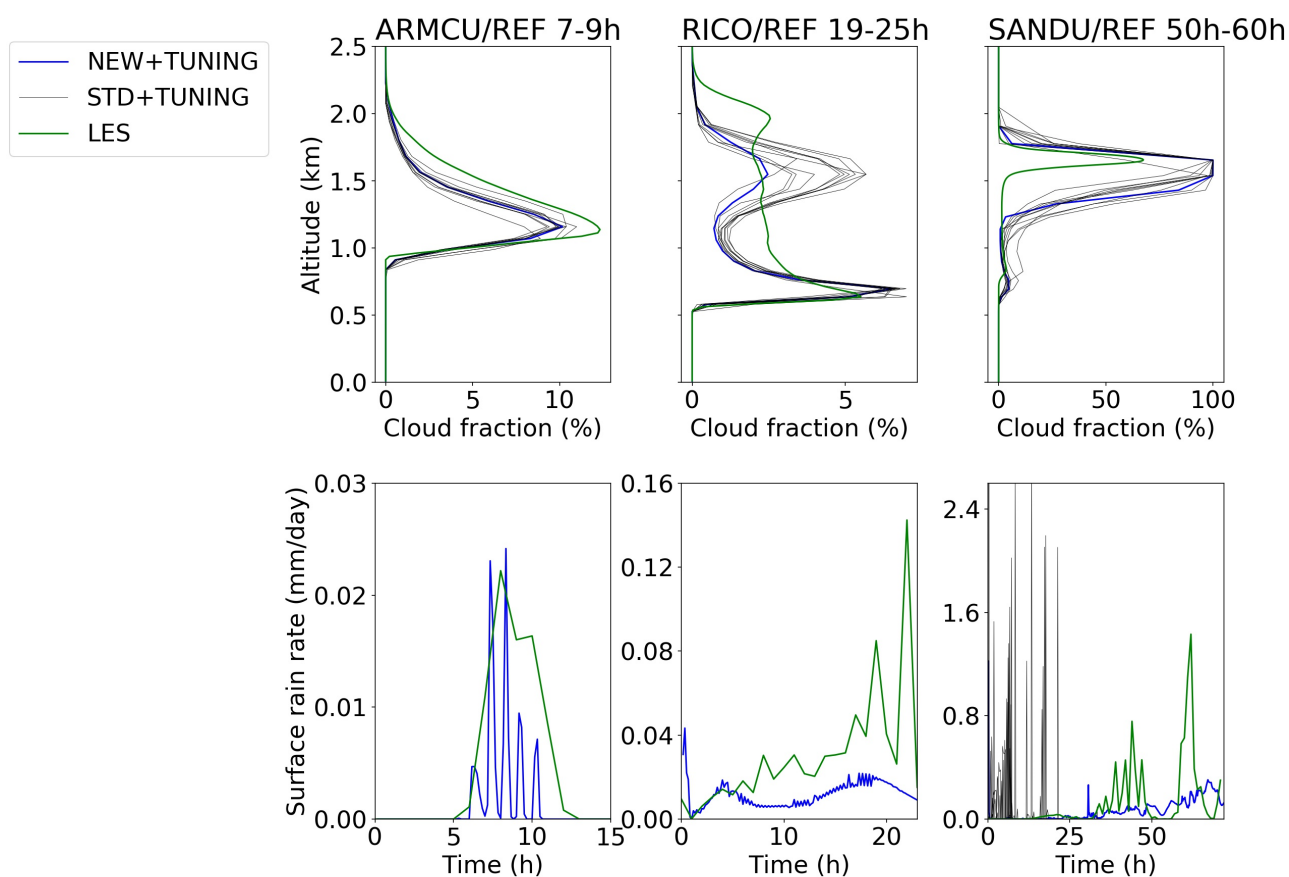


Figure 5.7: Cloud fraction (top) and surface rain rate (bottom) in ARM/REF, RICO/REF and SANDU/REF in one “best” simulation of the modified version of LMDZ (NEW+TUNING, blue) after 45 waves of tuning using *htexplo*, in ten best simulations using the standard version of LMDZ (STD+TUNING, black), and in a reference LES (green). The cloud fraction is averaged over the time interval specified for each case.

5.4.3 GCM tuning

As in Hourdin, Williamson, et al. (2020), for the global simulations, we use atmospheric simulations forced by SSTs and sea ice cover mean seasonal cycle. Simulations are run on the standard low resolution horizontal grid of LMDZ made up of 144 points in longitude and 143 in latitude, with a 95-level vertical grid. The parameters chosen are the same as the parameters used in the SCM tuning, and the initial NROY space for global simulations is the one obtained after 45 waves of SCM tuning. We perform two additional waves of 180 simulations each using the same SCM metrics considered for the 45 first waves, plus additional 3D metrics.

Table 5.3 lists the different 3D metrics used and the target values and tolerances to error for each metric. Most of the 3D metrics chosen target the shortwave (SW), longwave (LW) and net radiative fluxes at top-of-atmosphere computed in annual mean and averaged over spatial masks (see Fig. 5.8), using the CERES-EBAF L3b observational data set (Loeb et al., 2009) as reference. Three additional metrics are used, targeting the rain variability during the MJO (MJO), the frequency of daily precipitation above 50 mm/day ($PR > 50$) and the mean rainfall over the Sahel (AMMA) and using the Global Precipitation Climatology Project (GPCP) data set as reference (Adler et al., 2018). More details on the 3D protocol used can be found in Hourdin, Williamson, et al. (2020).

| | metrics | controls | target | error | unit | reference |
|--------------------|-------------|--|--------|-------------|----------------|---|
| Rad. metrics | glob.rt | | 1.577 | 0.25 | | |
| | glob.rlut | Top of atmosphere (TOA) | 239.7 | 25 | | |
| | conv.rsut | radiative fluxes: either total (rt), | 103.2 | 25 | | |
| | conv.rlut | upward shortwave (rsut), | 235.8 | 25 | | |
| | weak.rsut | upward longwave (rlut), | 81.8 | 25 | | |
| | weak.rlut | or shortwave cloud radiative effect (crest). | 264.3 | 25 | $W\ m^{-2}$ | CERES-EBAF L3b (Loeb et al., 2009) |
| | subs.rsut | The metrics are averaged | 84.9 | 25 | | |
| | subs.rlut | globally or over | 274.7 | 25 | | |
| | circAa.rsut | one of the masks | 23.2 | 25 | | |
| | circAa.rlut | shown in Fig. 5.8. | -44.1 | 25 | | |
| | etoa.crest | | -10.7 | 25 | | |
| Precip. metrics | AMMA | Mean rainfall over sahelian band | 1.16 | 0.04 | $mm\ day^{-1}$ | Global Precipitation Climatology Project (GPCP, Adler et al., 2018) |
| | MJO | Rainfall variability in the region of the Madden Julian Oscillation (MJO) | 3.33 | 1 | - | |
| | $PR > 50$ | Occurrence of precipitation rates superior to 50 mm/day | 0.34 | 0.1 | - | |

Table 5.3: Radiative and precipitation metrics used for GCM tuning at waves 46 and 47, along with the target value and tolerance to error for each metric. The reference data sets used to compute the target values for each metric are also given.

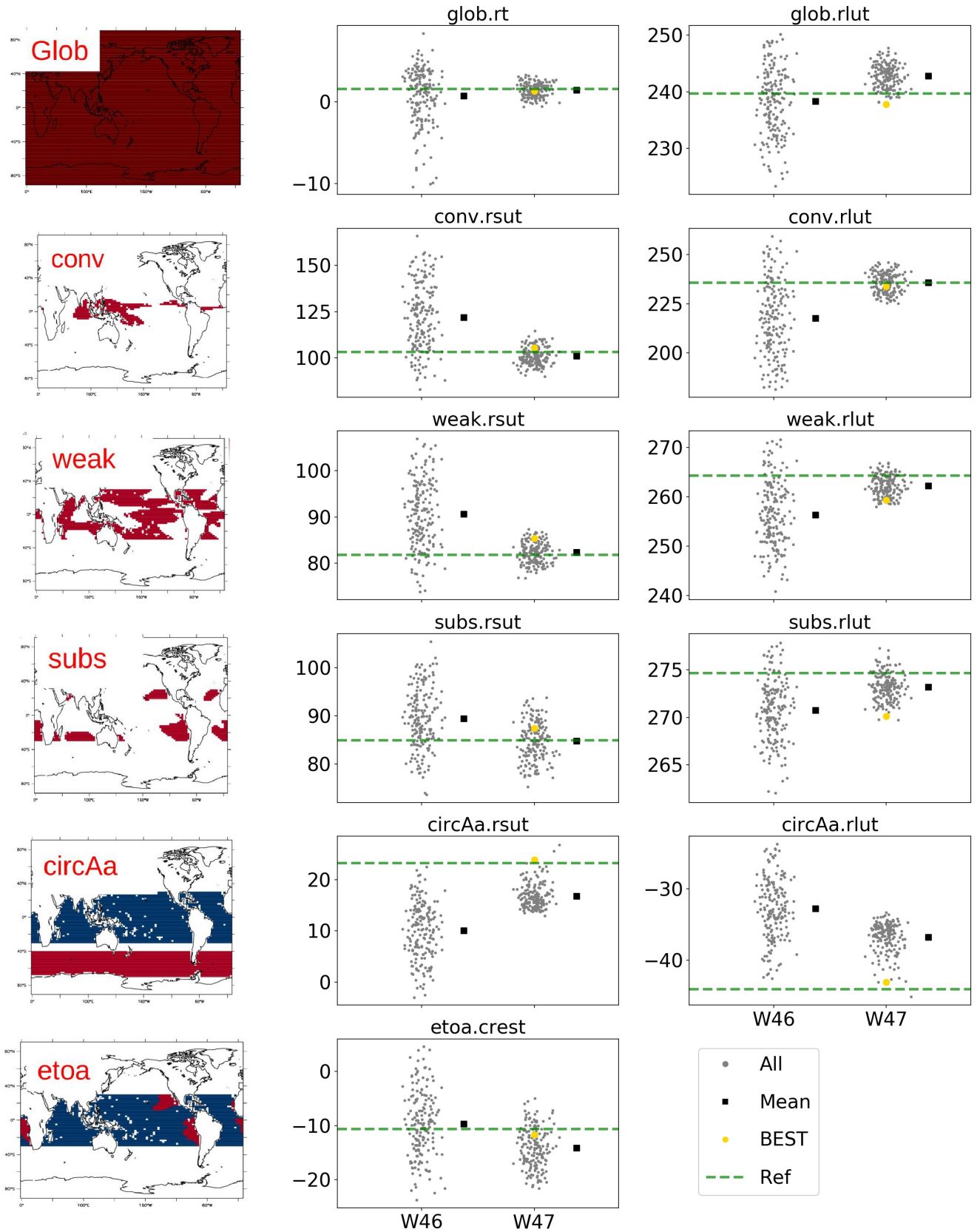


Figure 5.8: Radiative metrics used for the GCM tuning, targeting the total radiative flux (rt) at top of atmosphere, as well as the total upward shortwave (rsut) and longwave (rlut) fluxes at top of atmosphere. The metrics are averaged globally or over a mask (in red), or calculated as a difference between a red and a blue mask (anomalies). Grey dots represent the value of the different metrics for the 180 simulations wave 46 (W46) and 47 (W47), and the black squares, the average value of each metric for all simulations at each of the two waves. The golden points represent the values of the different metrics for the best simulation, as defined in Fig. 5.11.

In Chapter 4, we mentioned that simulating accurately the Earth’s energy budget in coupled GCMs was one of the main motivations of tuning. Here, this is done in practice by choosing an arbitrarily small tolerance to error in the metric that controls the total global radiative flux at top of atmosphere (TOA) (glob.rt) – 0.25 W m^{-2} instead of 25 W m^{-2} for the other radiative metrics. This tolerance to error is even smaller than the observational uncertainty of $\sim 4 \text{ W m}^{-2}$ given in Loeb et al. (2009). Moreover, the target value for tuning (1.5 W m^{-2}) is chosen to obtain a correct Earth’s energy budget in the coupled GCM, but is not consistent with the most recent estimate of 0.71 W m^{-2} from Johnson, Lyman, and Loeb (2016). This discrepancy is due to atmospheric responses to SST biases in the coupled GCM: in fact, the SSTs predicted by the coupled GCM do not match perfectly the present-day SSTs used as lower boundary condition in the atmospheric-only GCM. Due to these SST biases, it is necessary to tune the total energy budget of the atmospheric-only GCM to a value that satisfies the energy budget of the coupled model rather than to the value predicted by observations. This is the typical example where compensating errors are explicitly introduced in the model in order to guarantee key climate properties.

Figure 5.8 shows in grey dots the value of the different 3D radiative metrics considered for all simulations at waves 46 and 47. The green dashed line indicates the reference value given by observations for each metric. Wave 47 (W47) simulations are significantly closer to observations than those of wave 46 (W46). The effect of tuning is not as strong for the MJO, $PR > 50$ and AMMA metrics, shown in Fig. 5.9. For MJO and $PR > 50$, the black squares reveal on the contrary that W47 simulations are on average a little further from the observations than W46 simulations. This is due to the tuning protocol used, which gives preference to radiative metrics – and in particular to the total radiative flux glob.rt – over other metrics.

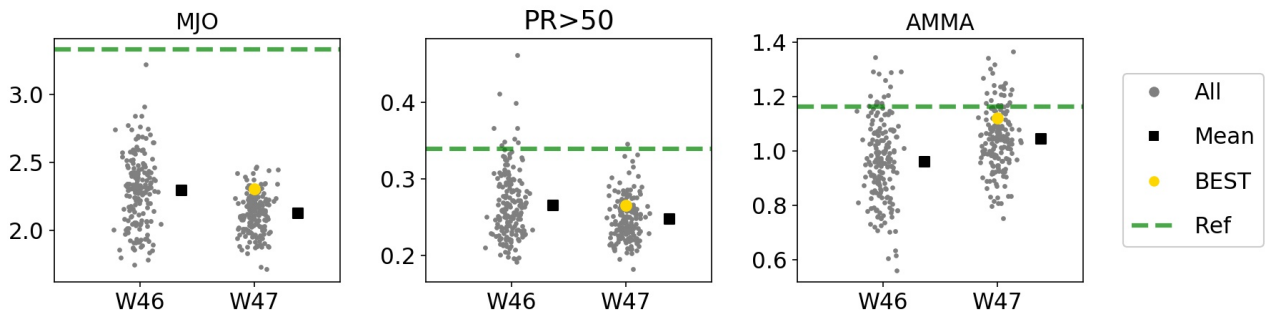


Figure 5.9: Same Figure as Fig. 5.8, but for precipitation metrics MJO, $PR > 50$ and AMMA.

As in Fig. 5.7, tolerances to error are used in *htexplo* to normalize the difference from the reference value. We define the BEST simulation as that with the lowest “score”, i.e. normalized error, averaged over the different metrics (1D+3D). The value of the different metrics for this BEST simulation is indicated by a golden point in Fig. 5.8 and 5.9.

The many ways to compare GCM results with observations

Simulating a correct Earth’s energy budget is crucial to avoid a drift of the model climate that would make it very far from the current climate and explains why we used glob.rt as a main tuning target. Nevertheless, a correct radiative budget is not the only criterion to

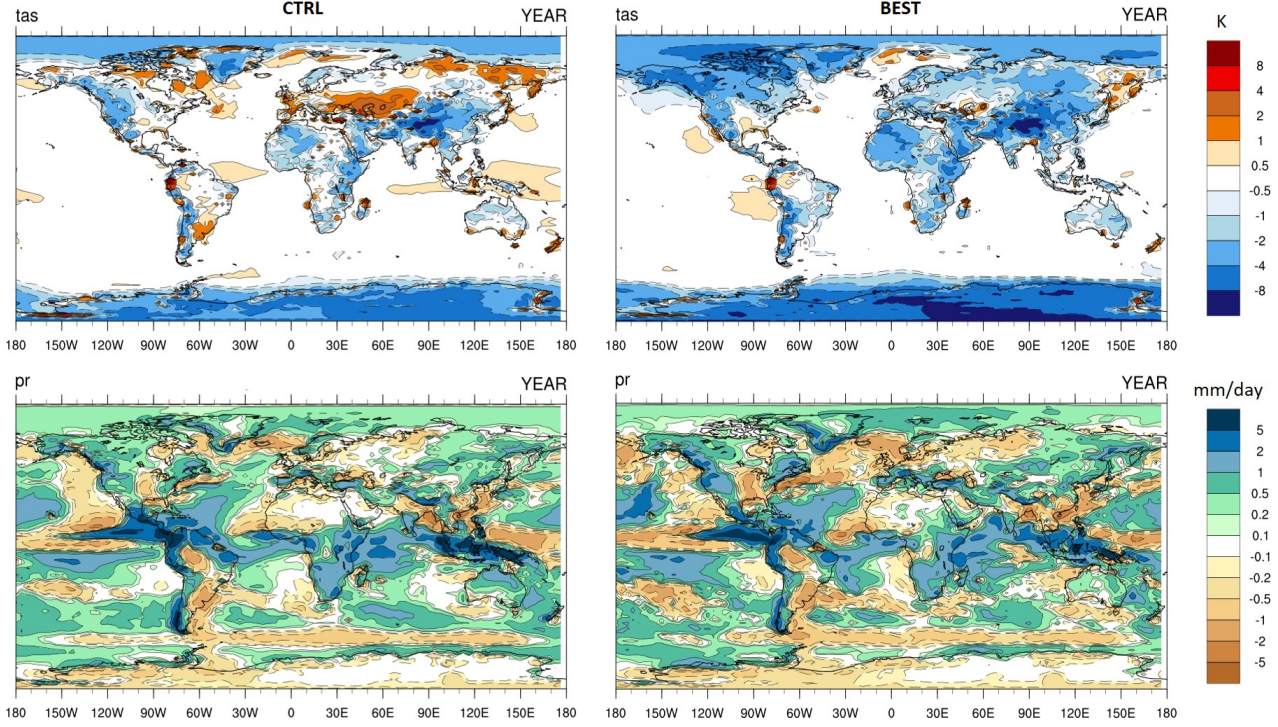


Figure 5.10: Biases in surface precipitation (pr, top) and temperature (tas, bottom) averaged over a year in the standard version of LMDZ (CTRL) and the best simulation of the second wave of GCM tuning (BEST). The observational data sets used as references are the one provided by ERA Interim for surface temperature (Dee et al., 2011) and by GPCP (Adler et al., 2018) for surface precipitation.

consider to evaluate the quality of a GCM simulation. Many other variables can be used, such as surface temperature, surface precipitation, the 500-hPa geopotential height, etc. There are different methods for comparing each of these variables to observations. In climate models, we are more interested in statistics in space and time than in local and instantaneous values. A choice has to be made on the time and space intervals chosen to calculate averages or other statistical values. Moreover, for some variables, the patterns present in the observations are also present in the model, but slightly shifted. Despite the presence of realistic patterns in the model, a dot by dot comparison between model and observations would attribute a poor performance to the model. To solve this issue, different tools have been developed, such as the use of spectral empirical functions (EOF, see Pritchard & Somerville, 2009; O. T. Schmidt, Mengaldo, Balsamo, & Wedi, 2019). With all these variables and tools, the ways in which GCM results and observations can be compared are virtually endless. This explains the difficulties in evaluating whether one version of a model is better than another in terms of results: it all depends on what is being compared.

In Fig. 5.10, we use the surface temperature (tas) and precipitation (pr) biases to compare the standard version of LMDZ (CTRL) and this BEST simulation. To calculate these biases, we averaged the surface temperature and precipitation given by the model and by observations over a year. The two simulations shown have many similarities: they share for instance positive biases in precipitation over the Eastern Tropical Pacific and over Indonesia, and negative

biases of temperature over Antarctica. However, some biases seem more pronounced in one simulation or the other. For instance, the simulation BEST seems to have a more pronounced global cold temperature bias over the continents than CTRL, but more realistic precipitations over Indonesia.

Even if we were to consider only these two variables – the surface temperature and precipitation averaged over a year – it is not possible to objectively determine which simulation is more realistic. Kuhn’s criteria does not apply: in global simulations, previous model abilities are not preserved. Improvements in certain regions are accompanied by deteriorations in others, which makes the evaluation of GCM simulations difficult. The question “is the CTRL simulation more realistic than the BEST simulation?” involves subjective choices on which aspects of the results are considered most important.

To make these choices explicit, the solution adopted by the scientific community to evaluate the performance of GCMs consists in using metrics. Metrics can first be used to assess whether a simulation S_1 is better than another S_2 for one aspect of the model X . The question is then not “which simulation is the best” but rather “which simulation is the best at predicting X ”. Knutti et al. (2017) make this choice when analyzing CMIP multi-model ensembles. They weight climate model projections on one particular aspect (e.g. the future evolution of Arctic sea ice) by taking into account the past performance of models on selected variables (e.g. their ability to simulate accurately Arctic sea ice decline in the twentieth century). Similarly, in the GCM tuning process, *htexplo* uses metrics to target essentially one (important) aspect of the system: radiative fluxes at the top of the atmosphere.

An attempt to define the comparison of two GCM simulations

To compare exhaustively two GCM simulations, a more ambitious use of metrics has to be done: metrics have to be chosen in order to define the overall quality of a GCM simulation. Due to the number of possible metrics, this choice will be partly arbitrary and express the personal preferences of the climate modeler. Different weights may be assigned to different metrics and the assignment of these weights will also be partly subjective. Nevertheless, metrics and weights allow to compare quantitatively two simulations according to these subjective preferences. They give some criteria to determine whether a global simulation S_1 is better than another S_2 or not. More precisely, we formulate the following proposition to compare two simulations S_1 and S_2 .

Proposition 1: Under an ensemble of n metrics $(p_i)_{i \leq n}$ and associated weights $(m_i)_{i \leq n}$, we consider that a simulation S_1 is better than another S_2 if and only if the weighted average of the normalized errors of S_1 $(m_i[S_1])_{i \leq n}$ under the defined metric is smaller than the weighted average of the normalized errors of S_2 $(m_i[S_2])_{i \leq n}$ under the same metrics. That is, mathematically:

$$S_1 > S_2 \Leftrightarrow \sum_{i \leq n} p_i m_i[S_1] < \sum_{i \leq n} p_i m_i[S_2] \quad (5.14)$$

In formula 5.14, different methods can be used to normalize the errors: in our example, *htexplo* divides them by the tolerances to error defined in Table 5.3. We apply Proposition

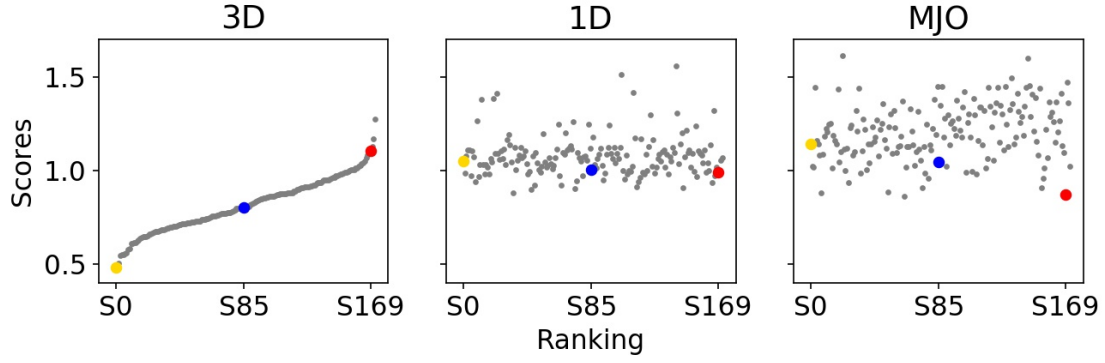


Figure 5.11: Scores, i.e. error from observations divided by tolerances to error defined in 5.3, in 3D metrics, 1D metrics and MJO metric for the 180 simulations of W47. The x-axis ranks simulations according to their 3D score. In each panel, the *best* simulation in terms of 3D score (S0) is indicated in gold, and one *intermediate* and one *worse* simulations in terms of 3D score are shown in blue and red respectively.

1 to compare the simulations run during W47. Three different ways of comparing these 180 simulations are shown in Fig. 5.11. 3D refers to the average of the normalized errors of all 3D metrics (radiative + precipitation), 1D of all SCM metrics (defined in 5.2), and MJO of the MJO metric only. The x-axis ranks the simulations according to their 3D score, and the y-axis gives the value of their score according to each formula (3D, 1D or MJO). The 3D score is further used to define a *best* simulation (S0, in gold), an *intermediate* simulation (S85, in blue) and a *worse* simulation (S169, in red). Contrary to their 3D score, the 1D score of these three simulation is comparable, and the ranking of these three simulations is reversed for the MJO metric. This illustrates that there is no absolute ranking of GCM simulations, but that the ranking is conditioned by the metrics and weights used when applying formula 5.14.

An attempt to define the comparison of two versions of a GCM

Despite its dependence on the chosen metrics and weights, formula 5.14 gives a quantitative method to compare two simulations. What about the comparison of two GCMs or, in our case, two versions of a GCM containing different parameterizations? The fact that one particular simulation of a model version V_1 is better than one particular simulation of another version V_2 does not prove that V_1 is better than V_2 . It is illustrated in Fig. 5.11, where some simulations of the new version of LMDZ might be better than the standard version, and some worse depending on the metrics and weights considered. This remark may sound trivial but in scientific papers, it has always been the rule rather than the exception to compare two particular simulations to evaluate the progress brought by a new parameterization (e.g. Rasch & Kristjánsson, 1998; Yoo, Park, Kim, Yoon, & Kim, 2015). *Htexplo* opens the door to another definition, that we give in the following.

Proposition 2: Let us call V_1 a particular version of a GCM and $S_1 = V_1(P_1)$ a simulation given by this version under a set of parameters P_1 in the parameter space Ω_1 . We will consider that V_1 is better than another version V_2 if and only if there is a set of parameters $P_{1,0}$ in Ω_1 such that $S_{1,0} = V_1(P_{1,0})$ is better than any simulation $S_2 = V_2(P_2)$, with P_2

belonging to the parameter space Ω_2 . Mathematically, we can write:

$$V_1 > V_2 \Leftrightarrow \exists P_{1,0} \in \Omega_1, \forall P_2 \in \Omega_2, V_1(P_{1,0}) > V_2(P_2) \quad (5.15)$$

that is using eq. 5.14:

$$V_1 > V_2 \Leftrightarrow \exists P_{1,0} \in \Omega_1, \forall P_2 \in \Omega_2, \sum_i p_i m_i[V_1(P_{1,0})] < \sum_i p_i m_i[V_2(P_2)] \quad (5.16)$$

When comparing two versions of a model, Proposition 2 distinguishes the differences in results due to the tuning of each version from those due to differences in their parameterizations. Conceptually, it clarifies what we aim to do when we evaluate the progress brought by a new parameterization to a given climate model. Nevertheless, it raises two main practical difficulties:

1. It requires first the existence of **a tuning protocol that explores extensively the space of parameters**. Due to the number of parameters involved in GCMs and the many values they can take, it is impossible in practice. The choice of parameters and parameter values used in the tuning process will always be in part arbitrary. A given tuning protocol might be more favorable to some versions of the model than to others. Therefore, some differences between two versions of a model after re-tuning may be due to the fact that the tuning protocol used is more adequate for one version of the model than for the other. We faced this problem at the beginning of our study: initially, we did not use **CLTAU** as a tuning parameter, and we were not able to find SCM results with the new version of LMDZ comparable with those given by the standard version, when exploring the space of parameters with *htexplo*. This problem was solved when **CLTAU** was used with appropriate range of values in the tuning process.
2. The second practical difficulty lies in the **definition of metrics and weights given to each metrics**. There is no consensus yet on the most relevant metrics to evaluate the performance of climate models. Yet, as we have seen in Fig. 5.11, the choice of metrics and associated weights determines the relative performance of two simulations, thus of two versions of a model.

Due to these practical difficulties, we will not use formula 5.16 in practice to compare the standard and the new version of LMDZ. However, it inspires our approach in what follows. We identify two differences in the new version of LMDZ compared to the standard version that do not appear to be due to differences in tuning, but rather to differences in parameterizations. These two properties are indeed shared by many simulations of W47 in the new version of LMDZ and reduce biases that were resistant to tuning in the standard version of the model.

Reduction of excessive rainfall rates over Indonesia and over the northeast coast of South America In the standard version of LMDZ, very high average rain rates are observed over Indonesia and over the northeast coast of South America. The purple patches in Fig. 5.12 middle panel (CTRL) show that the rain rates exceed 14 mm/day in average at these locations. On the contrary, in observations, the rain rates are always lower than 10 mm/day. The new parameterization seems to significantly reduce this bias, even if it does not entirely correct it. Indeed, we see in Fig. 5.12 that the purple patches are reduced in BEST compared to CTRL, which means that the rain rates over Indonesia and the northeast coast of South America are

significantly lower in BEST than in CTRL (even if they are still too high compared to observations).

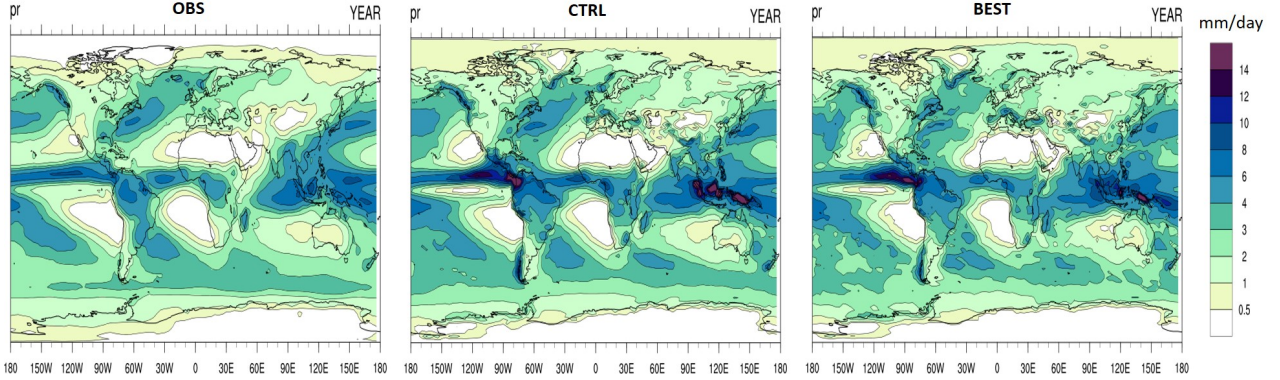


Figure 5.12: Surface precipitation in observations (OBS), in the standard version of LMDZ (CTRL) and in the BEST simulation of the new version. As in Fig. 5.10, precipitation observations are from GPCP (Adler et al., 2018). Precipitation in CTRL and BEST is averaged over a year of simulation.

This result was expected given the effects of the parameterization in SCM case studies. The SANDU case suggested indeed that the new version tended to reduce high precipitation rates compared to the standard one (see Fig. 5.7). Admittedly, the SANDU case is a case of shallow clouds containing only liquid precipitation, while the precipitation over Indonesia and the northeast coast of South America is certainly due in part to deeper convective clouds forming solid precipitation. However, the cloud and precipitation overlap parameterization also affects solid precipitation from deep convection. The decrease in surface precipitation rate over Indonesia and over the northeast coast of South America suggests that the effect of the parameterization on deep clouds is similar to what we observed in the SANDU stratocumulus case: it tends to reduce extreme precipitation rates.

Correction of summer continental warm bias over mid-latitudes Contrary to the first one, this second bias correction was not expected. In the standard version of LMDZ, the temperature at the surface reveals a strong summer continental warm bias in mid-latitudes. This bias was documented in Boucher et al. (2020); Cheruy et al. (2020) and Hourdin, Rio, et al. (2020). In the new version, as can be shown in Fig. 5.13, this bias is significantly reduced. We show here only the BEST simulation but this property is shared by the different simulations of W47. Simulations of the new version of LMDZ are significantly colder in summer over continents than the standard version of LMDZ.

How to interpret this behavior? In the ARMCU and RICO case studies, we noted that the new parameterization increases the frequency of light precipitation in the model. The coupling between this light precipitation and the surface certainly exerts an influence on surface temperatures and could explain the differences observed. Further research is underway to test this hypothesis.

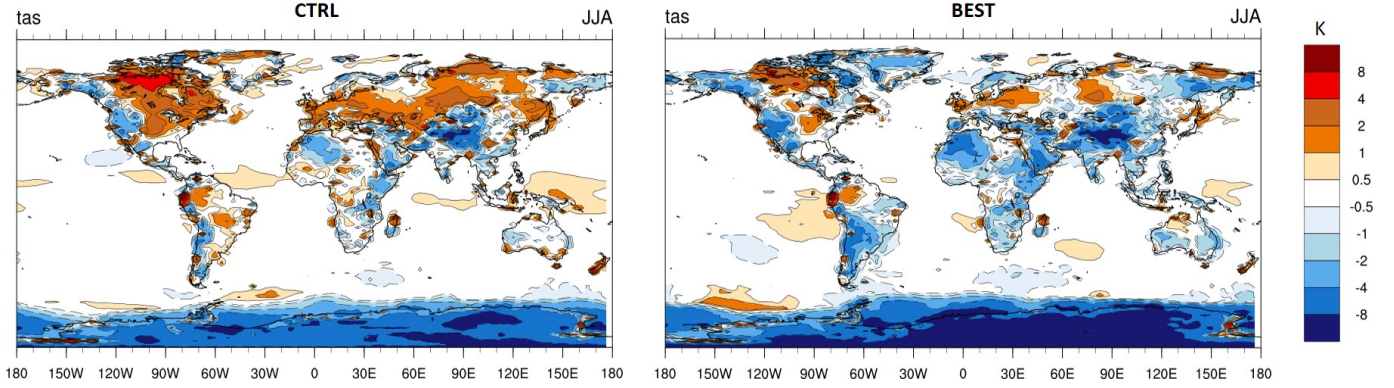


Figure 5.13: Biases in surface temperature (*tas*) averaged over a year in the standard version of LMDZ (CTRL) and in the BEST simulation of the new version. As in Fig. 5.10, ERA Interim is used as reference for the surface temperature (Dee et al., 2011).

5.5 Conclusion

This chapter is at the frontier between climate modeling and epistemology. On the one hand, we presented the scientific results we obtained by adaptating the parameterization of Jakob and Klein (2000) to LMDZ. We detailed the process we used to re-tune LMDZ following the addition of this parameterization using the tool *htexplo*. This was one of the first times *htexplo* was used to tune a GCM following the addition of a new parameterization – this test was successful in many ways. In SCM case studies, after 45 waves, *htexplo* reduced the parameter space sufficiently so that the simulations in the version of LMDZ with the new parameterization were of comparable or even better quality than those of the standard version of LMDZ. Then, this parameter space was used to perform the 3D tuning of the model, through atmospheric GCM simulations forced by SSTs and sea ice cover mean seasonal cycle. Two additional 3D tuning waves were conducted. Simulations obtained at wave 47 (after 45 waves in SCM + 2 waves in 3D) performed well for most of the global metrics considered. Furthermore, two important biases seem to be reduced in the simulations of the new version compared to the standard one: excessive rain rates over Indonesia and the northeast coast of South America, and a summer continental warm bias at mid-latitudes. These results strongly support the use of the parameterization proposed in future reference versions of LMDZ. They also illustrate the potential of automatic tuning methods such as *htexplo* to guide the development and testing of new parameterizations in GCMs.

On the other hand, and in parallel to these scientific results, we conducted an epistemological reflection throughout the chapter. In particular, we questioned what it means to improve a GCM and how progress can be defined for parameterizations. We first distinguished conceptual progress brought by a new parameterization from progress in terms of model results. We gave four criteria to evaluate conceptual progress – what we called the CISC criteria for Coherence, Interpretability, Simplicity and Comprehensiveness – and applied these criteria to the cloud and precipitation overlap parameterization implemented in LMDZ. This led us to conclude that this parameterization made significant conceptual progress in LMDZ. We then explored the meaning of progress for model results and identified two main difficulties that arise when one tries to assess the progress made by a parameterization on the results of a GCM. First, the results of a GCM depend largely on its tuning – so to compare two versions of a GCM, we need to distinguish differences that are due to tuning from those that are not. Second, there are

many ways to evaluate the results of a GCM and progress in some results is often associated with degradation in others. We argued that metrics can be used to objectively define which model results should be prioritized and proposed a definition of the comparison of two versions of a GCM independently of their tuning once a set of metrics and associated weights is defined (Prop. 2). Although this definition is difficult to apply in practice, it guided us in comparing the version of LMDZ containing the new parameterization and the standard version of the model.

Our epistemological work shows that the benefits of using automatic tuning methods such as *htexplo* should not be underestimated. By making the tuning process more standardized and reproducible, *htexplo* not only helps climate modelers to save the tedious and time-consuming process of re-tuning the model manually when a new parameterization is added. It also changes the way we look at GCMs. As we have seen in this chapter, a GCM is no longer one particular simulation, but an ensemble of plausible simulations defined by a parameter space and some structural limitations within this parameter space. The meaning of “improving a GCM” is thus questioned. Work remains to be done to clearly define metrics and a tuning protocol that rigorously assess whether a new parameterization is better than another or not. But with *htexplo*, there is hope that in the near future, GCM improvement will rest on solid epistemological grounds – even if the associated “progress” will probably have a different meaning than that given by Kuhn.

Chapter 6

The Coupled Model Intercomparison Project – History, uses, and structural effects on climate research

Contents

| | | |
|------------|--|------------|
| 6.1 | Preamble | 147 |
| 6.2 | Abstract | 148 |
| 6.3 | Introduction | 148 |
| 6.4 | History of climate model intercomparison projects | 149 |
| 6.4.1 | AMIP and the genesis of climate model intercomparison projects | 149 |
| 6.4.2 | The beginnings of CMIP – CMIP 1, 2 and 2+ | 151 |
| 6.4.3 | CMIP3 – A new era in climate change research | 152 |
| 6.4.4 | CMIP5 and 6 – the most recent phases of CMIP | 153 |
| 6.5 | Uses of CMIP results and related controversies | 154 |
| 6.5.1 | Exploring future climate change and the associated uncertainties | 155 |
| 6.5.2 | Comparing CMIP simulations with observations | 157 |
| 6.5.3 | Interpreting model results | 158 |
| 6.6 | Structural effects of CMIP on climate research | 160 |
| 6.7 | Conclusion | 162 |

6.1 Preamble

Over the past 30 years, projects to compare the results of different GCMs have become increasingly important in climate science research. In particular, the Coupled Model Intercomparison Project (CMIP) is now a reference tool for comparing simulations of coupled GCMs under different conditions. Studies based on CMIP results provide a convenient way to summarize what the most comprehensive GCMs have to say about anthropogenic climate change and logically play a key role in the IPCC reports. Yet, each phase of CMIP required significant human and computing resources from modeling groups. At the time of writing, about 50 modeling groups

participated in the last phase of CMIP, which represents almost all of the major climate modeling groups in the world. With the progressive appearance of GCRMs and other competitors to parameterizations and GCMs, we may wonder whether such an endorsement will continue in the next phases of CMIP. The purpose of this chapter is to conduct a historical and epistemological study that might inform the debate on the future of CMIP. In particular, we assess the role that CMIP has played historically, how CMIP results are analyzed today, and which structural effects CMIP has had on climate research.

This chapter is based on a paper published in *Wiley Interdisciplinary Reviews: Climate Change* in April 2020. The full reference is: Touzé-Peiffer, Ludovic, Anouk Barberousse, and Hervé Le Treut. “The Coupled Model Intercomparison Project: History, uses, and structural effects on climate research.” *Wiley Interdisciplinary Reviews: Climate Change* 11.4 (2020): e648.

6.2 Abstract

The results of the sixth phase of the Coupled Model Intercomparison Project (CMIP) are currently being analysed and will form the basis of the IPCC Sixth Assessment Report. Since its creation in the mid-1990s, CMIP has had an increasing influence on climate research. While the principle behind it has always remained the same – comparing different climate models under similar conditions – its design and motivations have evolved significantly over the phases of the project. This evolution is closely linked to the one of the IPCC, since, historically as well as today, the results of CMIP have played a major role in the panel reports. This role increased the visibility of CMIP – over time, more and more people started to be interested in CMIP and to analyze its results. Despite this success, the way CMIP is used today raises methodological issues. In fact, CMIP has promoted a particular way of doing climate research, centered on a single tool, Global Coupled Models (GCMs), and creating a gap between model developers and model users. Due to the debates regarding the interpretation of multi-model ensembles and the validation of GCMs, whether the emphasis on this particular way of studying climate is serving the progress of climate science is questionable.

6.3 Introduction

Intercomparison projects used in climate science are based on a simple idea: run a set of numerical climate models under the same conditions and compare their results. The Atmospheric Model Intercomparison Project (AMIP), which started in 1990, was the first attempt to coordinate such an activity. Different intercomparison projects have followed: the Coupled Model Intercomparison Project (CMIP), in particular, is now considered as “one of the foundational elements of climate science” (Eyring et al., 2016, p. 1937). Since its creation in the mid-1990s, it has evolved over five phases, involving all major climate modeling groups in the world. In addition to their role in climate research, these phases have held a central place in the reports of the Intergovernmental Panel on Climate Change (IPCC). For the expert panel, whose objective is to synthesize the state of current knowledge about climate and climate change, CMIP results were a goldmine – they provided an easy way to have a summary of what the most comprehensive climate models had to say about these questions. The figures based on their results were also perfect illustrations for the Summary for Decision Makers accompanying each IPCC report – for instance, in the fourth Assessment Report (AR4), among the 7 figures of the

Summary for Decision Makers, 4 were based on CMIP results.

If CMIP has been a great opportunity for the IPCC, its role for climate research is more controversial. CMIP has indeed favored the use of one type of model, Global Coupled Models, at the expense of other simpler, less comprehensive models. Climate science is not the only area of environmental science to look at complex systems – but it is probably the only one where the most comprehensive models have such a supremacy (Shackley, Young, Parkinson, & Wynne, 1998). Ecologists, biologists or economists all use models of varying complexity (Gabaix & Laibson, 2008; Jørgensen, 2008). The supremacy of GCMs in climate science raises questions a fortiori when we look at the numerous issues surrounding the validation of GCMs and the interpretation of their results (e.g. Knutti, Furrer, Tebaldi, Cermak, & Meehl, 2010; W. S. Parker, 2011; Shackley et al., 1998). The problem gets worse when looking at the particular use of GCMs CMIP has promoted – by making the outputs of its simulations freely available, CMIP has created a growing gap between model developers and model users. It has promoted the analyses considering GCMs as black boxes and made it extremely difficult for these analyses to be relevant for the improvement of climate models.

Our paper aims at tackling these different issues by providing an overview of the historical evolution of CMIP, what it has become today, and the effects it has had on climate research. We will start by analyzing how the project and the motivations behind it have evolved over time. We will see that part of this evolution comes from the intertwining between the history of CMIP and that of the IPCC. In a second part, we will study a set of papers based on the results of the fifth phase of CMIP (CMIP5) in order to distinguish three main uses of CMIP results. We will see that each of them raises methodological issues. Finally, borrowing a concept from Pitt (2000), we will define CMIP as an infrastructure for climate science and we will explore the various ways in which CMIP, as an infrastructure, has shaped climate research. Considering these different effects altogether, we will finally question the value of CMIP on climate research as a whole.

6.4 History of climate model intercomparison projects

6.4.1 AMIP and the genesis of climate model intercomparison projects

The genesis of climate model intercomparison projects lies in numerical weather forecasting. In the 1970s, intercomparison projects between numerical models used for weather predictions were pushed by the Global Atmospheric Research Program (GARP) - a program established in 1967 with the goal of coordinating weather and climate research internationally. One of the first decisions of the Joint Organizing Committee leading the program was to create a Working Group on Numerical Experimentation (WGNE), in order to promote a dialogue between modeling groups (Gates, 2015). The WGNE aimed in particular to facilitate intercomparison projects between atmospheric models used either for weather forecasting or for climate research.

At first, such intercomparison projects concerned mainly atmospheric models used for operational weather prediction. To compare different models or different versions of the same model under similar conditions, intercomparison protocols were designed and led to the publication of several papers in the 1970s (e.g. Baumhefner & Downey, 1978; D. D. Houghton & Irvine, 1976). Comparatively, fewer intercomparison projects emerged in climate research. One of the first example of juxtapositions of results from different atmospheric climate models was provided

in 1975 by the report of the U.S. GARP Committee's Panel on Climatic Variation (National Academy of Sciences, 1975). This report used the simulations available at that time to deduce common diagnostics between models. Even if it did not design any protocol to compare climate models under similar conditions, it inspired further research on the subject (Gates, 1992).

The first climate model intercomparison protocols appeared in the following years with a series of stand-alone initiatives. Among others, we can cite the Intercomparison of radiation codes in climate models (ICRCCM) workshop, which led to the publication of a paper comparing atmospheric radiative processes in climate models of varying complexity (Luther et al., 1988). Another example is given in Potter and Gates (1984), where the seasonal response of two atmospheric models is compared.

Until the end of the 80s, these different initiatives remained isolated. It led the Joint Scientific Committee (JSC) of the World Climate Research Program (WCRP) to call for a systematic and comprehensive intercomparison of atmospheric climate models. Since such an intercomparison would be based on a series of expensive simulations, having access to powerful computing facilities was necessary for the emergence of the project. Under the aegis of W. Lawrence Gates, the JSC successfully convinced the U.S. Department of Energy to provide the computer facilities of the Lawrence Livermore National Laboratory (LLNL) to support the project. It led to the creation of the Program for Climate Model Diagnosis and Intercomparison (PCMDI) at LLNL in 1989, with the official goal of “increasing understanding of the differences among climate models” (Gates, 1992, p. 1963). One year later, in 1990, the Atmospheric Model Intercomparison Project (AMIP) – the first major experiment of this program – was officially endorsed by the JSC.

AMIP was designed to compare the response of atmospheric general circulation models (GCMs) – the models simulating the behavior of the atmosphere at a global scale – on seasonal and interannual time scales. All atmospheric GCMs could participate. The decade 1979-88 was chosen as the simulation period, and the protocol imposed to all models specific boundary conditions – more precisely standardized values for the solar constant and atmospheric CO₂ concentrations, as well as observed mean sea surface temperature and sea-ice distributions. Moreover, the outputs of the simulations had to be given in a standard format (Gates, 1992).

There were officially two main motivations behind AMIP. The first one was to “undertake the systematic intercomparison and validation of the performance of atmospheric GCMs on seasonal and interannual time scales under as realistic conditions as possible” (Gates, 1992, p. 1963). In other words, AMIP aimed first at identifying output differences between atmospheric GCMs under the same protocol, and comparing them with observations in order to validate the performance of these models. However, the original ambition of AMIP was also “to support the in-depth diagnosis and interpretation of the model results” (Gates, 1992, p. 1963), that is to say, not only to find out what the differences between atmospheric GCMs were, but also to understand them. This second motivation led to the organisation of 26 diagnostic subprojects to analyze AMIP outputs in the years following the project (Gates et al., 1999, Appendix A).

AMIP was a major step for climate modeling – thanks to it, climate modelers had for the first time access to an institutional structure to compare and evaluate the performance of their model under similar conditions. The experiment became quickly “the most prominent international effort devoted to the diagnosis, validation, and intercomparison of global atmospheric models’

ability to simulate climate” (Gates et al., 1999): by 1995, 31 modeling groups had taken part in the experiment, representing almost the entire atmospheric modeling community. This massive endorsement paved the way to subsequent intercomparison projects, to the Coupled Model Intercomparison Project (CMIP) in particular.

6.4.2 The beginnings of CMIP – CMIP 1, 2 and 2+

As early as the late 1960s, it had been recognized that ocean played a key role in climate, and attempts had been made to couple atmospheric and ocean GCMs (Manabe & Bryan, 1969). However, such coupled models were complex, and required a high computational capacity. Thus, it is only in the 1980s that coupled models started to be developed in more and more laboratories to represent the dynamic interactions between the atmosphere, ocean and cryosphere. As the most comprehensive models, they were sometimes seen as “potentially the most useful tools in simulating global climate, studying present-day climate fluctuations and addressing the problem of anthropogenic climate change” (G. A. Meehl, 1995).

Therefore, at the end of the 80s, even if they were still facing strong uncertainties and systematic errors, some global coupled models have been used to study the impact of an increase of anthropogenic CO₂ on climate (J. Houghton, Jenkins, & Ephraums, 1990). For instance, *the Supplementary Report to the IPCC Scientific Assessment* (J. Houghton, Callander, & Varney, 1992) compared the temperature rise associated with a transient CO₂ doubling in four different coupled models. The report recognized the limits associated with this set of simulations, but still used them to confirm its statements about the rise of temperature associated with a CO₂ increase in the atmosphere (J. Houghton et al., 1992).

As a consequence, CMIP was born at the confluence of two influences. On one side, AMIP had shown the potential of intercomparison projects to coordinate and organize research around atmospheric GCMs. Therefore, it was tempting to organize a similar project for coupled GCMs. On the other side, more simulations from coupled models were needed to make statements about anthropogenic climate change more robust.

The original structure of CMIP reflected these two motivations. In fact, CMIP was initially divided into two complementary phases (G. A. Meehl, Boer, Covey, Latif, & Stouffer, 1997):

1. CMIP1, which started in 1996, and transposed the main objectives of AMIP – measuring and understanding the ability of atmospheric GCMs to simulate current climate – to coupled models.
2. CMIP2, which compared climate change simulated by coupled models under a 1% per year CO₂ increase. Starting in 1997, this second phase was thus directly in line with the comparison performed in *the Supplementary Report to the IPCC Scientific Assessment* (J. Houghton et al., 1992).

Due to limitations in data processing and archiving capabilities at that time, CMIP1 and CMIP2 included only a few output fields, and at a coarse temporal resolution: for example, surface temperature, precipitation, and sea level pressure were averaged over one month. This rough sample was a strong limitation for analyses based on experiments. Consequently, the Working Group on Coupled Modeling (WGCM) – a subgroup of the World Climate Research Programme (WCRP) playing a leading role in CMIP – launched a new phase, CMIP2+, in 1999

to include many more model fields, and daily data if possible. However, this new phase represented significant additional work for the modeling groups, and in the end CMIP2+ was able to collect only 12 complete sets of outputs (G. Meehl, Covey, McAvaney, Latif, & Ronald, 2005).

CMIP1 and 2 – but not CMIP2+, which was not completed in time – played a substantial role in the Third Assessment Report (TAR) of the IPCC. In particular, detailed analyses of CMIP models were presented in the chapters “Model Evaluation” (McAvaney & et al., 2001) and “Projections of Future Climate Change” (Cubasch & et al., 2001). In return, the IPCC Third Assessment Report made some recommendations about future phases of CMIP. In particular, it called for: “GCM simulations with a greater range of forcing scenarios and an increased ensemble size to assess the spread of regional predictions” (Giorgi & et al., 2001, p. 586). Taking into account these recommendations, the WGCM decided to design a new phase of CMIP.

6.4.3 CMIP3 – A new era in climate change research

A major novelty of CMIP3, compared to previous phases of CMIP, was to include climate change scenario experiments, that is to say projections of future climate change under different emission scenarios. Such simulations are of great interest for decision makers, because they can be used as a basis for choosing between different mitigation and adaptation strategies.

Scenarios had already been used in the IPCC First Assessment Report. However, the computational capacity available at that time did not make it possible to run these scenarios with GCMs; instead, very simplified models called “box-diffusion models”, which were thought to give the same results as GCMs when globally averaged, had been used (J. Houghton et al., 1990).

In its subsequent reports, the IPCC continued to steer climate research around emission scenarios. In particular, in the preparation of the Third Assessment Report (TAR), it produced a Special Report on Emission Scenarios (SRES) (Nakicenovic & Swart, 2000). With its set of 40 scenarios, this report aimed to cover a wide range of assumptions about the main demographic, economic, and technological driving forces of future greenhouse gas and sulfur emissions. The idea was originally that the climate modeling community would use these scenarios in coupled model simulations which could figure in the TAR. However, the SRES was approved only in 2000 – that is to say one year before the publication of the TAR – and most modeling groups could finally run only two scenarios (A2 and B2) (Cubasch & et al., 2001; Giorgi & et al., 2001).

To avoid such coordination and timing issues in subsequent IPCC reports, it was decided that scenario experiments would henceforth be part of CMIP (G. A. Meehl et al., 2007). Thus, CMIP3 included three different climate change scenarios, corresponding respectively to the B1, A1B, and A2 scenarios of the SRES. More generally, CMIP3 was also more focused on climate change than the previous phases of CMIP – among the twelve CMIP3 experiments, ten tested the dynamic response of climate to various CO₂ concentrations (stable or evolving with time).

A main motivation behind these experiments was to help the Fourth IPCC report (AR4) to provide “a better assessment of the state of human knowledge on climate variability and climate change from the models” (G. A. Meehl et al., 2007, p.1384). As a consequence, CMIP3 was planned early enough in order for the analyses based on model experiments to be used in the Fourth IPCC report (AR4). Providing assessments for the IPCC reports was thus a

motivation at the core of CMIP3, whereas it was a side objective for the previous phases of CMIP.

In CMIP3, the role of CMIP to organize and coordinate climate research – although already present in the first phases of CMIP – also acquired a new dimension. Indeed, CMIP3 was the first phase of CMIP to give open access to all the data from its experiments. While in CMIP1, 2, and 2+, only a few modeling groups around the world have had access to the data and analyzed them, the data of CMIP3 were made accessible to any student or researcher around the world. It represented more than 30 terabytes of data. This new openness “brought global coupled climate model intercomparison and analysis to an internationally coordinated level never before achieved in the field of climate science” (WGCM, 2006). As such, it ushered “a new era in climate change research” (G. A. Meehl et al., 2007).

6.4.4 CMIP5 and 6 – the most recent phases of CMIP

The success of CMIP3 put the climate modeling community at the center of contradictory interests. More and more scientists outside the climate research community were interested in using CMIP results for their own areas of expertise. Therefore, CMIP5¹ was designed in order to satisfy not only the motivations of the climate modeling community, but also those of many different users:

the integrated set of CMIP5 simulations attempt to address major priorities of several different communities, and incorporates some of the ideas and suggestions of many individuals and from a number of workshops and meetings. These workshops involved scientists with a wide range of interests, including climate modeling, biogeochemistry modeling, integrated assessment modeling, climate change impacts, climate analysis, climate processes, and climate observations. (Taylor, Stouffer, & Meehl, 2012, p. 486)

In order to take into account the various requests expressed, CMIP5 included more experiments than CMIP3. The analyses of the simulations conducted in these different experiments have formed the basis of the IPCC Fifth Assessment Report (WGCM, 2012, p. 6). However, the deadlines imposed by the redaction of the IPCC report have put a lot of pressure on researchers. Moreover, CMIP5 has been extremely demanding in terms of computing and time resources, and blocked other research in modeling centers (Eyring et al., 2016).

Therefore, when thinking about the design of CMIP6, there has been a common will from the modeling community to decouple CMIP experiments from the IPCC, and to reorganize CMIP towards a few precise scientific questions relevant for climate and climate change. More precisely, it was decided that CMIP6 should be centered around three main scientific questions:

“How does the Earth system respond to forcing? What are the origins and consequences of systematic model biases? How can we assess future climate change given internal climate variability, climate predictability, and uncertainties in scenarios?” (WGCM, 2014, p. 13)

¹CMIP4 was skipped in order to make the numbering of CMIP phases in line with the numbering of IPCC reports.

These questions reflect a will of the research community to take control of CMIP again, and use it to better understand climate processes: “Ultimately scientific progress on the most pressing problems of climate variability and change will be the best measure of the success of CMIP6” (Eyring et al., 2016, p. 1949). In particular, the importance given to scenarios in previous phases of CMIP was questioned. Indeed, although they are central in IPCC reports, scenarios are of little use in understanding climate processes themselves, as they involve many different assumptions that make it difficult to interpret their results. Therefore, it was decided that the scenarios would not belong to the core experiments of CMIP6 – the experiments that all participating groups should perform – but to the *CMIP-Endorsed Model Intercomparison Projects (MIP)* – secondary experiments on a voluntary basis.

6.5 Uses of CMIP results and related controversies

In the previous section, we have seen that even if the principle behind CMIP has always remained the same – comparing climate models by submitting them to a common set of simulations – its design and the motivations behind it have evolved significantly over the phases of the project.

However, the evolution of CMIP design and the motivations behind it is only one part of the story. To fully grasp what CMIP has become, we also need to understand how CMIP results have been used by climate scientists. For that, we looked at peer-reviewed papers based on CMIP outputs: more precisely, we consulted a set of 280 papers based on CMIP5 results published between 2012 and 2018 in six leading climate journals². In this set of publications, based on the title of the papers and their abstracts, we distinguished qualitatively three main uses of CMIP results: the exploration of future climate change and associated uncertainties, the comparison of CMIP simulations with observations and the interpretation of model results. Table 6.1 gives a summary of how often these different uses come in the various journals from which the papers were taken.

The three uses we distinguished are not exclusive – in particular, we counted 25 papers exploring future climate change that first compare CMIP simulations with observations in order to assess their quality. Nor are these categories exhaustive – among other uses, we noted in particular the study of paleoclimates or present climate. Besides, there are two kinds of papers based on CMIP results we did not consider in our study. First, papers analyzing CMIP results from only one or two models. Secondly, papers not using CMIP results as such, but proposing statistical methods to analyze them or calculate the associated uncertainties.

Our attribution of uses is in part subjective, and one might adopt a different classification. However, we hold that a large portion of the literature based on CMIP results explores future climate change and associated uncertainties, compares CMIP simulations with observations and/or tries to interpret the model results. Yet, we will see in the next part that each of these uses raises methodological questions.

²Climatic Change, Climate Dynamics, Environmental Research Letters, Geophysical Research Letters, Journal of Climate, Journal of Geophysical Research.

| Journal | Number of papers in the data set | Number of papers using CMIP results mainly to: | | |
|---------------------|----------------------------------|--|--|---------------------------------|
| | | Explore future climate change and associated uncertainties | Compare CMIP simulations with observations | Interpretation of model results |
| Clim. Dyn. | 52 | 21 | 19 | 22 |
| Clim. Change | 7 | 7 | 2 | 0 |
| Environ. Res. Lett. | 12 | 7 | 3 | 3 |
| Geophys. Res. Lett. | 64 | 28 | 14 | 29 |
| J. Climate | 92 | 41 | 26 | 30 |
| J. Geophys. Res. | 53 | 21 | 22 | 14 |
| Total | 280 | 125 | 86 | 98 |

Table 6.1: Summary of main uses of CMIP results for peer-reviewed papers ordered by journal. For a detailed list of all peer-reviewed papers analysed, see Touzé-Peiffer et al. (2020).

6.5.1 Exploring future climate change and the associated uncertainties

Exploring future climate change is the most widespread use of CMIP results we have identified in our set of papers. In the 125 papers tackling this issue, many different themes were addressed: among others, we can cite regional impacts of climate change (e.g. Luomaranta et al., 2014; Penalba & Rivera, 2013; Zomer et al., 2014), decadal predictions (e.g. Gaetani & Mohino, 2013; Guemas, García-Serrano, Mariotti, Doblas-Reyes, & Caron, 2015; G. A. Meehl et al., 2014), or the consequences of climate change on one particular climate phenomenon, for instance the Asian monsoon (e.g. Jayasankar, Surendran, & Rajendran, 2015; Srivastava & DelSole, 2014; Zou & Zhou, 2015), the ENSO (e.g. Stevenson, 2012; Taschetto et al., 2014) or the poleward expansion of Hadley Circulation (e.g. Hu, Tao, & Liu, 2013).

However, the interpretation of multi-model ensembles provided by CMIP is controversial. Let's consider for instance, Fig. 6.1, which gives the spread of the temperature and precipitation in India during the summer monsoon according to a set of CMIP5 models. We see that CMIP5 models simulate a future warming of Indian landmass at the end of 21st century by about $1.19 \pm 0.79^\circ\text{C}$ for RCP2.6 and $3.99 \pm 1.27^\circ\text{C}$ for RCP8.5. Concerning the change in precipitation, there is a relatively larger model spread, with a projected change of precipitation of 0.39 ± 0.79 mm/day for RCP2.6 and 0.95 ± 1.13 mm/day for RCP8.5 (Jayasankar et al., 2015). Seemingly, this range provides a measure of the uncertainty about future climate change – in this example, it seems that there is much more uncertainty about the impacts of climate change on precipitation than on temperature.

Nevertheless, such a statement raises epistemological issues. First, since the participation in CMIP is made on a voluntary basis, the set of models participating in CMIP has not been

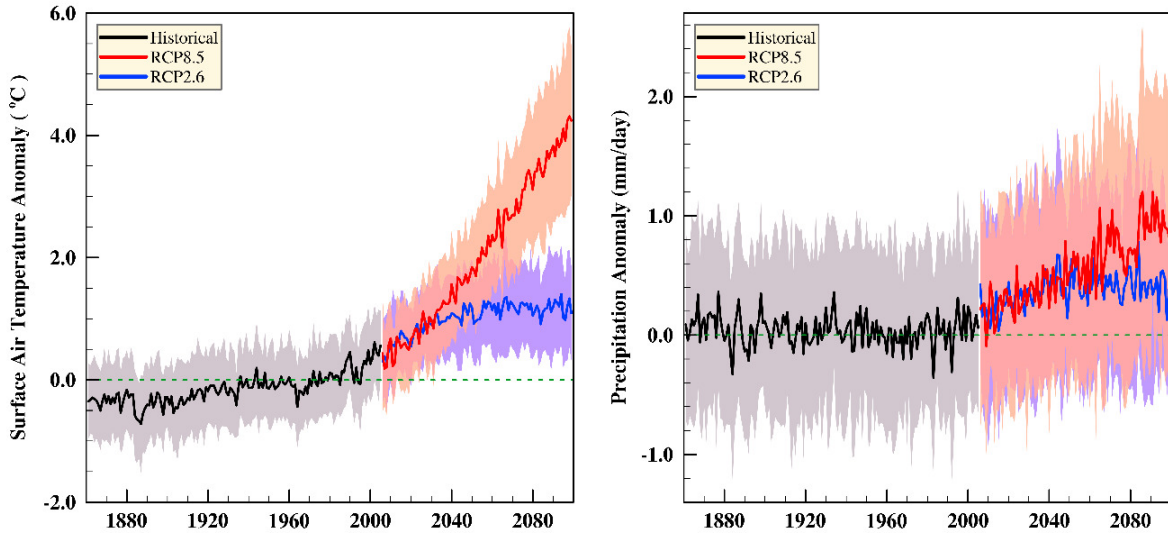


Figure 6.1: JJAS anomalies with respect to the base period (1961–1990) in CMIP5 historical simulations and 21st century projections of (a) surface air temperature and (b) precipitation for India. The intermodel spread is shown in shades (gray for historical runs, purple for RCP2.6 simulations, and light orange for RCP8.5 simulations). The solid lines represent the multi-model average for historical (black), RCP2.6 (blue), and RCP8.5 (red). Courtesy Jayasankar et al. (2015).

designed to span the uncertainty about future climate change. The distribution of CMIP multi-model ensembles is in a large part arbitrary. It led Tebaldi and Knutti (2007) to describe multi-model data sets as “ensembles of opportunity”. Since these ensembles of opportunity do not systematically explore the uncertainty about future climate change, some scientists consider them only as a lower bound of this uncertainty (e.g Stainforth et al., 2007). How low is this bound, however, remains unknown.

As a corollary, it is difficult to interpret the agreement between models participating in CMIP. As documented by Pirtle et al. (2010), in many scientific papers, confidence in model projections about future climate change is justified by the agreement between GCMs in multi-model ensembles. For instance, in Fig. 6.1, all models agree that the surface air temperature anomalies in India with respect to the base period (1961–1990) are below 6°C under the business as usual scenario (RCP8.5). Yet, if we consider the spread in CMIP projections only as a lower bound of the uncertainty, we would not conclude with confidence that this result stands as well for the real world.

A less ambitious goal would be to use CMIP simulations to have a “best guess” of future climate change. A natural candidate, chosen in many studies, is the average of the projections, as in Jayasankar et al. (2015) (see Fig. 6.1, where the average is represented in solid lines). However, in some situations, the rough average masks the signal entirely. For precipitation for instance, when temperatures rise, many climate models predict large-scale drying in the subtropics and moistening at high latitudes, but at slightly different locations. As Knutti et al. (2010) showed, the problem is that when models are averaged, they tend to cancel each other out. Thus, at some latitudes, all models show significant drying over a relatively large fraction of the land surface, but the average does not.

More generally, the relevance of the gross arithmetic mean for climate projections – what Knutti (2010) called “model democracy” – has often been questioned for two main reasons: 1) Climate models are not independent of each other – on the contrary, as Masson and Knutti (2011) and Knutti, Masson, and Gettelman (2013) have shown, they share pieces of code and common ideas. 2) Climate models do not have the same degree of agreement with observations: some perform better than others. These disparities between models motivated the work of Knutti et al. (2017) who proposed, as an alternative to the gross arithmetic mean, a weighting scheme that takes into account both the large differences in model performance and the interdependencies between models. Whether similar weighting schemes should be used more frequently remains controversial³.

Taking CMIP results as a lower bound of the uncertainty or using them as a “best guess” are not the only ways to interpret CMIP multi-model ensembles. Both climate scientists and philosophers of science (see for instance W. S. Parker, 2011; Tebaldi & Knutti, 2007; Winsberg, 2018a) have proposed other alternatives, but none of them has yet reached consensus. Although CMIP multi-model ensembles are often used to explore future climate change and associated uncertainties, how to interpret thus remains an open question.

6.5.2 Comparing CMIP simulations with observations

Another common use of CMIP results consist in comparing them with observations in order to assess the performance of the corresponding GCMs (86 papers in our set). A variety of observations can be used, such as station data, satellite data, proxy data or reanalysis data. However, this measure of performance of CMIP models raises another bunch of analytical problems.

The first one is related to the fact that, in complex climate models, many parameters are poorly constrained by observations and are adjusted in order to satisfy key climate metrics. Hourdin et al. (2017a) provide an excellent overview of the tuning of parameters in GCMs and the epistemological problems it poses. As they explain, the agreement with observations can be improved by changing parameters not directly relevant for the problem at stake. For instance, they show that the global top-of-atmosphere energy balance can be adjusted by changing a parameter controlling the fall velocity of ice crystals, a parameter which, at first glance, seems far from the issue at hand. It is thus possible to get the right result for the wrong reasons – the agreement with observations may result from compensating errors and do not necessary prove that processes are well represented in the model. As noted by Tebaldi and Knutti (2007), Frisch (2019), and others, this limitation is particularly strong when the same data sets are used to tune a model and to evaluate its performance.

Another issue concerns observations themselves. In our set of papers, the large majority of studies do not compare CMIP simulations with model-independent observations, but only with reanalysis. Since reanalysis models are based on numerical methods, assumptions, and parameterizations similar to those of real climate models, they all exhibit biases of various kinds (Edwards, 2010). Even if the analysis is continually corrected by available observational data, reanalysis models thus transmit part of their biases to reanalysis products. Therefore, an agreement between climate models and reanalysis data might just illustrate a common bias in

³For a detailed discussion of this issue and other challenges in combining projections from multiple climate models, see Knutti et al. (2010).

both climate and reanalysis models. This problem gets worse for processes for which we have few observations, such as the hydrological cycle (Tebaldi & Knutti, 2007) – the reanalysis is then loosely constrained by observations, but mainly model-derived.

As a consequence, although the agreement between models and observations is valuable to identify systematic biases in climate models, it is not a guarantee of their reliability. Due to the number of parameters in GCMs and to the interdependencies between GCMs and observations, the agreement between GCMs and observations validates GCMs only in a very weak sense.

6.5.3 Interpreting model results

Originally, the official motivations behind AMIP were formulated as such: “The basic purpose of AMIP is to undertake the systematic intercomparison and validation of the performance of atmospheric GCMs on seasonal and interannual time scales under as realistic conditions as possible, and to support the in-depth diagnosis and interpretation of the model results.” (Gates, 1992, p. 1963). The hope was that this “in-depth diagnosis and interpretation of the model results” would help to identify the causes of success and failures of participating climate models, and thus to improve their performance. In particular, Gates had identified the parameterization of convection and precipitation as an outstanding modeling problem and called for further analysis of AMIP results to reduce errors related to it (Gates et al., 1999).

In our set of papers, we counted 98 papers attempting to interpret model results. Yet, none of them look at the details of the parameterizations involved. Most of these studies stay at the level of model outputs; by analyzing correlations between the outputs of each model, they exhibit causal relationships between them. However, where this causality comes from in the details of the corresponding models is not addressed. Kent, Chadwick, and Rowell (2015) provide an example of such a study; their objective is to understand uncertainties in future projections of seasonal tropical precipitation. They investigate correlations between precipitation, global mean temperature, pattern in sea surface temperature and a few other variables. The influence of the precipitation schemes used by the different models is left completely out of the picture. Another example is provided by studies who try to assess the effect of one or several feedbacks under CO₂ increase (e.g. Long & Collins, 2013; Qu & Hall, 2014). These studies diagnose and interpret climate feedbacks and their effect solely on the basis of simulation results, but never discuss the underlying parameterizations.

According to Lenhard and Winsberg (2010), CMIP analyses stay at the level of model results because climate models exhibit a form of “confirmation holism”. This concept is traditionally defended in philosophy of science as the idea that a single hypothesis can never be tested in isolation, but that such tests inevitably depend on other theories or hypotheses. As Pierre Duhem – one of the first to formulate this theory – writes it:

In sum, the physicist can never subject an isolated hypothesis to experimental test, but only a whole group of hypotheses; when the experiment is in disagreement with his predictions, what he learns is that at least one of the hypotheses constituting this group is unacceptable and ought to be modified; but the experiment does not designate which one should be changed. (Duhem, 1954, p. 187)

As Lenhard and Winsberg (2010) explain, climate models face a particularly strong form of confirmation holism because there is at the same time a high modularity in their development

and lots of interactions between their different modules. As they put it: “The complexity of interaction between the modules of the simulation is so severe that it becomes impossible to independently assess the merits or shortcomings of each submodel.” (Lenhard & Winsberg, 2010) To illustrate their point, Lenhard and Winsberg (2010) take the example of AMIP and other intercomparison projects, and point out the difficulties these intercomparison projects have historically had to diagnose the causes of the successes and failures of climate models. In their view, because of these difficulties, model convergence is unlikely: in the foreseeable future, there will continue to be a plurality of models making divergent projections. Policy makers should therefore accept this plurality, and not wait for a unanimous voice from the climate modeling community.

While we agree with some of the conclusions from Lenhard and Winsberg (2010), we would qualify their claim that climate models themselves face a strong form of confirmation holism. Indeed, a climate model is not just the sum of the code (and the assumptions behind it) and the results it gives for a particular set of simulations. On the contrary, it is a dynamical entity with which it is possible to interact. When climate scientists want to study a climate model, they can initialize it with various conditions, change the parameters of the model, compare the results of simulations with data from observations or high-resolution simulations, use simplified or idealized versions of the model or other models, etc. Thanks to these various interactions, climate scientists can acquire a knowledge about the behaviour of a climate model, what it is doing and why.

This knowledge is most of the time collective, because it results from collaboration in research laboratories between different individuals working on separate but complementary aspects of the same climate model.⁴ However, if this knowledge is collective, it stays usually at the level of one research team working on one model. Indeed, due to the complexity of the models involved in CMIP, acquiring knowledge about the behaviour of a climate model takes time and scientists generally focus their efforts on one particular model. In fact, in the literature, we can find many studies investigating the link between the results of a model and its parameterizations (e.g. Hourdin et al., 2013; Notz, Haumann, Haak, Jungclaus, & Marotzke, 2013). But most of these studies are done for one model only – when many models are studied, as in CMIP multi-model ensembles, the details of the parameterizations involved are almost never taken into account. There are, of course, some rare exceptions: in particular, studies comparing radiation codes in different climate models, such as Oreopoulos et al. (2012) and Pincus et al. (2015), where the authors analyze not only the model results, but also the corresponding parameterizations and the assumptions they make⁵. Based on the set of papers we have studied, we claim nevertheless that the large majority of papers using CMIP results do not look at the details of the parameterizations involved. In other words, our analysis suggests that CMIP has promoted analyses which do not discuss the content of GCMs, but consider them as black boxes. Hence, a strong form of confirmation holism does exist: however, unlike Lenhard and Winsberg (2010), we argue that it is not so much GCMs per se, but GCMs as they are used in CMIP that are facing it.

⁴This collective aspect of knowledge is not a specificity of climate modeling. On the contrary, as underlined by Hardwig (1985), nowadays, epistemic dependence among scientists is pervasive.

⁵This is due to the fact that the representation of radiation in GCMs is a very different problem from the other parameterizations, and much better grounded. For more details about what makes radiation parameterizations unique, see Pincus and Stevens (2013), Pincus et al. (2015), and Pincus, Mlawer, and Delamere (2019).

In this section, we have examined how CMIP results have been used in scientific papers. We will see in the next section that this gives only a partial view of CMIP influence on climate research – beyond analyses based on their results, intercomparison projects have indeed had structural effects on climate research.

6.6 Structural effects of CMIP on climate research

CMIP should not be reduced to a set of simulations performed every 5 or 6 years by the main modeling groups in the world. All the social interactions, questions, and scientific coordination it creates have to be taken into account. CMIP is indeed a conglomerate of technical tools, common objectives, dedicated workshops, and so on – what we will call an infrastructure for climate research.

Let’s define more precisely what we mean by this. For that, we will rely on the concept of technological infrastructure introduced and studied by the contemporary philosopher Joseph Pitt. Pitt defines a technological infrastructure as “an historically determined set of mutually supporting artifacts and structures that enable human activity and provide the means for its development” (Pitt, 2000, 129). As far as science is concerned, Pitt explains that his definition does not include only shared machines or techniques, but also all the work relations among scientists which makes the doing of science possible. There is therefore a social component in Pitt’s definition of the “technological infrastructure of science”. There is also an historical and historiographical aspect behind this concept – it can be used to follow and understand advances of science:

the mechanism that makes the discoveries of science possible and scientific change mandatory is the technological infrastructure within which that science operates, and that to understand why a science worked the way it did, and why it works the way it does, you need to understand its context, which happens to include in important ways its technological infrastructure. (Pitt, 2000, p. 132)

As an infrastructure, CMIP has shaped climate science in many ways. First, the existence of CMIP has focused the effort of the scientific community on the construction of GCMs. Twenty years ago, Shackley et al. (1998) already underlined that GCMs were commonly considered as the “best climate models”. We argue that CMIP has reinforced this trend. In almost all major climate modeling groups in the world, each phase of CMIP has appeared as an international *rendez-vous*. To have a voice in it, every modeling group had to have the most recent version of its GCM ready. It created a pressure around the development of GCM, at the expense of other tools used to study climate.

This supremacy of GCMs is questionable, when considering the many issues exposed in section 2 regarding the validation of GCMs and the interpretation of their results. When modeling complex processes, there is always a trade-off to find between the inclusion of perceived complexity and the understanding of dominant processes, interactions, feedbacks, and uncertainties (Shackley et al., 1998). Yet, for GCMs, there is no trade-off: the balance is clearly in favor of the inclusion of perceived complexity. As a consequence, some climate scientists have argued that hierarchies of models of varying complexity would bring more insights about the climate system and should be more frequently used in climate science (e.g. Bony et al., 2013; Maher et al., 2019). Examples of commendable initiatives to compare GCMs with simpler models already exist, such as: 1) RCMIP (Reduced Complexity Model Intercomparison

Project), a unique CMIP6 sub-experiment, which provides a standard protocol for comparing simple models and emulators to the latest CMIP results 2) The use of single-column versions of GCMs compared to explicit high-resolution simulations to build and test parameterizations (e.g. Rio, Hourdin, Couvreux, & Jam, 2010; Rochetin et al., 2014).

Despite these efforts, voices have raised in the climate modeling community to protest that the pace of progress was too slow and that GCMs as they were used were not the appropriate scientific response to the challenges posed by global warming. Different alternatives in the practice of climate modeling have been proposed (e.g. Hurrell et al., 2009; T. N. Palmer, 2012; Shukla et al., 2009) – Katzav and Parker (2015) provide an overview and critical examination of these different approaches. More recently, T. Palmer and Stevens (2019) have argued for a new strategy based on higher resolution models and a new approach to parameterizations with stochastic modeling. Though in different ways, all of these proposals offer alternatives to the CMIP dogma.

In addition to focusing the climate community on GCMs, another main effect of CMIP on climate science is to have promoted connections between climate laboratories. It has helped them to talk to each other and share common references. As the historian of science Paul Edwards puts it:

By permitting regular, direct, and meaningful comparisons of the models with one another and with standardized data sets, [climate model intercomparison projects] have helped to transform climate modeling from a craft activity of individual laboratories into a more modular and standardized collective activity involving virtually all of the world's climate modeling groups; in theoretical terms, they linked a set of isolated systems and created a network. (Edwards, 2010, p. 350)

In practice, these connections were made through different means:

1. First, CMIP made its data freely available from all over the world thanks to a distributed structure, the *Earth System Grid Federation*. It has also supported the development of technical tools to convert the output data of various modeling group into a standardized format, making therefore their exchange easier.
2. Second, CMIP has spread some common scientific approaches in the climate modeling community, thereby facilitating the comparison of climate models with one another. We can cite for instance abrupt 4 times CO₂ experiments – simulations in which the CO₂ concentration in the atmosphere is immediately and abruptly quadrupled from its pre-industrial values – or transient simulations in which the CO₂ concentration is increased gradually at a rate of 1% per year. Thanks to CMIP, these two kinds of simulations have become standard tools to investigate future climate change with GCMs.
3. Last but not least, dedicated workshops, special issues in scientific papers, etc. have fostered the interest of climate scientists around CMIP results and led to various collaborations between modeling groups.

A side effect of the open-access to CMIP data is the creation of a growing gap between model developers and model users. Before AMIP and CMIP, the results of a GCM simulation were usually analyzed by the few people who had been involved in the development of the corresponding GCM (e.g. Manabe & Wetherald, 1975; Washington, Semtner, Meehl, Knight,

& Mayer, 1980). Model users had therefore a critical view of the strengths and weaknesses of the climate model they analyzed, because they had contributed to develop it. When data from intercomparison projects were made freely available, GCMs started to be analyzed by people who had not participated in their development. It resulted in a loss of understanding of climate model results, and increased the tendency to use GCMs only as black boxes (see 2.3).

Nevertheless, in CMIP6 overview paper, the authors claim that CMIP has favored “scientific progress on the most pressing problems of climate variability and change” (Eyring et al., 2016, p. 1949). Is it true? As an infrastructure, did CMIP succeed in making climate research more effective?

Here, we think it is important to distinguish between effectiveness and efficiency. Something is effective if it is adequate to achieve an objective. In contrast, something is efficient if it works in the best possible way. In other words, being effective is about doing the right things, while being efficient is about doing things right. A process can thus be efficient – for instance, if it is fast or cheap – but ineffective – if it is not well suited to the objective we want to achieve.

CMIP has certainly helped climate research to be more efficient. Indeed, it has pooled many time-consuming activities at the level of the research community, and therefore facilitated the work of climate scientists. For a single laboratory, building a simulation protocol is indeed a time-consuming and costly effort. Boundary conditions, forcings and the associated databases, output parameters and their format, as well as versions of the models involved have to be carefully defined. Thanks to CMIP, climate laboratories have shared this burden. The use of standardized format for the output data has also simplified a lot the analyses based on the results of simulations, and in particular, the comparison with data from observations. As Eyring et al. (2016) explain:

A key to the success of CMIP and one of the motivations for incorporating a wide variety of coordinated modeling activities under a single framework in a specific phase of CMIP (now CMIP6) is the desire to reduce duplication of effort, minimize operational and computational burdens, and establish common practices in producing and analysing large amounts of model output. (Eyring et al., 2016)

We thus agree that the production and the analyses of GCMs simulations have been made more efficient thanks to CMIP. But did it make climate research more effective? CMIP has focused the effort of the climate research community on one specific tool, GCMs, and has promoted the interpretation of this specific tool as a black box. Was it and is it still the most effective way to help climate scientists better understand climate variability and change? The debate stays open.

6.7 Conclusion

When looking at the history of CMIP, it seems that this project has known a growing success. It has promoted a coordination at an international level never before achieved in climate science. With all the major modeling groups in the world participating, CMIP has been massively endorsed by the research community. In addition, its results have played a key role in IPCC reports. However, while the analyses of CMIP6 are currently being undertaken, it might be time to pause and reflect about what CMIP has become and what it has brought to climate research.

First, CMIP has focused the attention of the climate research community on GCMs. As the ethnographer of climate modeling Simon Shackley observed, GCMs are often considered as the best models in climate science (Shackley et al., 1998) and the most useful one to predict future climate change. However, this view has been challenged by Bony et al. (2013); Shackley et al. (1998), and others. According to them, though valuable, GCMs should not be seen as the ‘panacea’ of climate science (Bony et al., 2013). Different models are useful for different purposes and simpler models can also provide valuable insights for understanding climate processes. They should therefore be more frequently used in complementary of more complex GCMs.

Another issue comes from the fact that CMIP promoted a particular use of GCMs as black boxes. Since CMIP data are freely available, anybody can analyse them. As a consequence, the large majority of CMIP analyses are conducted by scientists who have not been involved in the development of the corresponding GCMs, and have therefore a poor knowledge of the content of the climate models at hand. In fact, CMIP has created two distinct communities – model users and model developers – with few interactions between them. The existence of these two communities is an issue for the interpretation of CMIP results. In particular, CMIP results are almost never used to guide model improvement, whereas it was one of the main objective for AMIP.

We argue that there has not been enough explicit debate on the value of CMIP for climate research and policy guidance. We strongly encourage scientists to examine the consequences both for science and society of the particular form of research they have entered with CMIP. In the context of the challenges posed by climate change, given the limited means and computing resources available, there should be more discussions on the goals, epistemology and policy context of this tool at the core of climate research. Hopefully, our historical and epistemological perspectives about what CMIP was and what it is now will not close this debate but open it.

Chapter 7

Conclusion

7.1 Summary

In this thesis, I explored various epistemological issues associated with the parameterization of atmospheric convection. The originality of my work is that it is located at the interface between climate science and philosophy. I tried to make my epistemological reflection as close as possible to the questions atmospheric physicists ask themselves in their practice. During my Ph.D., I had the opportunity to participate in various atmospheric studies that led to interesting scientific results (Albright et al., 2020; Hwong et al., 2021; Madeleine et al., 2020; Rochetin et al., 2021; Stevens et al., 2021; Touzé-Peiffer et al., 2021). I chose in this thesis not to emphasize these scientific results as much as the epistemological reflections that accompanied them. I summarize below the main questions that I investigated.

Objects in the study of atmospheric convection

How to study atmospheric convection? Atmospheric phenomena all come from the same physical laws. However, one only has to look out the window to see that the atmosphere is always full of new faces. In particular, each cloud is unique: clouds never have exactly the same shape, the same brightness, the same appearance. Unlike clouds, air movements transporting water vapor and heat in the atmosphere are invisible. However, we know that there is as much diversity in these invisible air movements as there is in clouds. If we stopped at this diversity, we would not be able to learn anything from natural phenomena. We could not remember what we saw yesterday, understand what we see today and predict what we will see tomorrow. To learn, understand and predict, we need to name some “objects” that emphasize certain similarities linking phenomena despite their differences. Objects are the characteristic patterns that help scientists to orient themselves in the jungle of atmospheric phenomena. The main results of my study are 1) that objects are not just given by nature, but come also from a construction, 2) that objects define the way we study and understand atmospheric convection, both in observations and in simulations and 3) that it is critical to choose relevant objects to have the best angle of attack to understand atmospheric convection. This third point is an immediate consequence of the first two. To illustrate my argument, I used the example of an atmospheric phenomenon called “cold pools” and showed the role played by this object in our understanding of atmospheric convection.

Parameterization and tuning of atmospheric convection

Parameterizations of atmospheric convection are often pointed out as one of the major causes of uncertainty in GCMs (Gentine et al., 2018; Schneider, Teixeira, et al., 2017). They are sometimes described as *ad hoc*, i.e. not entirely justified by theory, and therefore not trustworthy. For the past 20 years or so, some scientists have called for parameterizations of atmospheric convection to be abandoned and have proposed alternatives to replace them. However, such parameterizations are still widely used today and continue to be of interest to other scientists. These different strategies to study climate reflect different epistemological conceptions of the role of climate models and of their convection parameterizations. In particular, parameterizations have sometimes been criticized because they are not deduced from first principles, but semi-empirically, by combining some theoretical principles and empirical assumptions derived from observations (Edwards, 2001; Petersen, 2012). I showed that this semi-empirical approach is not, in principle, less well founded than an approach based entirely on theory. On the contrary, it is a convenient way to think the parameterization problem using objects. Nevertheless, I mentioned that parameterizations impose an artificial separation of both scales and processes, two important limitations of their use in GCMs.

I also focused on the tuning of GCMs. The tuning of GCMs is different from a classical calibration procedure, which would simply aim at making an instrument that works well in theory also work in practice. Nor is it completely arbitrary, as some accuse, and used only to hide certain deficiencies of the model. I aimed to define the role and status of tuning in climate models between these two extreme positions. I explained in particular that the non-accuracy of parameterizations is crucial to understand why most of GCM parameters are poorly constrained by observational data, thus have to be tuned. I also clarified conceptually some terms used to speak about the tuning of GCMs, such as “compensating errors”, “overtuning” and “structural errors”. I suggested to avoid using this last term and proposed to replace it with “structural limitations”. Finally, I investigated the paradigm shift brought by automatic tuning methods: I explained notably that with them, the tuning process becomes more similar to a traditional calibration procedure.

The potential of automatic tuning methods is then studied in practice through a case study, where I developed a parameterization of cloud and precipitation overlap for the atmospheric model LMDZ. Through this case study, I asked a simple question: what does it mean to improve a GCM? To answer it, I distinguished between the conceptual progress that a new parameterization brings to a given GCM and the progress it brings on the model results. I explained that conceptual progress can be evaluated by considering criteria of Coherence, Interpretability, Simplicity and Comprehensiveness – what I called the CISC criteria. I then argued that progress in terms of model results have to be assessed independently of the tuning of a GCM, as differences in model results might be due to a different tuning rather than to differences in parameterizations. This creates significant methodological difficulties, especially if the tuning is carried out manually. I proposed a definition allowing to evaluate the progress brought by a new parameterization to a climate model independently of its tuning. Even if progress remains to be made, automatic tuning methods allow to consider applying this definition in practice in order to compare two versions of a GCM.

Historical and epistemological study of CMIP

Historically, one of the main roles of early model intercomparison projects was to improve GCMs, especially their parameterizations. Their principle was indeed to compare different GCMs under similar conditions in order to diagnose their respective strengths and weaknesses, and to evaluate in particular which parameterizations worked well and which ones did not. I explained that the Coupled Model Intercomparison Project (CMIP) has had a critical role in climate research, but not this one. In fact, CMIP has promoted analyses that consider GCMs as “black boxes” and do not look at the details of their parameterizations. CMIP has thus contributed to creating a gap between a (small) community of model developers and a (large) community of model users.

More generally, the way CMIP results are used today raises serious epistemological issues. First, when CMIP results are used to explore future climate change and associated uncertainties, it is not clear how to interpret the agreement between models. I explained that these difficulties come from the fact GCMs are dependent from one another, leading to possible common biases, and also that they have not been chosen to systematically sample the uncertainty. A second common use of CMIP results consists in comparing twentieth century simulations to observations in order to assess the performance of GCMs – and thus their ability to predict future climate change. Again, I reviewed some epistemological problems associated with this use of CMIP results – especially when the same observations are used to tune GCMs and evaluate their performance, or when reanalyses are used as observations.

Despite these difficulties, CMIP has been massively endorsed by the climate modeling community. I attributed this success in part to the IPCC, which advertised the analyses based on CMIP results. For the IPCC, CMIP was indeed an opportunity to summarize GCM projections about future climate change. In my analysis of CMIP, I tried to underline the tension between the need to provide information to decision makers via the IPCC reports and the need to advance science. Although CMIP has been an *efficient* way to provide statements for the IPCC reports, I questioned the fact that it is today the most *effective* way to drive progress in climate science.

7.2 Perspectives

My thesis work could provide food for thought on some concrete issues that the climate modeling community will have to face in the coming years. In the following, I provide an overview of some of these perspectives.

The future of CMIP

What future do we want for CMIP? What will CMIP7 look like? What about CMIP8, 9, 10, if CMIP continues until then? I firmly believe that AMIP and the early phases of CMIP, despite their limitations, have been extremely useful to the climate modeling community. These standardized model comparisons have forced climate modelers to make the results of their models accessible to everyone. Such transparency made it possible to identify biases common to many GCMs. As summarized by Stouffer et al. (2017) and Bock et al. (2020), analyses based on CMIP results have shown for instance the tendency of climate models to simulate land surfaces that are too warm and too dry during the summer, their difficulties to simulate accurately the

structure of the intertropical convergence zone or the Walker circulation, some problems related to the simulation of tropical and subtropical low-clouds, especially stratocumulus decks over the eastern parts of ocean basins, etc.

CMIP also demonstrated the robustness of some results about climate change. As discussed in I. M. Held and Soden (2006), many features of climate change could be expected from simple thermodynamic arguments or idealized models. For instance, in Chapter 3, we mentioned that in a simplified framework, Manabe and Wetherald (1967) have shown many results about climate change which are still relevant today. Nevertheless, CMIP proved the robustness of these results in complex GCMs as well. It gave confidence that such simple results were not proved wrong when the general circulation of the atmosphere and the coupling with oceans and other components of the climate system were taken into account.

Now that these biases and some robust features have been identified, one might ask: has CMIP had its day? Climate predictions from the last phases of CMIP have not changed much. Numerous studies investigating various aspects of CMIP results have found the same biases in CMIP3, CMIP5 and CMIP6 (e.g. Beadling et al., 2020; Harvey, Cook, Shaffrey, & Schiemann, 2020; Tian & Dong, 2020). Similarly, the spread of equilibrium climate sensitivity has not changed significantly since the early assessments of J. G. Charney et al. (1979). CMIP also did not resolve the biases identified in GCMs: as we have seen in Chapter 6, CMIP decoupled the analyses of GCMs from their improvement, thus promoting the use of models as black boxes. Consequently, most analyses based on the CMIP results offered little opportunity to improve the models, and in particular their parameterizations. Therefore, what will be the added value of the next phases of CMIP? What new results can we expect from CMIP7 and subsequent phases compared to CMIP6?

These questions seem all the more relevant as model development has slowed down in most climate modeling centers. A few years ago, Jakob (2010) already mentioned the lack of attractiveness of model development, and the difficulties in finding model developers among early career researchers. These challenges are even more important today because the alternative methods that are increasingly being used to replace GCMs – machine learning and high-resolution models in particular – are mobilizing part of the research effort in climate science. Contrary to these new prospects which lead to numerous publications, work on parameterizations is a long process, requires a high degree of specialization, and is difficult to exploit in scientific publications. The future will tell us if the time and effort invested in alternatives to parameterizations is worth it, but the fact is that, due to the lack of human resources, current progress on parameterizations is slow. Even if sensible progress can be noted in some GCMs in the past decade (see Rio et al., 2019, for an overview), most GCMs have not changed significantly between CMIP5 and CMIP6, apart from increases in the horizontal and vertical resolution. Major changes are not expected for the next phase of CMIP either. If the models are more or less the same from one CMIP phase to the next, what would be the point of re-running simulations every 5 or 6 years?

A potential solution would be to retain CMIP protocols, which provide a benchmark for the comparison of successive GCM versions, but to abandon the concept of “CMIP phases”. I propose that modelers should be able to submit a new version of their models to CMIP simulations whenever they want, i.e., whenever they believe that a new version is sufficiently different from the previous one and sufficiently mature to justify a new round of simulations. This would

create less pressure on model developers and allow them to develop their models according to their own planning and priorities, independent of the timeline imposed by CMIP. This would also prevent the “publish first” effect after each phase of CMIP, which leads to numerous papers of questionable scientific value, simply updating previous analyses with the new CMIP data.

Another challenge will be to compare the CMIP results with the new generation of coupled non-hydrostatic climate models with kilometer resolution, that we called global cloud-resolving models (GCRMs). DYAMOND (the DYNamics of the Atmospheric general circulation Modeled On Non-hydrostatic Domains Stevens, Satoh, et al., 2019) is the first intercomparison project of GCRMs. In some aspects, DYAMOND resembles CMIP: there is in both projects the same effort to run model simulations under similar conditions in order to compare their results. Yet, DYAMOND was not integrated into CMIP, but rather appeared as an alternative paradigm, and even a competitor to CMIP in terms of human and computer resources. In this context, how to relate CMIP to this new generation of models ? Should DYAMOND be integrated into CMIP? Or should we let the two intercomparisons evolve in parallel? With what objectives? CMIP will have to answer these many questions in the next few years.

Last but not least, I want to underline the new perspectives automatic tuning methods offer to CMIP. In Chapters 4 and 5, we have seen that with automatic tuning methods, a climate model is not just a single simulation, but an ensemble of simulations that are deemed plausible based on certain metrics. This gives a new meaning to the comparison of two climate models, which no longer becomes the comparison of a single simulation of each model, but the comparison of a set of plausible simulations of each model. One can therefore imagine that each version of a GCM submits several simulations to CMIP, which differ by their tuning. With the same number of models, many more worlds could then be explored with CMIP. This would allow to check more robustly some features of future climate change, in particular the spread in climate sensitivity. A reflection should therefore be initiated on how best to integrate these new opportunities into CMIP.

Destination Earth

While I was finishing this thesis, I became aware of the Destination Earth (DestinE) project that was being discussed at the European level. DestinE’s ambitious goal is to “construct highly accurate models, or ‘digital twins’, of the Earth to monitor and predict environmental change and human impact in support of sustainable development” (Bauer, Stevens, & Hazeleger, 2021, p. 80). The authors explain that such digital twins of the Earth, constrained by observations and the laws of physics, could be used to guide policy makers and businesses for a variety of applications. DestinE should help answer questions such as: “What is the economic impact of adding wind power plants in an area once Europe is carbon neutral? What agricultural policy is the most sustainable given future weather regime pattern changes?” (p. 81-82). The means to achieve such technical feats are rather elusive: Bauer, Stevens et Hazeleger explain that DestinE will exploit the potentials of “machine learning”, “extreme-scale data assimilation”, of an “highly efficient virtual environment” without precising explicitly how these means will be concretely used. Reading between the lines, we understand that the digital twins will be based on global models with a kilometric resolution – what we called GCRMs – coupled with models of human impacts and economics. There is thus first a technical debate as to whether it is possible to build such models, and on what time scale.

In addition to this technical debate, the use of such digital twins raises many epistemological questions. In particular, one sentence in the paper describing the project caught my attention:

“An exciting aspect of a digital twin is the potential to break the paradigm of classical Earth system prediction models with fixed and static flows of information managed by layers of experts. Here, the challenge will be to design a digital twin that allows users to intervene, extract information and influence the system trajectory across time and space, as done – albeit often unwittingly – in the real world.”
(Bauer et al., 2021, p. 82)

According to the authors, the interpretation of the results of digital twins will not be confined to a “layer of experts” and anyone could use these results for his or her own purposes. This ambition reveals a certain conception of the potential and role of models. Models are seen as predictive tools, whose results have value in themselves. Another conception of the role of models could be opposed to this one. One could argue that the value of a model comes more from the interpretation one gives to its results, than from the results themselves: in this case, models are used as a guide to formulate and test hypotheses rather than as the truth. Such use of models requires a good understanding of their limitations: a certain expertise on the functioning and the limits of these models seems therefore necessary to correctly interpret their results. On the contrary, allowing anyone to use models as “black boxes” – as DestinE proposes with its digital twins – seems undesirable, or even dangerous. In addition to the technical debate about whether or not it is possible to create digital twins of the Earth, there is thus an epistemological debate behind DestinE: which kinds of models do we want for the future and how do we want to use them?

This question does not only concern DestinE. The modeling of atmospheric convection – and more generally, of climate – is today at a turning point. Parameterizations of atmospheric convection, which have been the dominant paradigm on which GCMs have been built for more than 50 years, are being questioned. In particular, some scientists propose to replace parameterizations with high-resolution simulations – in the form of GCRMs or super-parameterizations – or machine learning. These different options are opposed on technical grounds, but also on epistemological grounds and reflect different conceptions on the role of models. Clarifying the role and limitations of current GCMs was an important motivation behind my Ph.D. thesis. To continue this reflection, I propose in the following three epistemological guidelines for climate modelers in order to make the best use of numerical models to study climate.

Three epistemological guidelines for climate modelers

Numerical models have proven to be extremely useful for a number of applications, including weather forecasts and climate projections. Nevertheless, it is increasingly recognized in many scientific fields that models can be misused, sometimes with adverse policy consequences (see for instance Horner & Symons, 2020, for an overview of the risks associated with epidemiological models). To conclude this thesis, I propose here three guidelines that might help avoid misuse of models in climate research.

All models are not equivalent

In my thesis work, I focused on the convection parameterizations used in GCMs. I studied the basis of these parameterizations as well as the specific problems that their use poses in climate

models. I mentioned in Chapter 1 that the epistemological problems associated with convection parameterizations are very different from those associated with other parts of GCMs, e.g., their large-scale dynamics, or the parameterizations used to represent radiation or microphysics. The different components of a GCM involve different assumptions, justifications and methods, and have to be interpreted accordingly. Therefore, not all parts of climate models should be put in the same box.

The need to recognize the diversity of the models seems even more important when climate models are coupled with socio-economic models, as DestinE intends to do. In particular, we must be aware that we have an extremely valuable tool for simulating atmospheric and oceanic flows: the fundamental equations of thermodynamics and fluid mechanics. Thanks to such equations, we can simulate with a relative fidelity the general circulation of the atmosphere and the oceans, which allows us to forecast the weather with a fairly good accuracy. Unfortunately, there is no similar equations to describe social and economic interactions. There is therefore an epistemological gap between the part of models that is used to simulate atmospheric and oceanic flows, and the part of the models that integrate socio-economic activities.

I find it very regrettable that Bauer et al. (2021) do not highlight this fundamental distinction. On the contrary, Bauer, Stevens and Hazeleger claim that some of the methods that have proven useful for weather prediction, data assimilation in particular, will be equally useful for constructing a digital twin of the Earth integrating many other components. This gives the illusion that the problems we face in forecasting weather and in simulating our human activities are comparable. Unfortunately, this is not the case: simulating atmospheric and oceanic flows is radically different from simulating our human activities. Therefore, it remains to be demonstrated that the methods that have proved their worth for the former will be equally useful for the latter.

Assess and accept the limitations of your model

All models have some usefulness, but also some limitations. These limitations have to be carefully assessed in order not to give wrong interpretations of model results. There are many examples of misinterpretations of models in the literature, sometimes with direct political consequences. For instance, as documented by Sarewitz and Byerly (2000), a complex model composed of 286 sub-models and thousands of parameters had been built in the 1980s to evaluate the risk of disposing of radioactive waste at a certain site in the United States, the “Yucca Mountain repository”. The selection of this site suffered from many political, economic and regulatory controversies in the United States. In front of the many oppositions, decision makers relied on the results of the model to guarantee “one million years” of safety if the radioactive waste were buried at this location. Nevertheless, analyses then revealed that a single key variable – the so-called “percolation flux”, which determines the time it takes to water to percolate down to the underground repository – was uncertain by three orders of magnitude, leading to similar uncertainties on the ability of the site to keep waste safe over long periods of time.

This example shows how models can be instrumentalized to give an illusion of objectivity in political and economic decisions. To avoid overconfidence in a model, its limitations must be carefully assessed and taken into account when interpreting its results. We believe it is important to emphasize this point because, as DestinE’s ambitions illustrate, there are more and more applications for models. This can give the misleading impression that models can provide quantitative answers to any problem. On the contrary, I argue that qualitative reasoning

based on an in-depth analysis of certain key hypotheses seems in many situations preferable to quantitative results given by a model. We should not ask too much of models. As the computational resources available continue to grow, I think it is increasingly important for scientists to recognize that models are not systematically the right way to answer a question. When decision-makers ask for quantitative results, the best answer is sometimes to tell them: “the models will not give you the answer to this question”.

Don’t let your model become a black box

When a model is used as a black box, only its inputs and outputs are considered. The content of the model is not understood by the model user. Without such understanding, model users cannot assess the assumptions on which the model is based. This is problematic because some assumptions that are justified for a certain use of a model may seem absurd for other uses. To determine the legitimate uses of a model, as well as its limitations, it seems therefore necessary not to use a model as a black box.

Two characteristics of a model favor its use as a black box: its complexity, and the openness of its results to non-expert users. When a model is very complex, even those who helped build it may have difficulties understanding its content. Only expert users, who have a detailed understanding of the assumptions on which the model is based and its limitations, can make a relevant interpretation of model results. Therefore, the more complex a model, the higher the level of expertise required to analyze it. In practice, in climate science, the opposite is often the case. Simple models are typically analyzed by those who developed them, whereas more complex models, such as GCMs, are analyzed by users who were not involved in their development. I have noted and criticized this use of GCMs as black boxes in CMIP. Similar criticisms apply to DestinE, which proposes highly complex models with which decision makers can directly interact.

In contrast to such paradigms, I believe that simple models that are well-interpreted can be highly relevant to current research in climate science. During my thesis, I had the chance to explore the history of atmospheric and climate modeling. This led me to read numerous papers published at the dawn of computer age in the 1960s and 1970s. I was genuinely impressed by the level of understanding that emerged from these articles despite the means available at the time, and in particular the limited computer resources. The illuminating paper of Manabe and Wetherald (1967), which quantified the response of the atmosphere to a doubling of CO₂ concentration, is probably one of the best examples of how very simple models can provide valuable insights about the climate system and climate change. As we explained in Chapter 3, many results given in this paper are still relevant today. Manabe and Wetherald had a sufficient understanding of atmospheric phenomena first to include the main atmospheric processes in their model and make it no more complicated than necessary, and second, to give an insightful interpretation of its results. Inspired by such attempts, perhaps should we try to understand in depth the tools we already have, instead of building new tools that we don’t fully understand.

Appendix A

Publication list

Peer-reviewed research articles

Touzé-Peiffer, L., Barberousse, A., & Le Treut, H. (2020). The Coupled Model Intercomparison Project: History, uses, and structural effects on climate research. *Wiley Interdisciplinary Reviews: Climate Change*, 11(4), e648.

Madeleine, J. B., Hourdin, F., Grandpeix, J. Y., Rio, C., Dufresne, J. L., Vignon, E., ... & **Touzé-Peiffer, L.**, Bonazzola, M. (2020). Improved representation of clouds in the atmospheric component LMDZ6A of the IPSL-CM6A Earth system model. *Journal of Advances in Modeling Earth Systems*, 12(10), e2020MS002046.

Touzé-Peiffer, L., Albright, A. L., Fildier, B., Pincus, R., Vial, J., & Muller, C. (2021). Atmospheric radiative profiles during EUREC⁴A. *Earth System Science Data*, 13(2), 617-630.

Hwong, Y. L., Song, S., Sherwood, S. C., Stirling, A. J., Rio, C., Roehrig, R., ... & **Touzé-Peiffer, L.** (2021). Characterizing convection schemes using their responses to imposed tendency perturbations. *Journal of Advances in Modeling Earth Systems*, 13(5), e2021MS002461.

Stevens, B., Bony, S., Farrell, D., Ament, F., Blyth, A., Fairall, C., ..., **Touzé-Peiffer, L.**, ..., & Holanda, B. (2021). EUREC⁴A. *Earth System Science Data Discussions*, 1-78.

Rochetin, N., Hohenegger, C., **Touzé-Peiffer, L.**, & Villefranque, N. (2021). A physically-based robust definition of convectively generated density currents: detection and characterization in convection-permitting simulations (2021). Submitted to the *Journal of Advances in Modeling Earth Systems*.

Touzé-Peiffer, L., Vogel, R., & Rochetin, N. (2021). Detecting cold pools from soundings during EUREC⁴A. submitted to the *Journal of Applied Meteorology and Climatology*.

Book & outreach articles

Touzé-Peiffer, L., Morel, P. (2018). Le climat : de l'observation à la modélisation – Brève histoire d'une épopée scientifique et technologique, *Éditions Matériologiques*.

Sommeria, G., **Touzé-Peiffer, L.** (2019). Le programme mondial de recherche sur le climat fête ses 40 ans, *La Météorologie*, 107, 10-13, 2019.

Data sets

Schulz, H., **Touzé-Peiffer, L.**, and Fildier, B.: EUREC4A GOES-16 ABI overview movies (VIS+IR), <https://doi.org/10.25326/225>, Aeris, 2021.

Touzé-Peiffer, L., Fildier, B., and Albright, A. L.: Atmospheric radiative profiles during EUREC4A, <https://doi.org/10.25326/78>, Aeris, 2021.

Open source codes

Fildier, B., **Touzé-Peiffer, L.**, and Schulz, H.: bfildier/EUREC4A_movies: v1.0.0, <https://doi.org/10.5281/zenodo.4777954>, zenodo, 2021.

Fildier, B., **Touzé-Peiffer, L.**, and Albright, A. L.: bfildier/Albright2020: v1.0, <https://doi.org/10.5281/zenodo.4010195>, zenodo, 2021.

Appendix B

Detecting cold pools from soundings during EUREC⁴A

B.1 Preamble

This chapter is based on a study on cold pools during EUREC⁴A that I carried out in collaboration with Raphaëla Vogel and Nicolas Rochetin. I wrote the first draft of the manuscript and performed most of the analyses. Raphaëla Vogel and Nicolas Rochetin guided me through the design and writing of the manuscript, and gave me extensive feedback before submission and during the review process.

Cold pools are one of the important object that can be studied with EUREC⁴A data. In the following study, we give a first look at cold pools during EUREC⁴A by using a novel method to detect cold pools from atmospheric soundings. This study illustrates how objects like cold pools give us a tool to analyze and understand observations. I therefore used it as a case study in my epistemological analysis of objects in Chapter 2. For the sake of clarity, I have cited only the results most relevant to our epistemological reflections in Chapter 2 and I give here the full study.

The corresponding paper has been submitted to the *Journal of Applied Meteorology and Climatology* in December 2020 and can be cited as: Touzé-Peiffer, L., Vogel, R., & Rochetin, N. (2021). Detecting cold pools from soundings during EUREC⁴A. *arXiv preprint* (<https://arxiv.org/abs/2104.09146>).

B.2 Abstract

This paper develops a novel method to detect cold pools from atmospheric soundings over tropical oceans and applies it to sounding data from the EUREC⁴A field campaign, which took place south and east of Barbados in January-February 2020. The proposed method exploits the fact that the air in a cold pool is denser than the air above it. It leads us to define cold pool soundings as those for which the mixed-layer height is smaller than 400 m. We first test this criterion by verifying its consistency with surface temperature and precipitation in a realistic high-resolution simulation over the western tropical Atlantic. Applying it to EUREC⁴A data, we then identify 7 % of EUREC⁴A dropsondes and radiosondes as cold pool soundings. In two selected case studies, we find that cold pools soundings coincide with mesoscale cloud arcs and temperature drops in the surface time series. Statistics for the entire campaign further

characterize the signature of cold pools in temperature, humidity and wind profiles. In the presence of wind shear, we show in particular that the spreading of cold pools is favored downshear, suggesting downward momentum transport by unsaturated downdrafts. These results support the robustness of our simple method in different environmental conditions and illustrate the new insights it offers for the characterization of cold pools and their environment.

B.3 Introduction

Below clouds, the partial evaporation of precipitation may cool the air sufficiently to generate unsaturated downdrafts, which spread horizontally when reaching the surface under the form of density currents. As these so-called "cold pools" expand, they lift the warmer adjacent air masses, leading in some situations to the creation of new convective cells (e.g. Craig Goff, 1976; Warner et al., 1979). This triggering has been observed to be particularly effective when two cold pools collide (Droegemeier & Wilhelmson, 1985; Feng et al., 2015; Meyer & Haerter, 2020) or when the vorticity created by a cold pool counteracts that from the low-level wind shear (Rotunno et al., 1988; Weisman & Rotunno, 2004). In addition, since they induce strong gusts near the surface, cold pools are suspected to enhance surface fluxes, and thus to modify the properties of the subcloud layer (SCL) (Langhans & Romps, 2015; Tompkins, 2001). For all these reasons, cold pools are thought to play an important role for the organization and the propagation of convection (Kurowski et al., 2018; Schlemmer & Hohenegger, 2014; Tompkins, 2001).

Cold pools are ubiquitous in regions of precipitating convection, both over lands and oceans (Bryan & Parker, 2010; Zuidema et al., 2017). Most past attempts to detect cold pools in observational data used methods based on surface time series. Analyzing data from the Tropical Ocean Global Atmosphere (TOGA) Coupled Ocean-Atmosphere Response Experiment (COARE; Webster & Lukas, 1992), Young et al. (1995) for instance defined the beginning of a cold pool time period through the onset of any rain shaft of at least 2 mm h^{-1} and the end of a cold pool time period by the end of the subsequent surface temperature recovery. During the Rain in Cumulus over the Ocean campaign (RICO; Rauber et al., 2007), which took place in the eastern Caribbean from December 2004 to January 2005, Zuidema et al. (2012) used a similar method but did not impose any threshold on surface rain rate, in order to consider all precipitating events. More recently, de Szoeke et al. (2017) detected cold pools over the central Indian Ocean during the Dynamics of the Madden-Julian Oscillation experiment (DYNAMO; Yoneyama, Zhang, & Long, 2013) as temperature drops in the surface time series measured by the research vessel Roger Revelle. A slightly modified version of this method was used in Vogel (2017) to detect cold pools from 2011 to 2017 at the Barbados Cloud Observatory (BCO), a site exposed to relatively undisturbed trade-winds on the eastern side of Barbados (Medeiros & Nuijens, 2016; Stevens et al., 2016).

Surface measurements from research vessels or weather stations are useful to get data with high temporal resolution at a single point. They give precise information on the surface characteristics of cold pools, such as the wind shifts they induce, the amplitude of the associated temperature drops, their effects on surface fluxes, etc. However, they say little about their vertical structure, in particular their height, which is recognized as an important parameter to describe the triggering of new convective cells near cold pool edges (Grandpeix & Lafore, 2010; Jeevanjee & Romps, 2015). The distribution of temperature and moisture in and above cold pools is also poorly known (Zuidema et al., 2017), as is the distribution of horizontal and vertical winds. In a study of shallow marine cumulus convection, using large-scale simula-

tions, C. Bretherton and Blossey (2017) found mesoscale convergence under precipitating cloud clusters and associated circulations that could explain the mesoscale aggregation of shallow convection under these conditions. Detecting similar circulations above cold pools to test this hypothesis in observations remains a challenge.

The field study EUREC⁴A (Elucidating the role of clouds-circulation coupling in climate; Bony et al., 2017) might help overcome some of these difficulties. EUREC⁴A was held in January and February 2020 over the Atlantic ocean east and south of Barbados (Stevens et al., 2021) and involved a wide variety of platforms: among others, four research aircraft, four research vessels, surface observatories and a battery of uncrewed aerial and seagoing systems. Aircraft and research vessels supported the implementation of a large-scale sounding array. In total, more than 2000 atmospheric profiles were measured during EUREC⁴A using radiosondes and dropsondes. This unprecedented dataset represents a unique opportunity to study the properties of cold pools over tropical oceans, provided a robust method to detect cold pools from atmospheric soundings exists.

The lack of such a method motivates the present study, which proposes a detection method of cold pools based on the height of the mixed layer and applies it to EUREC⁴A data. The paper is structured as follows: first, in Section B.4, we describe the EUREC⁴A radiosonde and dropsonde data used in this study, then, in Section B.5, we present our detection method of cold pools and test it in a high-resolution simulation. Finally, we assess our detection method with EUREC⁴A data and use it to provide a first analysis of cold pools during the field study.

B.4 Radiosonde and dropsonde data

During EUREC⁴A, atmospheric soundings were released in two different regimes: the “Tradewind Alley” and the “Boulevard des Tourbillons” (Stevens et al., 2021). The Tradewind Alley is a corridor up to 50°W east of Barbados, extending from approximately 11°N to 17°N. In January and February, free-tropospheric subsidence prevails in these latitudes and the Tradewind Alley can thus be seen as a trade-wind regime, with shallow convective clouds under a capping inversion (Stevens et al., 2016). The Boulevard des Tourbillons is a corridor further south extending from the northern coasts of Brazil to Barbados. It was initially chosen to characterize large-scale oceanic eddies observed along the coast of South America and their impact on air-sea interactions (Bony et al., 2017; Stevens et al., 2021). Located in areas of deeper convection, the Boulevard des Tourbillons is also an opportunity to extend the EUREC⁴A sounding array down to the Intertropical Convergence Zone.

In the Tradewind Alley, both radiosondes and dropsondes were used to characterize the trade-wind environment. Radiosondes were launched from land at BCO (13.16°N, 59.43°W), and from two research vessels – the German research vessel Meteor, cruising between 12-14.5°N at a fixed longitude of 57°W, and the American research vessel from NOAA Ronald H. Brown (RH-Brown), moving along transects further to the east. Two research aircraft, the German HALO and the WP-3D Orion (P-3) from NOAA, complemented this set of measurements by launching dropsondes. HALO flew at an altitude of about 10 km along the “EUREC⁴A circle”, a circular path with approximately 200 km diameter centered at 13.3°N, 57.7°W. When launching dropsondes, the P-3 flew at about 7 km along both linear and circular patterns around the EUREC⁴A circle, as well as further to the east close to the nominal position of the RH-Brown. No dropsondes were launched in the Boulevard des Tourbillons, but radiosondes

were launched from the French research vessel L’Atalante and from the German research vessel Maria S. Merian (MS-Merian).

Equipped with a GPS receiver, all these radiosondes and dropsondes measured the pressure, temperature, humidity and horizontal wind along their fall or ascent. For radiosondes equipped with parachutes, we count the ascent and the descent as two separate profiles. In this study, we use quality controlled data (Level-3) interpolated on a common altitude grid with bin sizes of 10 m (George et al., 2021; Stephan et al., 2020). In addition to the filters already used in the initial dataset, we are adding a filter to keep only soundings with at least 30 measurements of temperature, pressure and humidity below 500 meters. This filter is needed to get enough data in the lowest atmospheric layers to distinguish cold pools from their environment. It mainly affects descending profiles from radiosondes as their signal was sometimes lost in the first hundreds of meters above sea level due to Earth’s curvature (Stephan et al., 2020). It removes in total 441 radiosonde profiles, resulting in an input dataset set with 1068 atmospheric profiles from dropsondes and 1106 from radiosondes.

B.5 Presentation of the method and test in a high-resolution simulation

Most methods which have been proposed to detect cold pools from observations used the fact that the air inside a cold pool is colder than the air *around* it. The present detection method uses the fact that cold pool air is colder than the air *above* it. In terms of virtual potential temperature, a cold pool is indeed colder, thus denser than the SCL air on top of it. Consequently, a sharp increase in virtual potential temperature θ_v is expected at the top of a cold pool. Over tropical oceans, in convective regimes, this contrasts with the SCL outside cold pools, which tends to be well-mixed in θ_v up to cloud base (e.g. Cuijpers & Duynkerke, 1993; Pennell & LeMone, 1974). It motivates the use of a cold pool detection method based on the height of the mixed layer (H_{mix}).

Following Canut et al. (2012) and Rochetin et al. (2021), we define H_{mix} as the lowest altitude Z above $Z_{min} = 100$ m where the virtual potential temperature θ_v is higher than its mass-weighted average from Z_{min} to Z by a fixed threshold $\epsilon = 0.2$ K:

$$\begin{aligned} \theta_v(Z) &\geq \tilde{\theta}_v + \epsilon \\ \text{with } \tilde{\theta}_v &= \frac{\int_{Z_{min}}^Z \rho(z) \theta_v(z) dz}{\int_{Z_{min}}^Z \rho(z) dz} \end{aligned} \quad (\text{B.1})$$

ρ being the density of the air. Setting Z_{min} at 100 m is necessary due to the presence of unphysical temperature peaks below 100 m for a few radiosondes. The virtual potential temperature is calculated assuming that the air of the lowest layers is not saturated, so that the mixing ratio of liquid water in the air can be neglected. It is then approached as: $\theta_v = \theta(1 + 0.61r)$, r being the mixing ratio of water vapor. The calculation of H_{mix} thus requires only the vertical profiles of pressure, temperature and humidity at a single point and is directly applicable to soundings.

Figure B.1 shows the H_{mix} distribution for EUREC⁴A soundings. The histogram reveals a negatively skewed distribution, with a median of 720 m. Assuming that the left tail of the

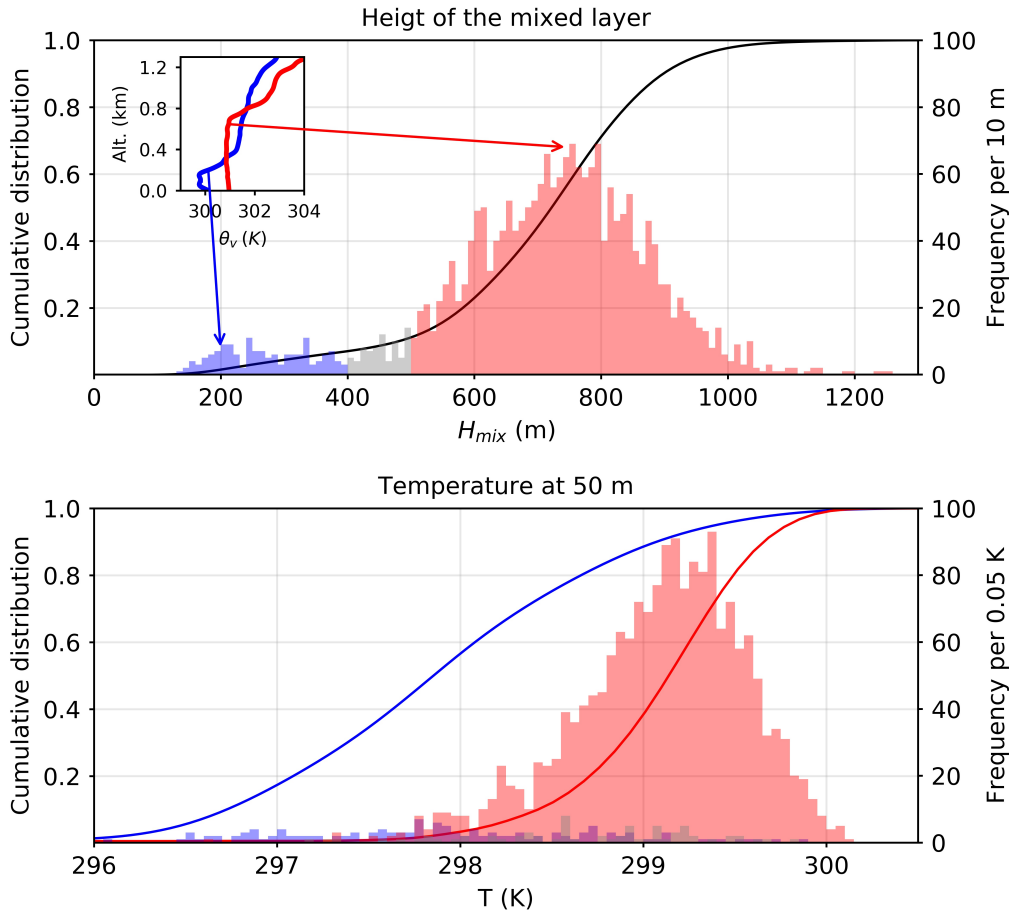


Figure B.1: (top) Cumulative distribution function (black) and histogram for the height of the mixed layer for all radiosondes and dropsondes launched during EUREC⁴A. The bins of the histogram are 10 m wide. Colors indicate cold pool soundings (blue), unclassified soundings (grey) and environmental soundings (red). The upper left panel shows two examples of θ_v profiles, one in a cold pool and the other in the environment. (bottom) Cumulative distribution functions for cold pool (blue) and environmental (red) temperatures at 50 meters, and the corresponding histograms with a bin width of 0.05 K.

distribution is due to cold pools, we choose to define “cold pool soundings” as those with H_{mix} lower than 400 m (7% of soundings, in blue), and “environmental soundings” as those with H_{mix} higher than 500 meters (90% of soundings, in red). With this definition, only 3% of soundings (in grey) are neither in cold pools nor in the environment. The blue part of the distribution peaks slightly above 200 m, consistent with the typical depth of cold pools observed by Zuidema et al. (2012) in trade-wind regimes. Since this value is more than three times smaller than the typical mixed layer height in the region (Fig. B.1), the mixed layer height seems to be a suitable criterion to distinguish cold pools from their environment during EUREC⁴A. In the cold pool recovery process, the height of the mixed layer gradually increases due to surface sensible and latent heat fluxes and entrainment warming and drying from above, as documented by Zuidema et al. (2012) during RICO. It could explain the large spread of the H_{mix} distribution below 400 m (in blue) and its smooth increase from 400 to 500 m (in grey).

The surface temperature distribution (bottom panel, Fig. B.1) reveals that cold pool sound-

ings are on average colder than environmental soundings. Nevertheless, a fixed surface temperature threshold – as done for instance in de Szoeke et al. (2017) with radiosondes launched during DYNAMO – would not be able to isolate cold pools. Indeed, given the spatial and temporal extent of EUREC⁴A, cold pools are not the only source of variability for surface temperature. In particular, measurements from research vessels reveal a sea surface temperature (SST) difference of more than 1 K between the northeast (50°W, 16°W) and the southwest (12°N, 59.5°W) of the Tradewind Alley during EUREC⁴A. Mesoscale variability in SST features was also observed in the Boulevard des Tourbillons during the field study (Stevens et al., 2021). Due to this variability, the surface temperature distributions for cold pool and environmental soundings overlap (as shown in Fig. B.1) and do not allow to clearly separate cold pools from their environment.

To show that H_{mix} provides on the contrary a robust detection of cold pools, we first test our detection method in a high-resolution simulation over the Atlantic Ocean, upstream of Barbados. This simulation was conducted using the large-eddy simulation version of the ICOsahedral Non-hydrostatic (ICON) model, with realistic boundary conditions, 313-m horizontal grid spacing and 150 vertical levels (Dipankar et al., 2015; Zängl et al., 2015). It was initialized on 11 December 2013 at 0900 UTC with 1.25km-resolution runs (see Vial et al., 2019a, for further details).

In high-resolution models, most cold pool detection methods rely on anomalies in surface temperature (or related quantities such as potential temperature or buoyancy), winds, or surface rain (see Drager & van den Heever, 2017, and references therein). To determine cold pool edges, they usually requires the entire 2 or 3D field of some thermodynamic or dynamic variables and are therefore unusable with sounding data. On the contrary, our detection method is based on 1D thermodynamic profiles and is thus entirely local.

Despite this simplicity, Figure B.2 shows that it gives consistent results with the surface temperature and precipitation fields. It represents three snapshots of the mixed layer height H_{mix} (left) and the temperature near the surface ($z \approx 50$ m) (right) on 11 December 2013 at 1600 UTC and on 12 December 2013 at 0330 UTC and 0930 UTC. In the right panels, we see that regions with negative temperature anomalies are co-located with significant surface precipitation (red dots) and also with important wind shifts (not shown), suggesting that these cold regions are in fact convective cold pools. In the left panels, we apply our detection method by circling in yellow regions where the mixed layer is less than 400 meters. Qualitatively, there is a really good agreement between cold pools detected with our method and regions strongly cooled by rainfall. H_{mix} is unaffected by large-scale temperature gradients and seems to discriminate cold pools everywhere in the domain. Conversely, precipitation alone is not a distinguishing factor of cold pools, as it is generally present only over a small part of cold pools, and any method based on a temperature threshold would likely be strongly sensitive to the temperature gradient between the northeast and the southwest of the domain. This gives us confidence to choose H_{mix} in the next section to detect cold pools from soundings during EUREC⁴A.

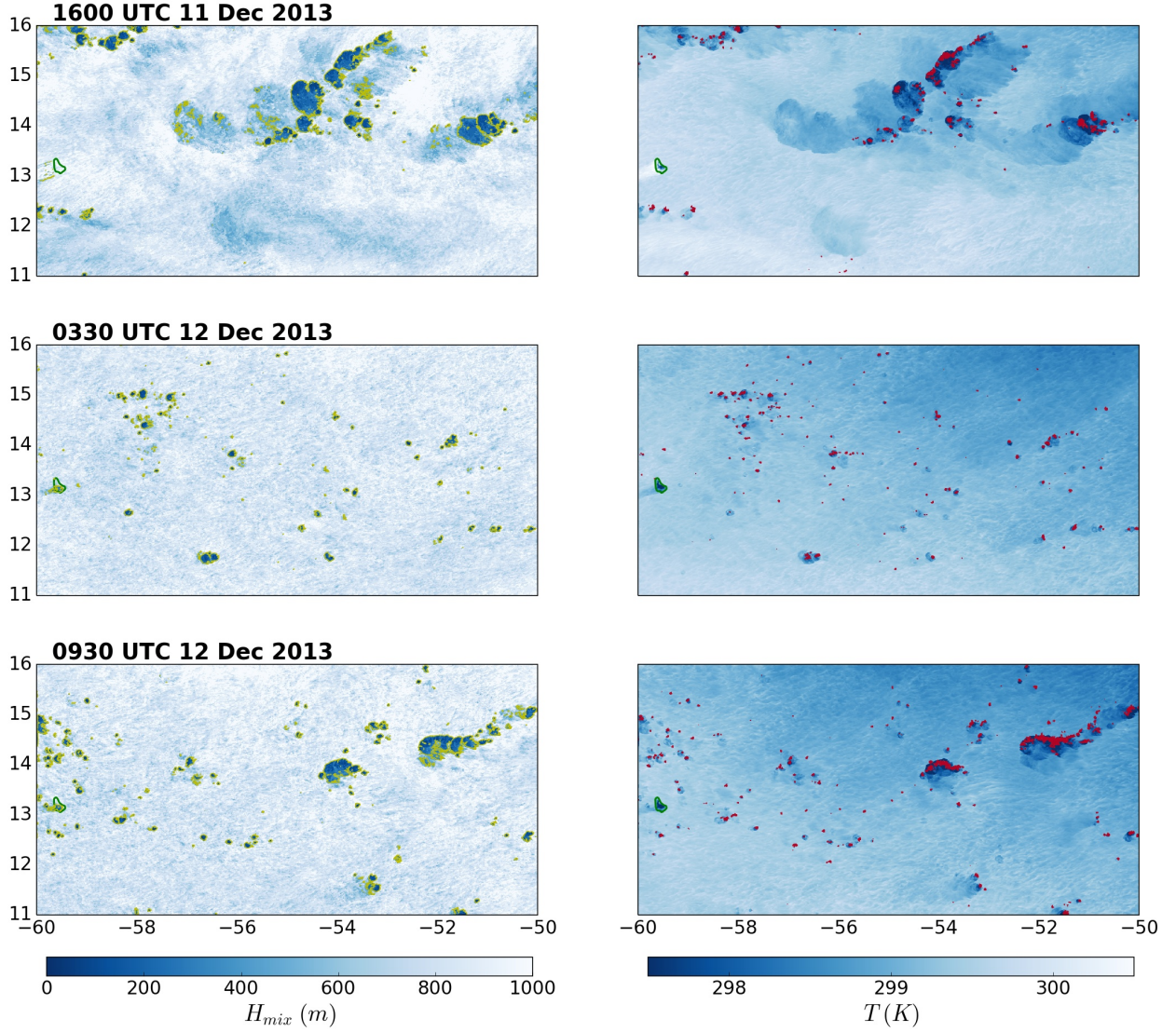


Figure B.2: Height of the mixed layer H_{mix} and surface temperature T at $z \approx 50$ m in an area of $5^\circ \times 10^\circ$ upstream of Barbados (circled in green) on December 11, 2013 at 1600 UTC and December 12, 2013 at 0330 UTC and 0930 UTC. In the left-hand panels, cold pools are circled in yellow as regions where H_{mix} is lower than 400 meters. In the right-hand panels, the red dots represent locations where surface rainfall is greater than 10 mm day^{-1} .

B.6 Application to EUREC⁴A data

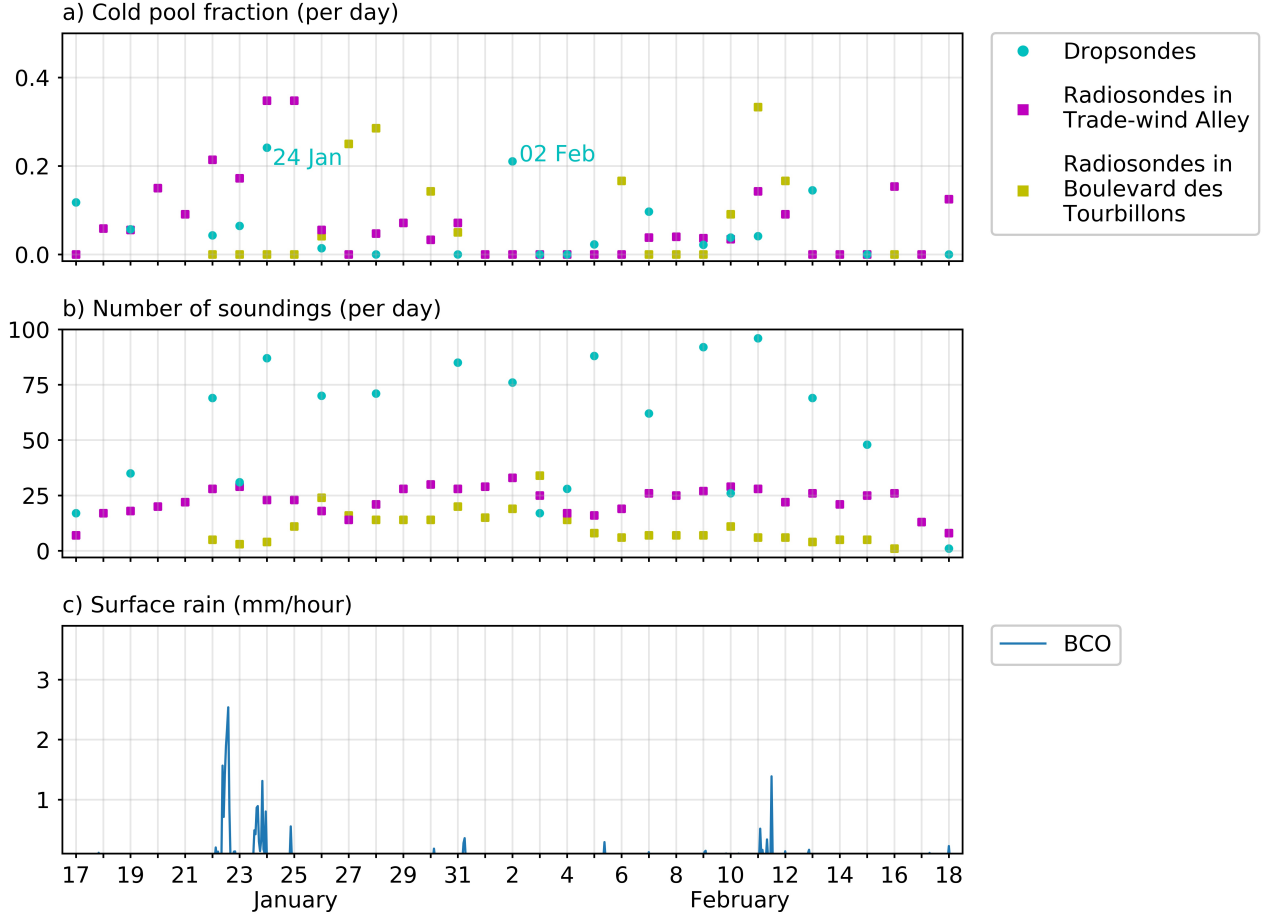


Figure B.3: (a) Cold pool fraction per day from dropsondes (cyan circles), and from radiosondes launched in the Trade-wind Alley (magenta squares) and in the Boulevard des Tourbillons (yellow squares). The two days selected for the case studies are indicated in cyan. (b) Number of soundings per day, with the same color code. (c) Hourly rain rate at BCO from the micro-rain radar at 325 m height (see Stevens et al., 2016, for details).

Applying our method to our set of 2174 EUREC⁴A soundings, we find that 149 sondes have fallen into cold pools. Figure B.3 shows (a) the cold pool fraction, (b) the number of soundings per day and (c) the hourly surface rain rate measured at BCO. As expected, we see some consistency between rain rates measured at BCO and cold pool fraction in dropsondes and radiosondes in the Trade-wind Alley, in particular around January 22-25 and February 11-12, when significant rainfall rates were measured at BCO. For the dropsondes, the largest cold pool fractions have been observed on January 24 and February 2. In the following, we will examine the cold pool characteristics on these two days in more detail.

B.6.1 January 24

On January 24, more than 20 % of dropsondes (18 out of 88, see Fig. B.3) launched by HALO fell into cold pools. The MODIS-Terra satellite image of that day (Fig. B.4) reveals the presence of many mesoscale arcs. As noted by previous observational and modeling studies

of trade-wind regimes (Seifert & Heus, 2013a; Zuidema et al., 2012), such mesoscale arcs are likely due to cold pools. Indeed, while the dense air inside cold pools is less able to support buoyancy-driven convection, it has long been known that the edges of an expanding cold pool can lift surrounding air masses and trigger new convective cells (e.g. Knupp & Cotton, 1982; Purdom, 1976; Weisman & Rotunno, 2004). Based on these previous studies, we expect relatively few clouds above cold pools, but mesoscale arcs of clouds near their border. The presence of many mesoscale arcs around the EUREC⁴A circle is thus consistent with the numerous cold pools detected that day from the HALO dropsondes.

Superposing the launch positions of dropsondes to GOES-16 visible channels reveals the cloud field dropsondes have actually sampled. Panel b (yellow box) shows the position of dropsondes launched by HALO when passing over a mesoscale arc in the northeastern part of EUREC⁴A circle, just before 1400 UTC. The bottom panels (d-e) give the potential temperature and zonal wind speed measured by these dropsondes. The cold pool sounding (blue) is the coldest, with the strongest zonal wind near the surface and the shallowest mixed layer. In terms of potential temperature, H_{mix} and zonal wind, the unclassified sounding (grey) represents an intermediate case between the cold pool and the environmental (red) soundings. Located inside the mesoscale arc, but closer to its edges than the cold pool sounding, it suggests that the mixed layer may partially recover its initial properties near the edge of a cold pool.

Panel c (cyan box) gives a second example of a cold pool sounding. In this example, due to the presence of many clouds on the satellite image, it is hard to say whether the cold pool sounding fell in a cloud arc or not. Our detection method seems nevertheless consistent with wind measurements, which reveal the presence of a cold outflow at the surface (with a difference of more than 3 m s^{-1} in near-surface zonal velocity between the blue and red soundings). Note that in this example, the surface temperature does not distinguish cold pools, as the blue sounding (1419 UTC) in panel c is less than 0.5 K colder than the neighboring red soundings (1415 and 1424 UTC), and slightly warmer than the red sounding in panel b (1400 UTC).

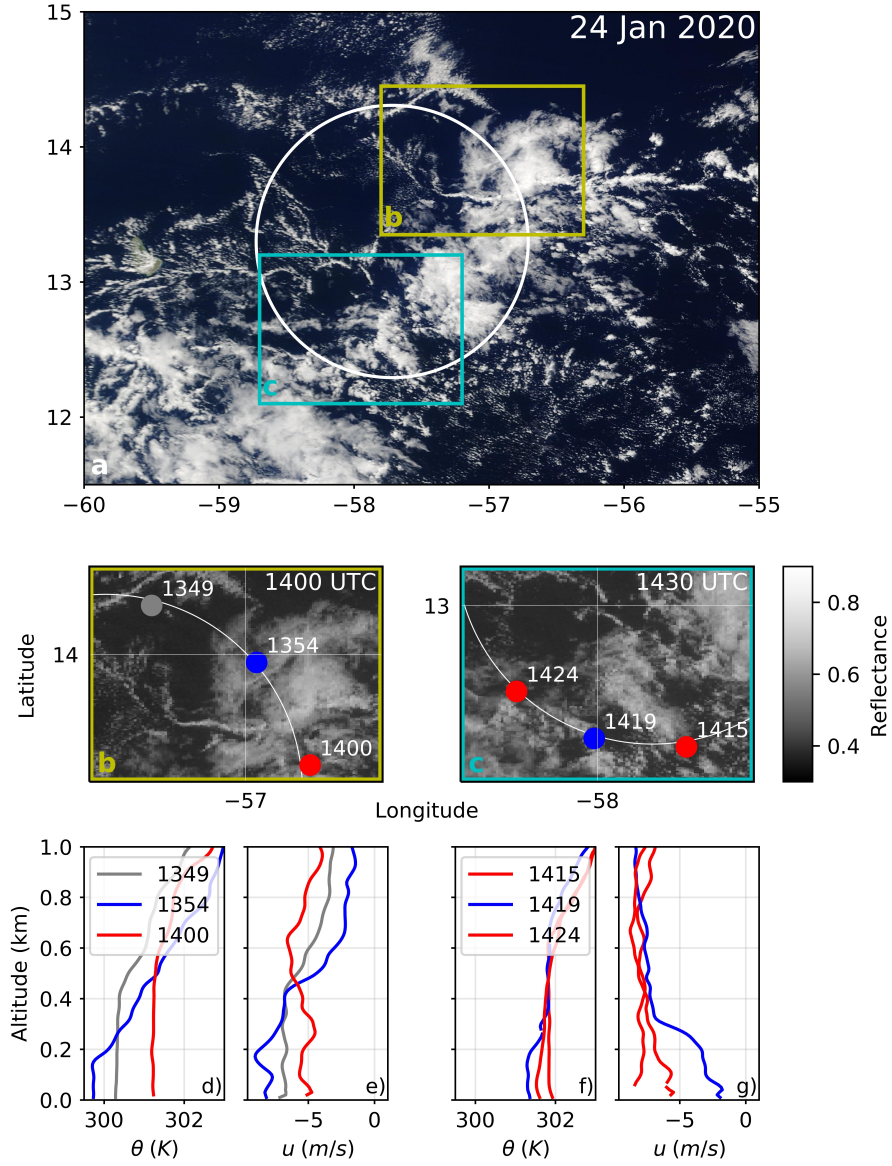


Figure B.4: (top) MODIS-Terra scene from Worldview upstream of Barbados on January 24, 2020. The EUREC⁴A circle is indicated in white. The rectangles mark the two regions shown in the lower panels. (middle) GOES-16 visible reflectance (channel 2) displayed at 1400 UTC (b) in the northeast quarter of the EUREC⁴A circle and at 14:30 UTC (c) in the southwest part of the circle. The ground position and launch time of dropsondes dropped in the 15 minutes preceding the satellite image is shown in blue for cold pool soundings, in red for environmental soundings and in grey for unclassified soundings. (bottom) Profiles of potential temperature (d-f) and zonal wind speed (e-g) from the surface to 1 km for the dropsondes highlighted in the middle panel.

B.6.2 February 2

On February 2, the satellite images reveal a cloud field with many “flowers”, characterized by the presence of mesoscale, quasi-circular, stratiform shallow clouds, capped by a strong inversion and separated by very dry clear-sky areas following the classification introduced in Stevens, Bony, et al. (2019). At 1700 and 1850 UTC, in the eastern part of EUREC⁴A circle, a mesoscale cloud arc is visible in all directions around one of the flowers. As on January 24, we suspect this cloud arc to be due to the presence of a cold pool at the surface. This claim is supported by the surface meteorological data from the research vessel Meteor, which captured the onset of the corresponding cold pool around 1110 UTC (see Fig B.5, panel a). Indeed, as shown in Fig. B.6, the surface air temperature (a) measured by the Meteor at 1110 UTC suddenly dropped by approximately 1.5°C, suggesting that the research vessel entered a cold pool. This drop is also visible in the virtual potential temperature (c). The ship remained in the cold pool until 1230 UTC when the temperature rose to 26.2°C, a slightly higher value than before the passage of the cold pool. The wind speed (d) changed significantly as the ship passed through the cold pool front: the absolute wind speed increased rapidly from 5 m s⁻¹ to 10 m s⁻¹ at the time of the temperature drop, before decreasing continuously to 2 m s⁻¹ and going back to its pre-cold pool value of 5 m s⁻¹ after the cold pool passed. The wind speed variations upon entering and leaving the cold pool thus have approximately the same amplitude. These results stand both for the meridional and the zonal wind, although the wind shifts have slightly stronger amplitudes for the zonal wind. The specific humidity (b) and the equivalent potential temperature (d) reveal that this cold pool is overall moister than its environment, especially near its edges, which are about 1 g kg⁻¹ moister than its core. Finally, the W-band radar (e) shows 2.5 km deep clouds with strong reflectivity near the edges of the cold pool. When the ship enters the cold pool, these clouds coincide with a high surface rain rate of more than 4 mm h⁻¹ (f), so we suspect the evaporation of precipitation below these clouds to feed this cold pool. Overall, these different features are consistent with previous observations of cold pools over tropical oceans (de Szoeke et al., 2017; Zuidema et al., 2012) and support the cold pool tagging shown in Fig. B.5.

Of the 100 dropsondes and radiosondes launched near the EUREC⁴A circle that day, 16 were detected falling in cold pools. 13 of them fell in or around the cloud cluster passing over the Meteor at 1110 UTC. The three other cold pool soundings all fell in or around other cloud clusters, located further east of the EUREC⁴A circle (visible for example at 1700 UTC in Fig. B.5). Mean sounding profiles in (blue) and out (red) of cold pools are shown in the bottom panels (e-g). Cold pool soundings are on average 1 K colder than the other soundings in terms of potential temperature. This difference is mainly observed below 400 meters. The air in the cloud layer above cold pools, from 800 m to 3 km, is significantly moister than the ambient air, consistent with the presence of cloud clusters on satellite images. In the SCL, we observe on the contrary slightly drier air between 200 meters and 800 meters in cold pools compared to their environment, which might be due to downdrafts transporting the relatively drier air from the cloud layer to the subcloud layer above cold pools. On this day, near-surface air in cold pools is on average moister than in their environment, consistent with Meteor observations. At the surface, cold pools propagate as density currents and induce significant wind shifts, visible in the relatively large standard deviation of the near-surface wind in cold pools (g). The consistency between satellite images, Meteor data and the thermodynamic profiles derived from soundings support the ability of our detection method to distinguish cold pools from their environment on February 2.

This robust distinction between cold pool and environmental soundings allows the study of not only the properties of cold pools, but also those of the convective system to which they belong. For instance, the longevity of the cloud cluster first detected at 1110 UTC by the Meteor (Fig B.6) and still visible on satellite images at 1850 UTC (Fig B.5) suggests intense convective activity above the cold pool detected in the southern part of the EUREC⁴A circle. To characterize the associated air circulations, we compute the divergence in and above this cold pool, and compare it to the divergence calculated over the EUREC⁴A circle from environmental soundings. For that, we use the method introduced by Lenschow and Sun (2007) and successfully tested during the Next Generation Aircraft Remote Sensing for Validation Studies (NARVAL2, Bony & Stevens, 2019). This method assumes that the wind field is stationary and that wind variations in longitude and latitude are linear at each vertical level, defining thus the large-scale wind $\mathbf{V} = (u, v)$ such that:

$$\mathbf{V} = \mathbf{V}_0 + \frac{\partial \mathbf{V}}{\partial x} \Delta x + \frac{\partial \mathbf{V}}{\partial y} \Delta y \quad (\text{B.2})$$

where \mathbf{V}_0 is the mean velocity over the area and Δx and Δy are the eastward and northward displacements from a chosen center point. Since \mathbf{V} , \mathbf{V}_0 , Δx and Δy are known, $\frac{\partial \mathbf{V}}{\partial x}$ and $\frac{\partial \mathbf{V}}{\partial y}$ can be calculated using a simple least square fit in equation B.2 and the divergence can then be derived as: $D = \partial_x u + \partial_y v$.

As in Bony and Stevens (2019), for HALO dropsondes out of cold pools, we choose as center point the center of the EUREC⁴A circle, indicated by a red cross in Fig. B.5. For cold pool soundings, the center point is taken at the center of the cold pool visible in the southern part of the EUREC⁴A circle. To estimate the center location, we take as a starting point the position of the Meteor where the minimum temperature was measured and advect it by the surface mean wind speed measured by the research vessel in a four-hour time slot centered on the cold pool period. In Fig. B.5, the position of the estimated “cold pool center” is indicated in each snapshot by a blue cross. We see that it matches approximately the center of mesoscale arcs observed at 1700 and 1850 UTC. With respect to, respectively, the center of the EUREC⁴A circle and the estimated center of the cold pool, we then compute the least square fit for the 29 dropsondes dropped in the EUREC⁴A circle out of cold pools from 1420 to 1845 UTC, and for the 11 dropsondes dropped in the main cold pool during the same period. This period is long enough to have a sufficient number of dropsondes to get a robust measure of divergence in and out of cold pools, but short enough to consider the stationarity assumption as valid at least out of cold pools, as the autocorrelation time-scale of large-scale divergence estimated by Bony and Stevens (2019) is approximately 4 hours. Assuming stationarity of the wind field above the cold pool is more controversial, but seems nevertheless a good approximation in this precise situation due to the duration of the convective system and the coherence between dropsonde measurements at different times.

The blue curve in Fig. B.5 h reveals a layer of strong divergence near the surface in the cold pool, culminating at more than $80 \times 10^{-6} \text{ s}^{-1}$ close to the ground, consistent with the spreading of the cold pool at the surface. On the other hand, from 500 m to 1.5 km, we observe a zone of convergence peaking slightly above 1 km with a value close to $-80 \times 10^{-6} \text{ s}^{-1}$. This converging air might feed convective updrafts or downdrafts near cloud base. Finally, between 2 and 3 km, there is again a divergent layer, with a magnitude of about $80 \times 10^{-6} \text{ s}^{-1}$. It corresponds to the altitude where the updraft air is detrained and where the stratiform outflow layer is located, though the divergence profile close to the inversion might also be influenced by local circulations induced by mixing with drier free tropospheric air. In contrast, the divergence plot out of

cold pools has a much smaller amplitude from the surface to the inversion. As expected, above the inversion, the divergence patterns in and out of cold pools follow each other closely (not shown), suggesting homogeneous large-scale wind patterns in the free troposphere, whether or not there are cold pools below.

The mesoscale divergence values above the cold pool are two or three times larger than the maximum large-scale divergence observed by Bony and Stevens (2019) during NARVAL2 over 180 km diameter circles. When averaged over a EUREC⁴A circle, they could therefore leave an imprint on the large-scale divergence, as shown by the black line which represents the divergence computed over a EUREC⁴A circle from 1841 to 1936 UTC using all dropsondes (12 in total, including 4 dropped in cold pools). The orange line shows the divergence calculated over the same circle with environmental soundings only. The comparison between the orange and black curves shows local differences of more than $10 \times 10^{-6} \text{ s}^{-1}$ between the two curves. The largest differences are observed between 500 m and 2 km, with smaller differences also present near the surface and just below the inversion, at 3 km. The differences between the black and orange curves are consistent with the differences between the blue and red ones, albeit of opposite sign: indeed, as the cold pool center is located outside the EUREC⁴A circle at 1850 UTC (see the blue cross on Fig. B.5, panel d), convergence over the cold pool at that time leads to divergence over the EUREC⁴A circle and vice versa. This example shows that mesoscale circulations around cloud clusters such as the one observed here could explain part of the variability of the large-scale divergence in the SCL and cloud layer noted in Bony et al. (2017).

In large-eddy simulations of marine shallow cumulus convection, C. Bretherton and Blossey (2017) found circulations similar to those observed around the main flower on February 2. According to the authors, these circulations might participate in the mesoscale aggregation of shallow convection below the trade-inversion. Indeed, in trade-wind regimes, C. Bretherton and Blossey (2017) show that such circulations induce relative moistening of the moistest subdomains, a form of gross moist instability. This hypothesis, if verified, would give a possible explanation for the persistence for several hours of the cloud clusters on February 2 despite the presence of cold pools below them.

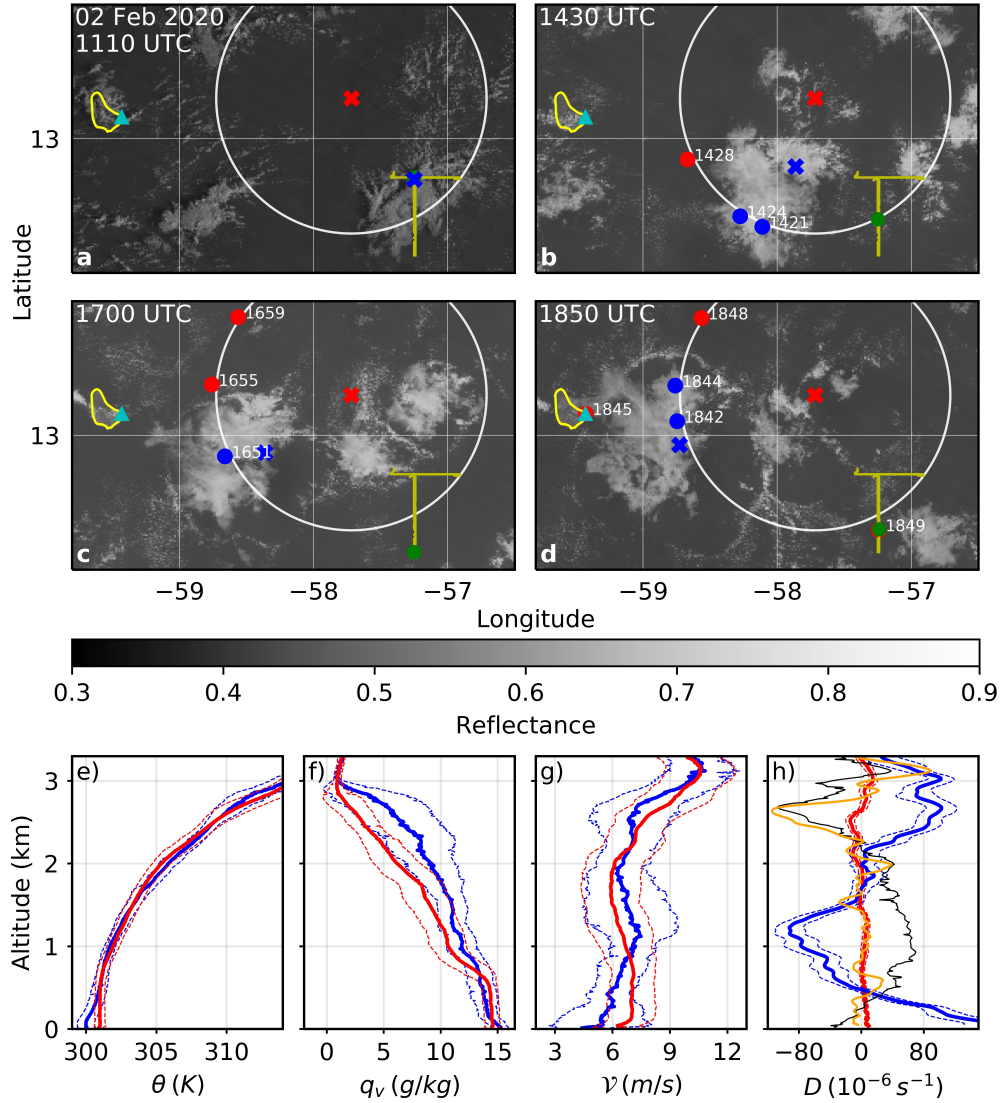


Figure B.5: (top) GOES-16 visible reflectance (channel 2), displayed on February 2 at (a) 1130, (b) 1430, (c) 1700 and (d) 1850 UTC in the Atlantic region upstream Barbados (circled in yellow). The EUREC⁴A circle is outlined in white. The ground position and launch time of dropsondes dropped in the 15 minutes preceding the satellite image is shown in blue for cold pool soundings and in red for environmental ones. The yellow line represents the path of the Meteor, with green-filled circle marking the position of the vessel at the time the satellite image was taken. The location of BCO is indicated by a cyan triangle. (bottom) (e) Mean potential temperature, (f) specific humidity, (g) wind speed and (h) divergence in cold pool and environmental soundings. In each panel, the standard deviation around the mean for the two types of soundings is represented by a dashed line. In panel h, the black line indicates an example of divergence calculated over an entire EUREC⁴A circle (from 1841 to 1936 UTC) using all dropsondes (12 in total, including 4 in cold pools), and the orange line indicates the divergence over the same circle by considering environmental dropsondes only. For the sake of clarity, the x-axis is stretched by a factor of 4 for these two curves, that is one should read ± 20 instead of ± 80 ($10^{-6} s^{-1}$).

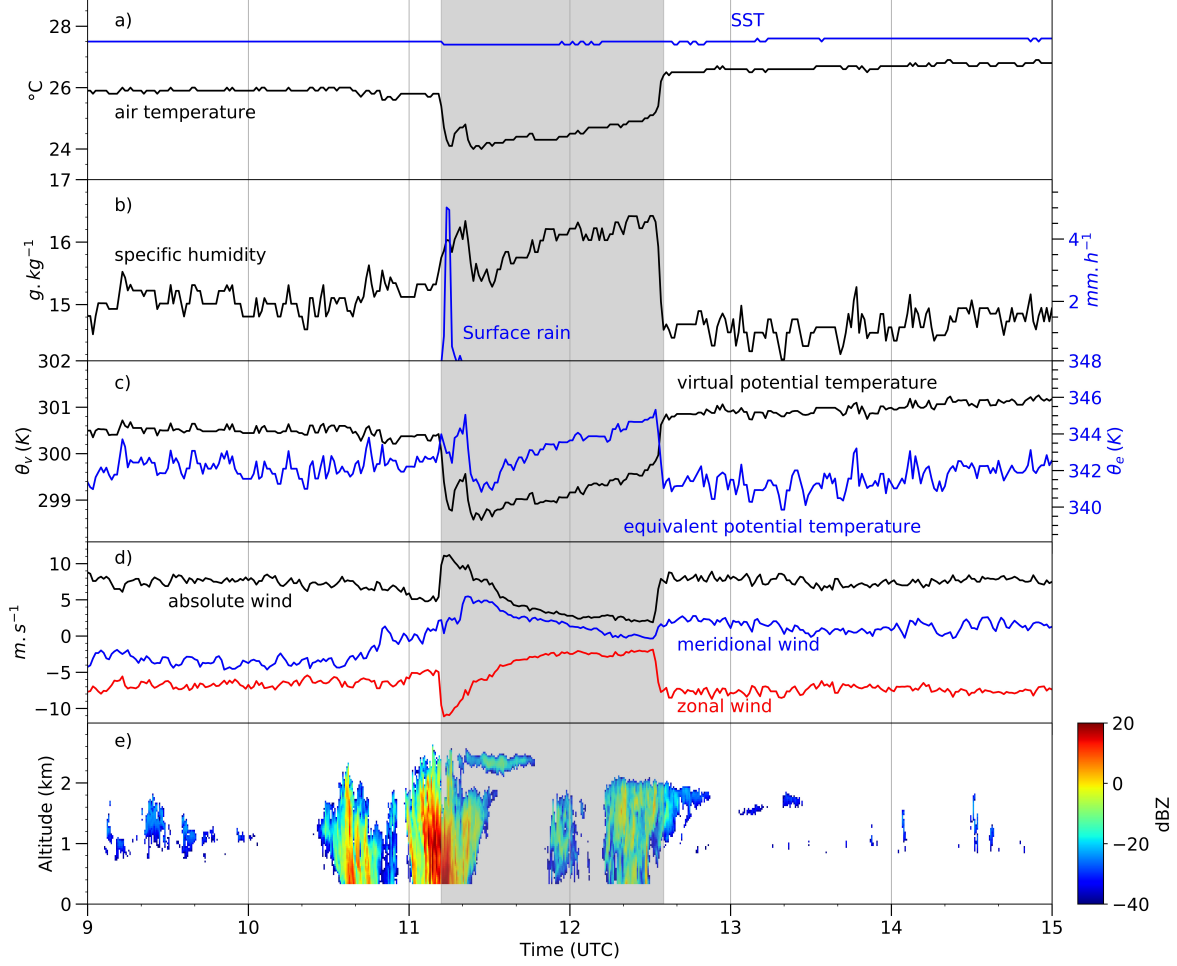


Figure B.6: Shipboard measurements from the Meteor on February 2, from 0900 to 1500 UTC: (a) surface air (black) and sea temperature (blue), (b) surface specific humidity (black) and surface rain rate (blue), measured using a ship rain gauge SRM 450. (c) virtual potential temperature (black) and equivalent potential temperature (blue), (d) Zonal (red), meridional (blue) and absolute (black) wind speed, (e) Vertically pointing shipboard W-band radar data at 94 GHz, measured using the Raman lidar system LICHT (Lidar for Cloud, Humidity and Temperature profiling) formerly at BCO and described in Stevens et al. (2016). Color bar indicates uncalibrated radar signal-to-noise ratio, best interpreted relative to itself. In the five panels, the grey shading marks the cold pool time period. Meteor data are freely accessible at <https://observations.ipsl.fr/aeris/eurec4a/>.

B.6.3 Statistics for other days

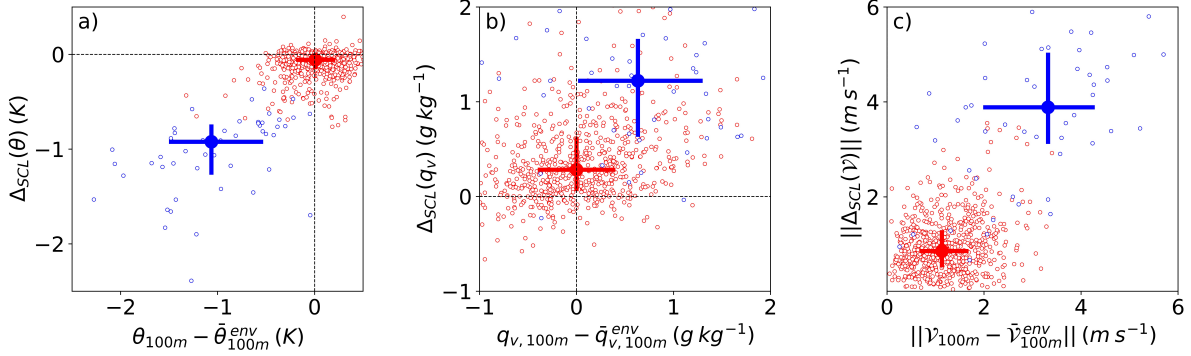


Figure B.7: Scatter plot of the vertical variations of (a) potential temperature, (b) specific humidity and (c) wind speed between 100 and 500 m vs spatial anomalies of the same variables at 100 m. The soundings considered are dropsondes launched in EUREC⁴A circles, and spatial anomalies are calculated with respect to the circle environmental mean, that is the mean calculated circle by circle with dropsondes launched out of cold pools only. Also reported are the median values for cold pools (blue) and their environment (red). The bars indicate the 25th and 75th percentiles of the distributions.

We will now look at a few statistics of cold pools for other days during EUREC⁴A, focusing first on the HALO dropsondes launched in the EUREC⁴A circles. In total, the aircraft dropped more than 700 dropsondes along 72 circles, that is about 10 dropsondes per circle. We use circle sounding data to compare the characteristics of cold pools to those of their close environment. In Fig. B.7, the x-axis represents the difference between the potential temperature (a), specific humidity (b) and wind speed (c) measured by each dropsonde at 100 m and the mean value at the same altitude for environmental dropsondes launched in the same EUREC⁴A circle. Mathematically, we define in the following $X_{z_0} = \frac{1}{100m} \int_{z_0-50m}^{z_0+50m} X(z) dz$ as the value of X at the altitude z_0 averaged over a 100 m depth and $\bar{X}_{z_0}^{env}$ as the mean circle value of X_{z_0} in environmental soundings. We average X_{z_0} over a 100 m interval in order to be less sensitive to local vertical heterogeneities. With these notations, the x-axes become respectively $\theta_{100m} - \bar{\theta}_{100m}^{env}$, $q_{v,100m} - \bar{q}_{v,100m}^{env}$ and $\|\mathbf{v}_{100m} - \bar{\mathbf{v}}_{100m}^{env}\|$. They show that cold pools are on average 1 K colder and 1 g kg⁻¹ moister than their environment, and that the wind difference between individual soundings and mean circle values is on average 2 m s⁻¹ larger for cold pools than for their environment.

The y-axis quantifies the imprint of cold pools on the vertical profiles for the same variables. For each sounding, the y-coordinate represents the difference $\Delta_{SCL}(X)$ in the SCL between 100 and 500 m, that is in and above cold pools when there is one. Mathematically, we define $\Delta_{SCL}(X) = X_{100m} - X_{500m}$ for both cold pool and environmental soundings. In the environment, on average, there is little difference in terms of potential temperature, specific humidity and wind speed between the two layers, consistent with what we would expect for a relatively well-mixed layer. On the other hand, cold pool soundings are on average 1 K cooler and 1 g kg⁻¹ moister at 100 m, and the average wind difference is about 3.5 m s⁻¹ between the two layers. These values are consistent with those of the x-axis and suggest that in a well-mixed layer the vertical imprint of cold pools can be used to estimate the differences between cold pools and their environment in terms of potential temperature, specific humidity and wind

speed.

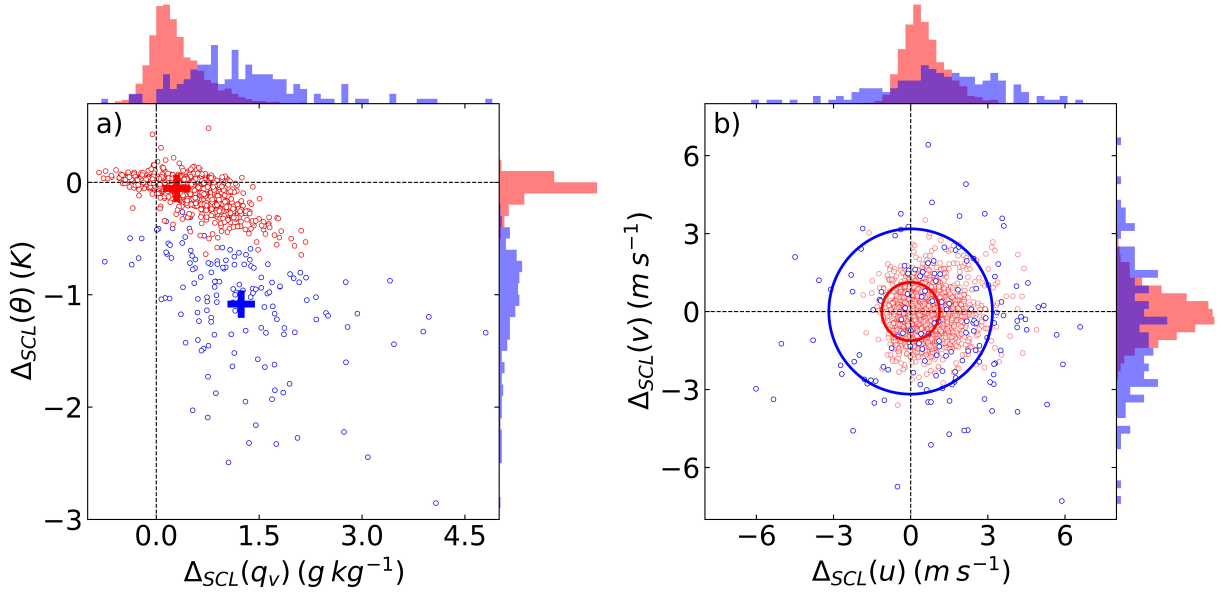


Figure B.8: (left) Scatter plot of the difference of potential temperature at 100 m and at 500 m $\Delta_{SCL}(\theta)$ vs difference of specific humidity between the same altitudes $\Delta_{SCL}(q_v)$ for cold pool (blue) and environmental (red) soundings. The crosses indicate the mean of each distribution. On the upper and right sides of the graph, marginal distributions of each variable are also reported for cold pools and their environment. The bins of the histograms are 0.1 g kg⁻¹ wide for the specific humidity and 0.1 K wide for potential temperature. To facilitate the comparison, the area under each distribution has been normalized to 1. (right) Same plot, but for the meridional and zonal winds, with a bin width of 0.3 m s⁻¹ for the corresponding histograms. The circles represent the mean of the absolute wind speed distributions, that is the mean of $\sqrt{(\Delta_{SCL}(u))^2 + (\Delta_{SCL}(v))^2}$ for cold pool and environmental soundings.

Based on these results, we take in Fig. B.8 the vertical imprint of cold pools as a proxy to estimate the intensity of cold pool anomalies. It allows us to generate statistics for all EUREC⁴A soundings, including the many radiosondes and dropsondes not launched in EUREC⁴A circles for which we do not have any environmental reference to assess cold pool properties. Consistently with Fig. B.7, cold pool soundings are on average 1 K colder at 100 m than at 500 m, 1 g kg⁻¹ moister and experience a wind difference between the two layers 2 m s⁻¹ larger than for environmental soundings. The coldest cold pools are also those for which the humidity difference is the most important, although potential temperature alone explains only a small part of the total humidity variance for cold pool soundings ($R^2 = 0.19$). The difference in humidity profiles in the SCL between cold pools and environment is also visible in Fig. B.9 a, which gives the equivalent potential temperature in cold pool and environmental soundings. This panel further shows that cold pool soundings are well mixed in equivalent potential temperature between 1 and 2.5 km, suggesting efficient convective mixing above cold pools.

In Fig. B.8 b, the zonal and meridional wind distributions show that cold pools spread in all directions, consistently with the conceptual picture of a density current propagating over a

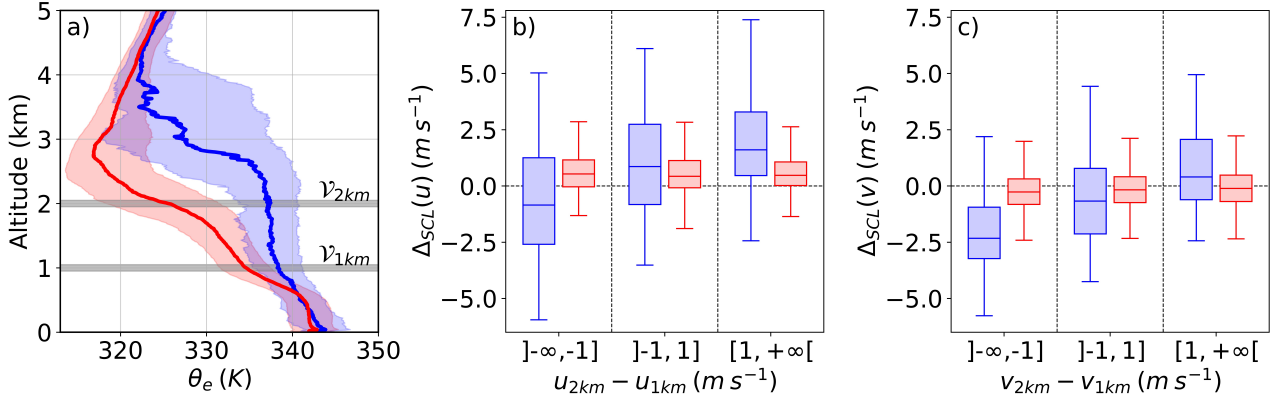


Figure B.9: (a) Profiles of equivalent potential temperature for all EUREC⁴A soundings launched in cold pools (blue) and in their environment (red). The bold lines represent the medians of each distribution and the color shading their interquartile range. The grey shading indicate the altitudes taken to define $\mathbf{V}_{1km} = (u_{1km}, v_{1km})$ and $\mathbf{V}_{2km} = (u_{2km}, v_{2km})$ used in the middle and right panels. (b) Box and whisker plots for the zonal wind difference between 100 and 500 m $\Delta_{SCL}(u)$ for three zonal wind shear ranges in the cloud layer. The wind shear ranges are defined by calculating for each sounding the difference between $u_{2km} = \int_{1.8km}^{2.2km} u(z)dz$ and $u_{1km} = \int_{0.8km}^{1.2km} u(z)dz$. Horizontal line within each box represent the median, box bottom and top are 1st (Q1) and last (Q3) quartile of the distributions, and the whiskers extend up to 1.5 interquartile range above Q3 and below Q1. (c) Same plot, but for the meridional wind difference between 100 and 500 m $\Delta_{SCL}(v)$ for three meridional wind shear ranges in the cloud layer.

solid boundary. In the presence of vertical wind shear, theories and numerical simulations nevertheless predict that cold pool spreading may not be perfectly isotropic, as vertical momentum transport might favor the propagation of cold pools downshear (Grant et al., 2020; Mahoney et al., 2009; Moncrieff, 1992). Fig. B.9 middle and right panels test this hypothesis by showing the zonal and meridional wind difference between 100 and 500 m ($\Delta_{SCL}(u)$ and $\Delta_{SCL}(v)$) for three zonal and meridional wind shear ranges in the cloud layer, between 1 and 2 km: $]-\infty, -1]$, $]-1, 1]$ and $]1, +\infty[$ (m s^{-1}). The box and whisker diagrams reveal that $\Delta_{SCL}(u)$ is gradually shifted to higher values as the zonal wind shear in the cloud layer increases. On the contrary, in environmental soundings, the zonal wind shear in the cloud layer has little, if any, influence on $\Delta_{SCL}(u)$. The same result stands for the meridional wind, suggesting significant momentum transport by downdrafts in cold pools in all directions. $\Delta_{SCL}(u)$ is higher than $\Delta_{SCL}(v)$ for both cold pool and environmental soundings because of strong zonal winds in the region, around -8.1 m s^{-1} at 100 m on average, compared to -1.7 m s^{-1} at 100 m for mean meridional winds. These strong winds create friction in the lowest layers and decrease u_{100m} with respect to u_{500m} , thus increasing $\Delta_{SCL}(u)$. The relation between the direction of propagation of cold pools and the vertical shear, as well as the vertical imprint of cold pools on dynamic and thermodynamic profiles illustrate some interesting results inaccessible with detection methods based on time series (de Szoeke et al., 2017; Young et al., 1995; Zuidema et al., 2012). Though preliminary, these results provide a first systematic and robust characterization of cold pools over the wide range of conditions observed during EUREC⁴A, demonstrating how valuable detecting cold pools from soundings can be for the study of cold pools and cloud-circulation coupling.

B.7 Conclusions

In this study, we presented a new method using the mixed layer height as a criterion for detecting cold pools from soundings over tropical oceans. This method is based on the analysis of more than 2000 radiosondes and dropsondes from the EUREC⁴A field campaign that took place over the western tropical Atlantic near Barbados in January-February 2020. The mixed layer height H_{mix} , defined using the virtual potential temperature, shows a left heavy tail for EUREC⁴A soundings that we attribute to cold pools. It leads us to classify the soundings with $H_{mix} < 400$ m as cold pool soundings, and those with $H_{mix} \geq 500$ m as environmental soundings.

We first test this criterion in a simulation over the Barbados region, performed with the LES version of the ICON model. In this simulation, cold pools are visually identified as regions with negative surface temperature anomalies and positive surface precipitation. They coincide with regions of shallow mixed layer, supporting the relevance of the mixed layer height as a proxy to identify cold pools in this LES simulation despite the presence of a strong SST gradient across the domain. Then, we apply our detection method to EUREC⁴A soundings, looking first at January 24 and February 2, two EUREC⁴A flight days during which the number of cold pools detected in dropsondes is particularly high. On these two days, the cold pools detected with our method are consistent with satellite images, surface time series from the research vessel Meteor, and thermodynamic and dynamic profiles measured by soundings. On February 2, the calculation of the divergence in and above one of the cold pools further reveals intense mesoscale circulations, consistent with the observed persistence of cloud clusters above cold pools that day.

Finally, we give a few statistics for the entire EUREC⁴A period, focusing initially on the dropsondes launched by HALO in the EUREC⁴A circle. By comparing, circle by circle, the cold pool and environmental characteristics, we find that at 100 m, the cold pool soundings are on average 1 K colder and 1 g kg^{-1} moister than the environmental soundings at the same altitude, and experience a wind deviation of more than 3 m.s^{-1} with respect to their mean environment. We further show that these differences have a similar amplitude as the vertical differences between 100 and 500 m for each cold pool sounding, while environmental soundings exhibit a small difference between these two altitudes, consistent with the well-mixed nature of the subcloud layer (SCL) in trade-wind regimes. This significant vertical imprint of cold pools in the SCL is then studied for all the EUREC⁴A soundings. The coldest cold pools turn out to be those with the largest humidity difference between 100 and 500 m. When there is shear in the cloud layer, we find that the spreading of cold pools is favored downshear, suggesting momentum transport by unsaturated downdrafts feeding cold pools.

These first results and the method described here pave the way for more comprehensive analyses of cold pools over tropical oceans. The quantity and quality of the measurements made during EUREC⁴A is a great opportunity to perform such analyses. In particular, drones and low-flying aircraft are likely to provide valuable information on the vertical structure of cold pools, which could complement sounding measurements. In parallel, time series from research vessels and surface weather stations could help study the characteristics of cold pools close to the ground and better understand their interactions with the surface. The interest in cold pools in tropical oceans is not new (Houze Jr & Betts, 1981; Leary & Houze Jr, 1979; E. J. Zipser, 1969), but EUREC⁴A offers an unprecedented dataset to finely characterize the temporal and spatial properties of cold pools and explore unanswered questions about their effect on the

organization and propagation of convection. Hopefully, the method and results described in this paper will help us make the best use of this dataset – and turn this opportunity into reality.

B.8 Data availability

ICON primary data can be accessed through the "Mistral" super computer of the German Climate Computing Center / Deutsche Klimarechenzentrum (DKRZ). BCO micro-rain radar data are accessible to the broader community, as detailed in Stevens et al. (2016). Access to primary data is provided here: <https://mpimet.mpg.de/en/science/the-atmosphere-in-the-earth-system/working-groups/tropical-cloud-observation/barbadosstation1/instrumentation-and-data>. Radiosonde and dropsonde data are described respectively in Stephan et al. (2020) and George et al. (2021) and available through the AERIS portal (<https://eurec4a.aeris-data.fr/>). Free access to Meteor surface meteorological data and Raman lidar data is also provided by AERIS. The MODIS-Terra picture on January 24 is taken from NASA Worldview (<https://worldview.earthdata.nasa.gov>).

Appendix C

Atmospheric radiative profiles during EUREC⁴A

Contents

| | | |
|------------|--|------------|
| C.1 | Preamble | 195 |
| C.2 | Abstract | 196 |
| C.3 | Introduction | 196 |
| C.4 | Data and methods | 197 |
| C.4.1 | Radiosonde and dropsonde data | 197 |
| C.4.2 | Radiative transfer calculation | 198 |
| C.5 | Preliminary results and discussion | 201 |
| C.5.1 | Variability across soundings | 202 |
| C.5.2 | Diurnal cycle and day-to-day variability | 202 |
| C.5.3 | Radiative signatures of mesoscale patterns of cloud organization | 205 |
| C.5.4 | Effect of sharp moisture gradients on radiative heating profiles | 207 |
| C.6 | Uncertainty assessment | 209 |
| C.7 | Conclusions | 211 |
| C.8 | Code and data availability | 211 |

C.1 Preamble

After a field campaign, it is common to publish data papers in peer reviewed journals in order to document the datasets that have been collected. This is the objective of the paper presented in this chapter, which describes a dataset freely available online composed of the radiative profiles computed for all the soundings measured during EUREC⁴A (more than 2000 soundings). Preliminary analyses about the variability of the calculated radiative profiles are also presented. In particular, we provide a first characterization of the radiative properties of sugar, gravel, fish, and flowers, which we mentioned in Chapter 2 as examples of objects that could focus and guide atmospheric research.

The paper was published in *Earth System Science Data* in February 2021 and can be cited as Touzé-Peiffer, L., Albright, A. L., Fildier, B., Pincus, R., Vial, J., & Muller, C. (2021). Atmospheric radiative profiles during EUREC⁴A. *Earth System Science Data*, 13(2), 617-630. Anna Lea Albright, Benjamin Fildier and I contributed equally to the analysis, figures and text. Robert Pincus, Caroline Muller and Jessica Vial helped in conceptualizing and guiding this project and contributed to the manuscript.

C.2 Abstract

The couplings among clouds, convection, and circulation in trade-wind regimes remain a fundamental puzzle that limits our ability to constrain future climate change. Radiative heating plays an important role in these couplings. Here we calculate clear-sky radiative profiles from 2504 in situ soundings (1068 dropsondes and 1436 radiosondes) collected during the EUREC⁴A field campaign, which took place in the downstream trades of the western tropical Atlantic in January-February 2020. We describe the method used to calculate these cloud-free, aerosol-free radiative profiles. We then present preliminary results sampling variability at multiple scales, from the variability across all soundings to groupings by diurnal cycle and mesoscale organization, as well as individual soundings associated with elevated moisture layers. We also perform an uncertainty assessment and find that the errors resulting from uncertainties in observed sounding profiles, and ERA5 reanalysis employed as upper and lower boundary conditions are small. The present radiative profile data set can provide important additional detail missing from calculations based on passive remote sensing and aid in understanding the interplay of radiative heating with dynamic and thermodynamic variability in the trades. The data set can also be used to investigate the role of low-level radiative cooling gradients in generating shallow circulations. All data are archived and freely available for public access on AERIS (Albright et al., 2020, <https://doi.org/10.25326/78>).

C.3 Introduction

The EUREC⁴A field campaign, which took place in January and February 2020 in the downstream trades of the western tropical Atlantic, was designed to elucidate the couplings among clouds, convection, and circulation in trade-wind regimes and understand the role of this interplay in climate change (Bony et al., 2017). Shallow trade-wind clouds cover large parts of tropical oceans (Wood, 2012), yet their response to warming remains largely unknown, and uncertainty in shallow convective processes are the cause for large uncertainties in climate projections (Bony & Dufresne, 2005; S. C. Sherwood, Bony, & Dufresne, 2014; Vial, Dufresne, & Bony, 2013; Zelinka et al., 2020). Among all physical processes involved in shallow convection, atmospheric radiative cooling emerges as key to the coupling between low-level circulations and convection. Understanding the dynamics driven by variations in radiative heating rates, and potential relationship to the mesoscale organization of clear and cloudy regions, was one motivation for the campaign (Bony et al., 2017).

A characteristic feature of the trade-wind vertical moisture profile is a sharp humidity gradient between the moist marine boundary layer and dry, subsiding free troposphere around two kilometers Malkus (1958); Riehl, Yeh, Malkus, and La Seur (1951). This characteristic vertical moisture structure has important implications for radiative cooling profiles, but it is difficult to observe with satellite remote sensing (Stevens et al., 2017). Indeed, moisture profile features,

such as the sharp decreases in moisture at the top of the marine boundary layer or elevated moisture layers, are smaller than typical weighting functions of even hyperspectral instruments (e.g. Maddy & Barnet, 2008; Menzel, Schmit, Zhang, & Li, 2018; Schmit, Li, Ackerman, & Gurka, 2009), especially in the lowest three kilometers, corresponding to the weakest absorption lines (Chazette, Marnas, Totems, & Shang, 2014). The lack of informative observations means that the vertical profile of water vapor in large-scale atmospheric analyses do not represent the fine-scale moisture structure indicated by soundings (Pincus et al., 2017). Errors in the vertical moisture structure estimated from passive remote sensing produce corresponding errors in radiative cooling profiles computed from retrievals and/or analyses, making in situ soundings especially valuable.

Here we calculate radiative profiles from 2504 soundings (1068 from dropsondes and 1436 from radiosondes) collected during EUREC⁴A, whose network of observations provided extensive sampling of the tropical trade-wind environment. Similar studies have been conducted over continents as part of the Atmospheric Radiation Measurement program (Kato et al., 1997; E. Mlawer, Clough, & Kato, 1998), over the western Pacific warm-pool region as part of the Coupled Ocean Atmosphere Response Experiment (Guichard, Parsons, & Miller, 2000), and over the western tropical Atlantic, albeit focused on transported Saharan dust layers (Gutleben, Groß, & Wirth, 2019). The present radiative profiles have the potential to complement and further what can be learned from calculations based on passive remote sensing. In addition, this data set may help in understanding how low-level gradients in radiative cooling fuel shallow circulations, as observed to emerge in remote sensing and large-eddy simulations (L’Ecuyer, Wood, Haladay, Stephens, & Stackhouse Jr., 2008; Seifert, Heus, Pincus, & Stevens, 2015; Stephens et al., 2012). These shallow circulations are speculated to influence the mesoscale spatial organization of shallow convection, a question at the core of EUREC⁴A (Bony, Schulz, Vial, & Stevens, 2020b; Stevens, Bony, et al., 2020).

In Section C.4, we describe the data, the radiative transfer code, and the procedure underlying the calculation of the radiative profiles. We then present initial results to open the discussion on questions that could be explored with these radiative profiles (Section C.5). Lastly, we perform an uncertainty assessment (Section C.6) and find that errors resulting from uncertainties in the sea surface skin temperature, in situ soundings, and ERA5 reanalysis used as boundary conditions are modest.

C.4 Data and methods

C.4.1 Radiosonde and dropsonde data

From January 8 to February 19, over 2500 atmospheric soundings were conducted using dropsondes and radiosondes over the western tropical Atlantic ocean south and east of Barbados. As the sondes fall or ascend, their simple autonomous sensors, equipped with a GPS receiver, measure the vertical profiles of pressure, temperature, relative humidity, and instantaneous horizontal wind. To calculate radiative profiles, we employ level-3 data, which have been interpolated into a common altitude grid with 10 meter spacing (George et al., 2021; Stephan et al., 2020). We select dropsondes and radiosondes that have measurements on more than ten atmospheric levels in total. This filter suffices to remove failed soundings and results in an input data set consisting of 1068 atmospheric profiles from dropsondes and 1436 profiles from radiosondes. The minimum and maximum levels z_{min} and z_{max} measured by each sonde are

reported in the final data set.

Figure C.1a shows the geographic and temporal distributions of the sondes used to calculate the radiative profiles. Radiosondes were launched from a network of one land station and four research vessels, within a region ranging from 51–60°W to 6–16°N. On land, radiosondes were launched from the Barbados Cloud Observatory (BCO), located on a promontory on the easternmost point of Barbados called Deebles Point (13.16°N, 59.43°W), where it is exposed to relatively undisturbed easterly trade-winds. The fleet of four research vessels includes the French research vessel *L’Atalante*, two German research vessels *Maria S. Merian* (MS-Merian) and *Meteor*, and the American research vessel from the National Oceanic and Atmospheric Administration (NOAA) *Ronald H. Brown* (RH-Brown). Dropsondes were launched from both the German High Altitude and Long Range Research Aircraft (HALO) and the United States Lockheed WP-3D Orion from NOAA (WP-3D). HALO typically flew at an altitude of 30,000 ft (approximately 9 km), following a circular flight pattern with 90 km radius centered at 13.3°N, 57.7°W. When launching sondes, the WP-3D flew at 24,000 ft (approximately 7 km), releasing sondes along both linear and circular patterns in the region covered by HALO, as well as further to the east close to the nominal position of the RH-Brown.

Radiosondes were launched every four hours, daily from January 8–February 19, 2020, approximately synchronously from each platform. Given that the time-lag between ascending and descending radiosondes is on the order of hours, and that there is substantial horizontal drift between the ascent and descent, we chose to compute separate radiative profiles for ascending and descending radiosondes. For dropsondes, HALO flight takeoffs were staggered at 5, 8, and 11 am local time, with flights lasting approximately eight hours, yielding roughly 72 sondes per flight. The WP-3D undertook three night flights, which allows for a better characterization of the diurnal cycle, together with the radiosondes launched during the night (Figure C.1b).

We refer the reader to Stephan et al. (2020) and George et al. (2021) for a complete description of the radiosonde and dropsonde data sets, respectively, and Bony et al. (2017) and Stevens et al. (2021) for an overview of the campaign scientific motivations and measurement strategy.

C.4.2 Radiative transfer calculation

The radiative transfer code used here, RRTMGP (Rapid Radiative Transfer Model for GCMs, Parallel) (Pincus et al., 2019), is a plane-parallel correlated- k two-stream model based on state-of-the-art spectroscopic data for gas and condensate optics. It is based on line parameters from Atmospheric and Environmental Research and the MT_CKD water vapor continuum absorption model (E. J. Mlawer et al., 2012). The calculation of radiative profiles from radiosonde and dropsonde data then proceeds in the following way:

1. vertical soundings of temperature, pressure, and water vapor specific humidity at 10 meter resolution are interpolated onto a 1 hPa vertical grid and then merged with temperature and specific humidity from ERA5 reanalyses in the following manner. Sonde measurements below 40 m are first truncated for all sondes: radiosondes do not provide data in this surface layer because of deck heating effects on ships (Stephan et al., 2020), and we apply the same filter to dropsondes for consistency. The ERA5 profiles at hourly and 0.25° resolution (European Centre for Medium-Range Weather Forecasts, 2017) are linearly interpolated temporally and spatially to the time, latitude, and longitude of the

sounding. ERA5 values are used above the highest level measured by each sonde to extend the observed soundings vertically to 0.1hPa and account for the effect of high-altitude thermodynamic variability on the radiative cooling profiles below. To obtain the lower boundary condition, we linearly interpolate the ERA5 sea surface skin temperature (SST_{skin}), also at hourly and 0.25° resolution (European Centre for Medium-Range Weather Forecasts, 2017), to the time, longitude and latitude where the sounding was launched;

2. CO_2 concentrations are set to the present day value of 414 ppm while CH_4 , O_3 and N_2O concentrations are taken from the standard tropical atmosphere profile of Garand et al. (2001);
3. the set of resulting profiles is then used as input to RRTMGP to derive upwelling and downwelling clear-sky radiative fluxes in the shortwave and longwave ranges of the spectrum. The calculation uses a spectrally-uniform surface albedo of 0.07 and a spectrally-uniform surface emissivity of 0.98, typical values for tropical oceans.

Dropsondes and radiosondes drift horizontally as they rise and/or fall (Figure C.1a), which could lead to slight errors due to aliasing of horizontal variability in moisture content into vertical variability. This potential source error is less pronounced for dropsondes than for radiosondes due to their faster speed of travel through the troposphere.

We compute radiative fluxes and heating rates only for the gaseous component of the atmosphere, without explicitly taking into account cloud or aerosol properties. These radiative profiles are therefore clear-sky and aerosol-free. The soundings do, however, capture the water vapor structure, including regions of high humidity in cloud areas and aerosol layers. Cloud cover in trade-wind regimes is relatively low, between 10% (Nuijens, Medeiros, Sandu, & Ahlgrimm, 2015) and 20% (Medeiros & Nuijens, 2016) for active clouds, so cloud-free, or clear-sky, profiles are representative of the thermodynamic environment. Taking into account the influence of cloud liquid water would require a number of ad hoc assumptions about microphysical and optical properties within clouds (see for instance Guichard et al., 2000). Similarly, we do not directly represent the radiative effect of mineral dust aerosols. The dominant aerosol radiative effect in this region has been shown to result from the covariance of aerosols with water vapor, such that aerosols tend to be associated with elevated moisture layers Gutleben, Groß, Wirth, and Mayer (2020); Gutleben et al. (2019). Dust aerosol layers are, moreover, more common in the summer than in winter (Lonitz et al., 2015). We leave open the possibility that direct scattering by dust aerosols has an additional role on radiative heating rates, but do not have the coincident data to appropriately address this question for all soundings.

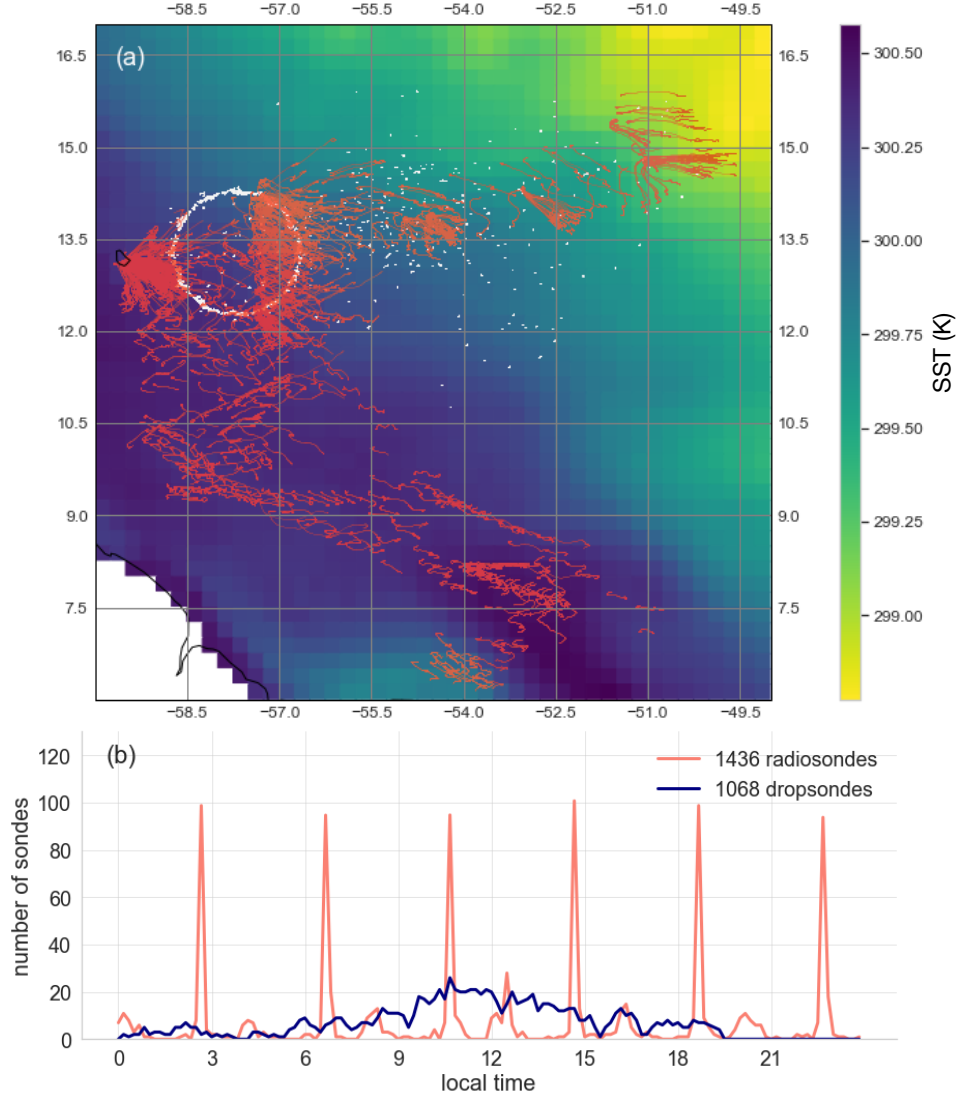


Figure C.1: (a) The EUREC⁴A sounding network: 1068 soundings from dropsondes (white) and 1436 from radiosondes (coral). We employ 807 dropsondes launched from HALO and 261 dropsondes from the WP-3D to calculate radiative profiles, as well as 276, 342, 147, 362, and 309 radiosondes launched from the Atalante, BCO, MS-Merian, Meteor, and RH-Brown, respectively. Background colors show sea surface skin temperature (SST_{skin}) from ERA5 re-analysis at 0.25° resolution averaged over January and February. (b) The diurnal distribution of the 1068 dropsondes (blue) and 1436 radiosondes (coral) with sonde launch-time binned in 10-minute intervals.

C.5 Preliminary results and discussion

This section includes a first exploration of the data set. We examine radiative variability at different scales – across all soundings, at the diurnal timescale, and according to different patterns of mesoscale organization – as well as in individual profiles showing the influence of sharp vertical moisture gradients on radiative heating rates.

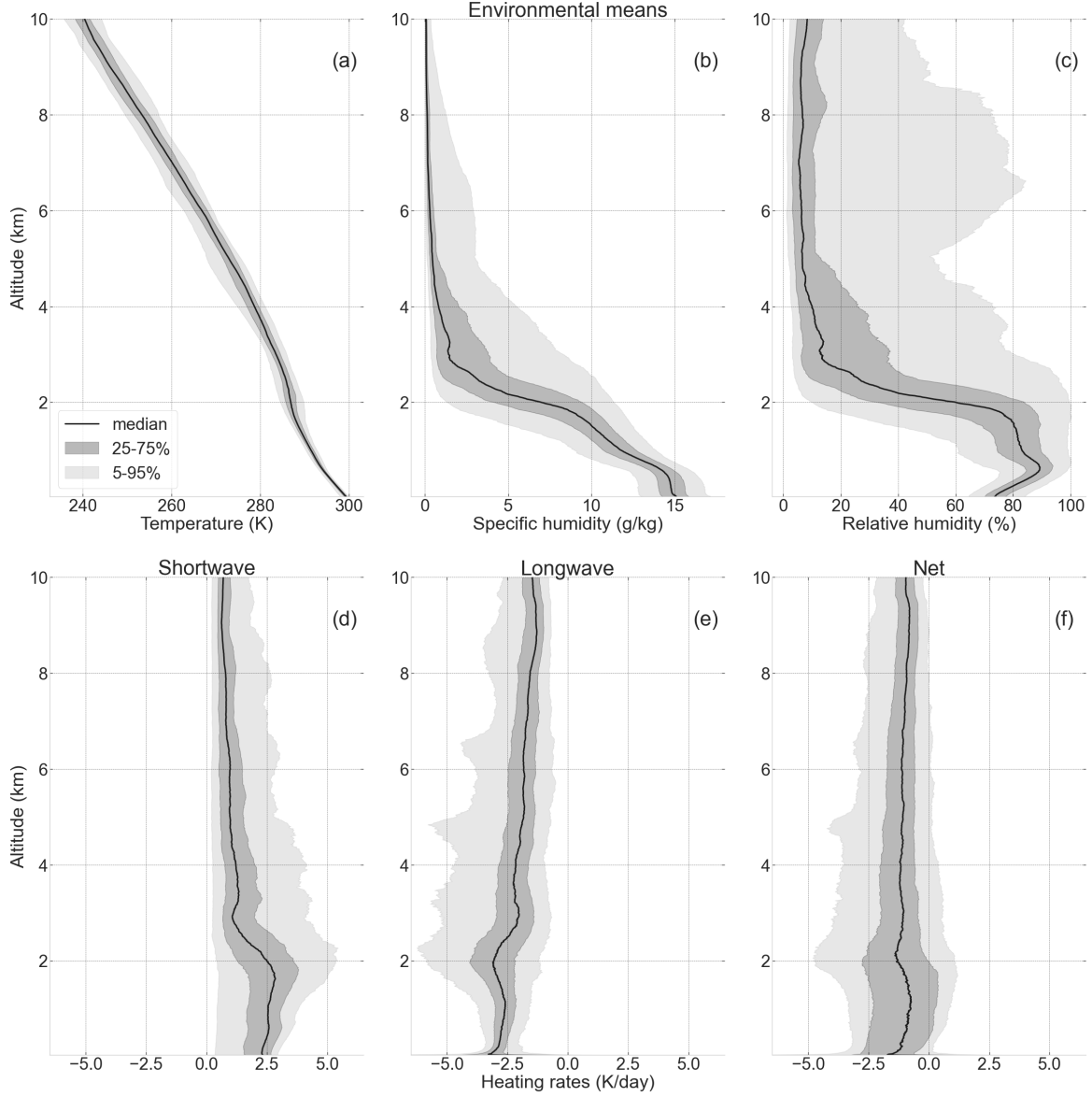


Figure C.2: Top: Temperature (a), specific humidity (b) and relative humidity (c) (with respect to ice for $T < 0^\circ\text{C}$) from EUREC⁴A dropsonde and radiosonde data. Bottom: Shortwave (d), longwave (e) and net (f) heating rates calculated from EUREC⁴A dropsonde and radiosonde data using the radiative transfer code RRTMGP. The center traces are the median profiles, and the medium and light grey shadings indicate the 25–75% and 5–95% intervals, respectively. For the shortwave, the median and the interquartile range are calculated using daytime values only.

C.5.1 Variability across soundings

A distribution of longwave, shortwave, and net heating rates, as well as large-scale thermodynamic quantities, are shown in Fig. C.2. Local extrema in the median shortwave, longwave, and net heating rates occur near 2 km (Fig. C.2d,e,f), associated with the rapid decrease in specific and relative humidity at this level (Fig. C.2b,c). The top of the planetary boundary layer, or interface between the moist marine boundary layer and dry free troposphere above, is expected to occur around 2km in the trades (Cao, Giambelluca, Stevens, & Schroeder, 2007; Malkus, 1958; Stevens et al., 2017). The spread in specific and relative humidity is greater than that in temperature, suggesting a strong role for moisture variability on the variability in radiative heating rates. On average, longwave cooling is stronger than shortwave heating, such that net heating rates are largely negative from the surface up to 10 km, with a median value around -1 K/day. Additional local minima in longwave heating are observed around 3 and 5 km between the 5% and 25% quantiles. These local minima could, for instance, correspond to the radiative effect of elevated moisture layers arising from convection detraining moisture at these higher levels, albeit less frequently, or aerosol layers associated with increased water vapor concentrations (Gutleben et al., 2019; O, Wood, & Tseng, 2018; Stevens et al., 2017; Wood, O, et al., 2018; Wood, Tseng, et al., 2018).

We next partition radiative heating variability into its variability in time (e.g. diurnal cycle, day-to-day variability) and regarding the spatial characteristics of the convection field (e.g. the spatial distribution of clear and cloudy regions).

C.5.2 Diurnal cycle and day-to-day variability

Figure C.3 gives an overview of the diurnal variability of radiative heating, which has been implicated in the diurnal cycle of convection and cloudiness (e.g., W. M. Gray & Jacobson Jr, 1977; D. A. Randall & Tjemkes, 1991; Ruppert & Johnson, 2016). Shortwave radiative heating follows the solar cycle. Longwave heating rates show less diurnal variability and have approximately the same amplitude (with an opposite sign) as shortwave heating rates during daytime. This compensation between longwave cooling and shortwave heating results in a daytime net heating rate that is slightly positive in the lower 2km. The daytime heating contributes to stabilizing the lower atmosphere, disfavoring convection. At night, strong radiative cooling destabilizes the lower troposphere and strengthens convection. The maximum nighttime longwave cooling occurs slightly above 2 km, with secondary cooling peaks occurring around 4 and 6km. During daytime, the peak in stabilizing radiative heating appears slightly below 2km. This difference in the height of peak radiative heating, albeit of different sign, could reflect differences in the height of the moist, convecting layer over the diurnal cycle: a shallower marine boundary layer during the day that deepens at night (Vial et al., 2019b). These considerations highlight the potential for subtle interactions among radiation, convection, and cloudiness on the diurnal timescale.

Fig. C.4 shows the day-to-day evolution of the shortwave (top), longwave (middle) and net (bottom) heating rates derived from radiosondes launched at BCO. In the shortwave and net heating rates, the daily stripes are due to zero shortwave heating during the night. In the longwave component alone, the amplitude of the diurnal cycle is less evident. Regarding the day-to-day variability, both in the shortwave and the longwave components, trends in the height-evolution of the radiative heating maxima appear to persist over several days. These trends are likely due to variations in humidity (e.g. Dopplack, 1972; Jeevanjee & Fueglistaler,

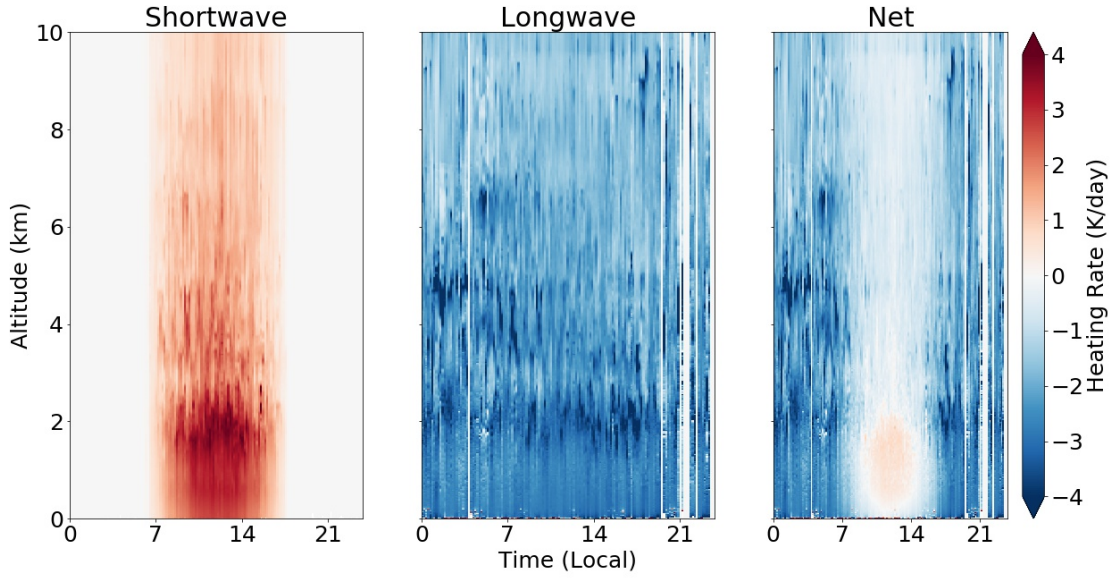


Figure C.3: Diurnal composite of shortwave (left), longwave (middle), and net (right) clear-sky heating rates binned in 10-minute intervals. Colored shadings indicate heating rates in units of K/day. The data are plotted with respect to local solar time to simplify interpretation of the diurnal cycle. White indicates the absence of data. We note that some variability, such as in the nighttime longwave radiative cooling variability, could result from different numbers of sondes launched throughout the diurnal cycle (as illustrated in Fig. C.1b).

2020) and are consistent with the presence of multi-day trends in moisture observed at BCO during the campaign (see Figure 13 in Stevens et al., 2021). At the end of the campaign, the rise in the peak of longwave cooling appears to correspond to the rising location of the interface between the moist, convecting layer below and dry free troposphere above (not shown). The persistence and evolution of radiative heating patterns could be tied to larger-scale synoptic moisture activity or to the evolution of mesoscale organization patterns.

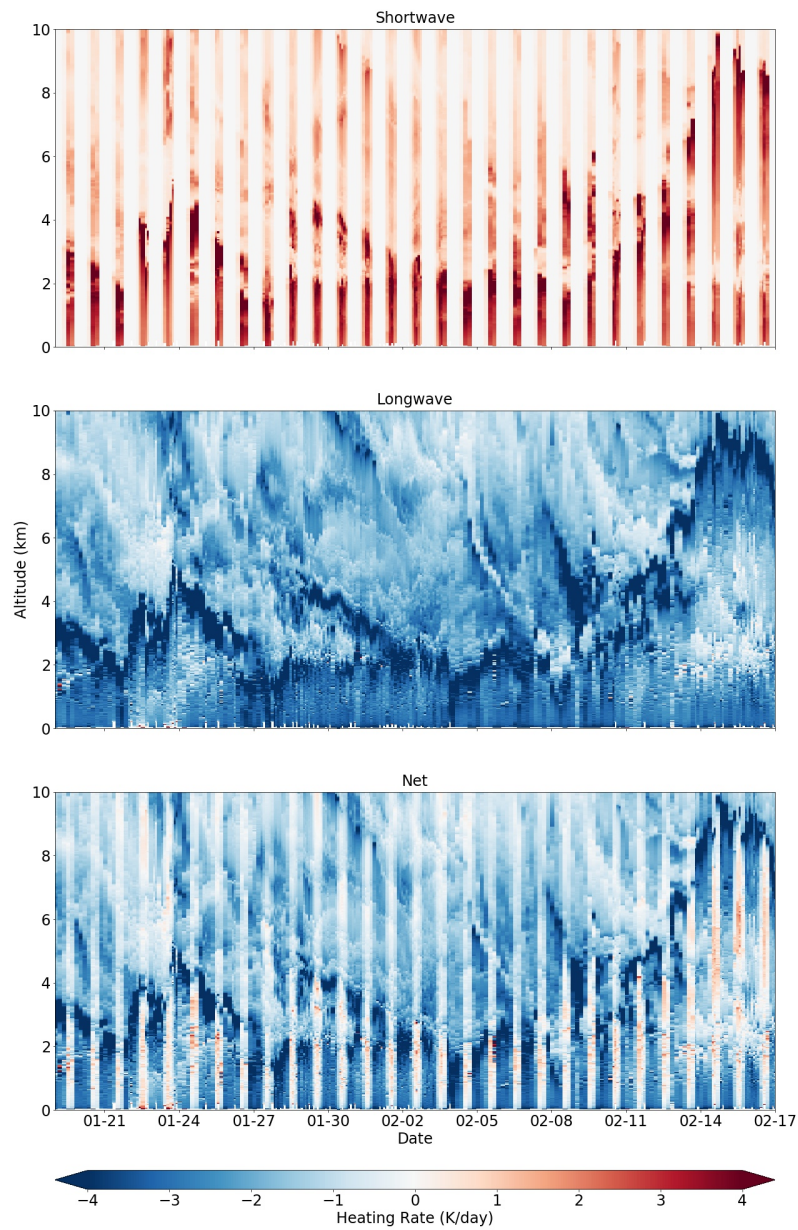


Figure C.4: Shortwave (top), longwave (middle), and net (bottom) heating rates at BCO during EUREC⁴A, from January 19 to February 17. The heating rates are calculated from radiosondes launched at BCO. In colors are heating rates with units of K/day. White indicates the absence of data.

C.5.3 Radiative signatures of mesoscale patterns of cloud organization

We next aggregate radiative heating rates spatially. Fig. C.5 illustrates four representative cases of the Fish-Gravel-Flower-Sugar classification established previously for mesoscale (20-2,000km) organization patterns of clear and cloudy regions (Bony et al., 2020b; Stevens, Bony, et al., 2020). These cloud organization patterns were identified visually from satellite imagery and correspond to differences in large-scale environmental conditions (Bony et al., 2020b). They are also observed to have different top-of-the-atmosphere radiative effects (Bony et al., 2020b). As outlined in Stevens, Bony, et al. (2020), Sugar refers to a ‘dusting’ of small, shallow clouds with low reflectivity and a random spatial distribution. Gravel clouds tend to be deeper than Sugar (up to 3-4km), have little stratiform cloudiness, precipitate, and organize along apparent gust fronts or cold pools at the 20-200km scales. Fish are skeletal networks (often fishbone-like) of clouds at the 200-2,000km scale with stratiform cloud layers; the Fish pattern is often associated with extratropical intrusions. Flowers are circular features defined by their stratiform cloud elements. Both Fish and Flowers are surrounded by large swaths of clear air.

We choose four days as an example of the large-scale environmental and radiative signature of each pattern, given the spatial pattern observable in the GOES-16 satellite images in the HALO flight path shown by the white circle. We plot daily-mean profiles for temperature, specific humidity, and relative humidity (Fig. C.5a,b,c), as well as shortwave, longwave, and net radiative heating rates (Fig. C.5d,e,f). These profiles were calculated from approximately 70 HALO dropsondes launched during the eight-hour flight on each day. We also plot the standard deviation of radiative heating for each flight (Fig. C.5g,h,i). As a first approximation, the standard deviation of daily radiative heating profiles acts as a proxy for spatial variability in radiative heating rates.

Spatial variability in radiative heating has been shown to drive shallow circulations (e.g. Naumann, Stevens, & Hohenegger, 2019) and affect convective organization (e.g. C. S. Bretherton et al., 2005; Muller & Held, 2012). In this illustrative example, the differences in the mean and standard deviation of the radiative heating rates hint at a role for differences in radiative cooling rates in the onset or maintenance of mesoscale patterns of organization. For instance, the ‘Fish’ pattern on January 22, 2020 is associated with a moister lower troposphere between 1 and 3km and slightly drier free troposphere above 4km. This vertical moisture distribution may give rise to the observed vertical variability in radiative heating rates, with larger peaks in the mean profile (Fig. C.5e) and standard deviation (Fig. C.5h) in radiative heating observable between 2 and 4km, likely corresponding to strong humidity gradients at these levels.

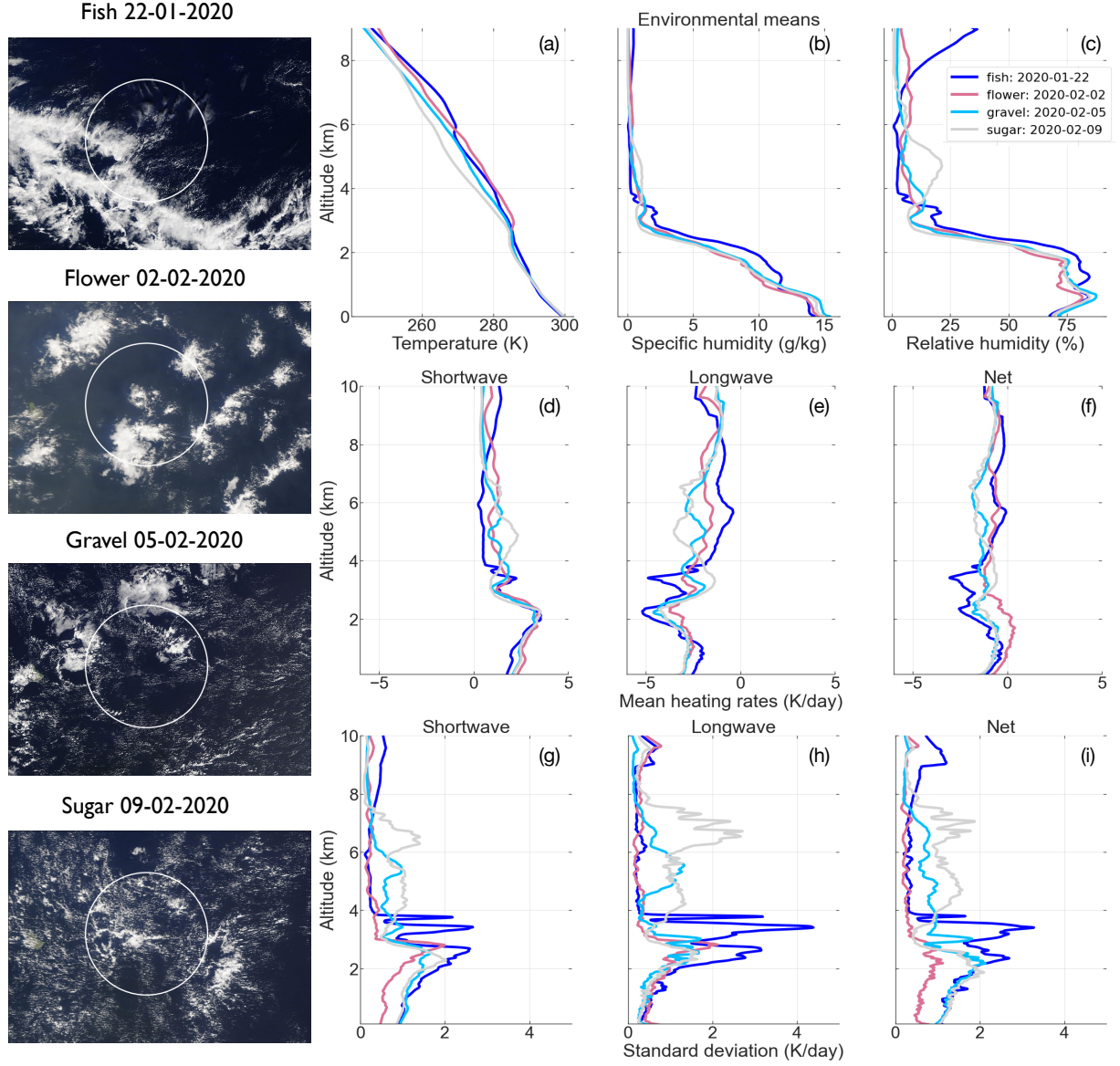


Figure C.5: Thermodynamic (a-c), daily mean radiative heating (d-f), and daily standard deviation of radiative heating (g-i) profiles classified by mesoscale organization pattern, using a characteristic example of each type as diagnosed from snapshots from GOES-16 infrared channel (left column). This figure employs HALO dropsondes launched in the circular flight pattern (shown by the white circle) on the chosen day, corresponding to roughly 70 dropsondes each. We focus on the spatial extent of the HALO flight pattern because the cloud organization pattern does not necessarily extend across the entire sampling domain Figure C.1a, nor have the patterns been shown to be scale-invariant.

C.5.4 Effect of sharp moisture gradients on radiative heating profiles

Figure C.6 highlights the radiative signatures of elevated moisture layers, which can persist for multiple hours at inversion levels (Gutleben et al., 2019; Stevens et al., 2017; Wood, O, et al., 2018). We focus in detail on two thermodynamic and radiative heating profiles of a particular elevated moisture layer extending to 4 kilometers, alongside GOES-16 images (Fig. C.6i,j) corresponding to these soundings. This structure persisted for at least four hours on January 24, 2020, and we plot thermodynamic conditions and radiative heating profiles sampled three hours apart, at 12:55 and 15:55 UTC (see Fig. C.6). A striking feature is the sharp peak in longwave cooling at the top of the moisture layer of nearly -20 K/day at 15:55 UTC, corresponding to the strong humidity gradient, with relative humidity decreasing by nearly 70% in 100 meters (Fig. C.6c,d).

Although we calculate clear-sky profiles only, the present work could be extended to account for radiative effect of cloud liquid water, which could be used, for instance, to investigate the radiative effect of geometrically- and optically-thin ‘veil clouds’ persisting at inversion levels (O et al., 2018; Wood, O, et al., 2018; Wood, Tseng, et al., 2018), such as those illustrated by the flight photographs (Fig. C.6a,e). Over global oceans, approximately half of low clouds do not fully attenuate space-borne lidar, suggesting that these optically-thin clouds contribute significantly to total cloud cover estimates (Leahy et al., 2012) and could have an important radiative impact (e.g., Wood, Tseng, et al., 2018).

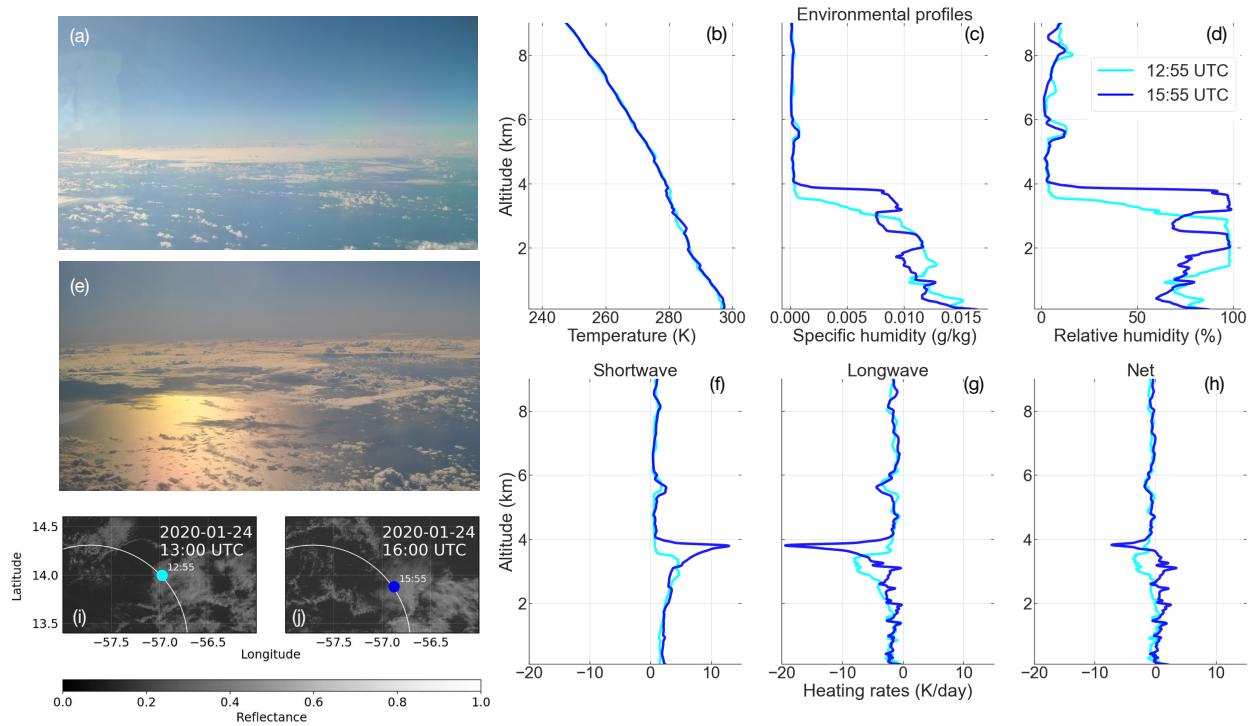


Figure C.6: Thermodynamic and radiative heating profiles associated with an elevated moisture layer persisting for multiple hours on January 24, 2020 in the HALO flight pattern. Plotted here are the temperature (b), specific humidity (c), relative humidity (d), as well as shortwave, longwave, and net radiative heating rate (f-h) profiles for two soundings sampled three hours apart, at 12:55 and 15:55 UTC. Alongside these profiles are photographs (a,b) taken from the HALO aircraft during the flight and GOES-16 satellite images (i,j), with the dropsonde location and launch time indicated by a circle along the circular flight pattern. Credit for the two flight photographs: J. Vial.

C.6 Uncertainty assessment

To evaluate the robustness of our results and ensure good use of this data set, we performed several uncertainty assessments by perturbing the SST_{skin} , in situ moisture data, and ERA profiles used. We also included in the data set the minimum and maximum levels z_{min} and z_{max} measured by each sonde. Unless indicated otherwise, the errors reported below correspond to a subset of profiles with valid data starting at 40 m (ie. $z_{\text{min}} \leq 40$ m) and during daytime, which corresponds to 1314 profiles. The daytime filter was required for relevant calculation of the error in the shortwave, and then kept for consistency for the longwave, but the magnitude of errors in the longwave is not affected by this filter (not shown).

We first test the sensitivity to the ERA5 SST_{skin} . To this end, we perturbed the original SST_{skin} by ± 0.42 K and recalculate all heating rates. This value is chosen as it corresponds to the root-mean-square-error (RMSE) between ERA5 SST_{skin} and Marine-Atmosphere Emitted Radiance Interferometers (M-AERI) measures taken during a series of cruises in the Carribean Sea from 2014 to 2019 (Luo & Minnett, 2020). Figure C.7 shows the RMSE between the original and perturbed radiative profiles (blue curves). In the longwave and net, the effect of the perturbation is strong in the first atmospheric layer, but then decreases rapidly and becomes negligible after a few hundred meters. Except for the first few atmospheric layers, the uncertainty around the SST_{skin} can therefore be safely neglected.

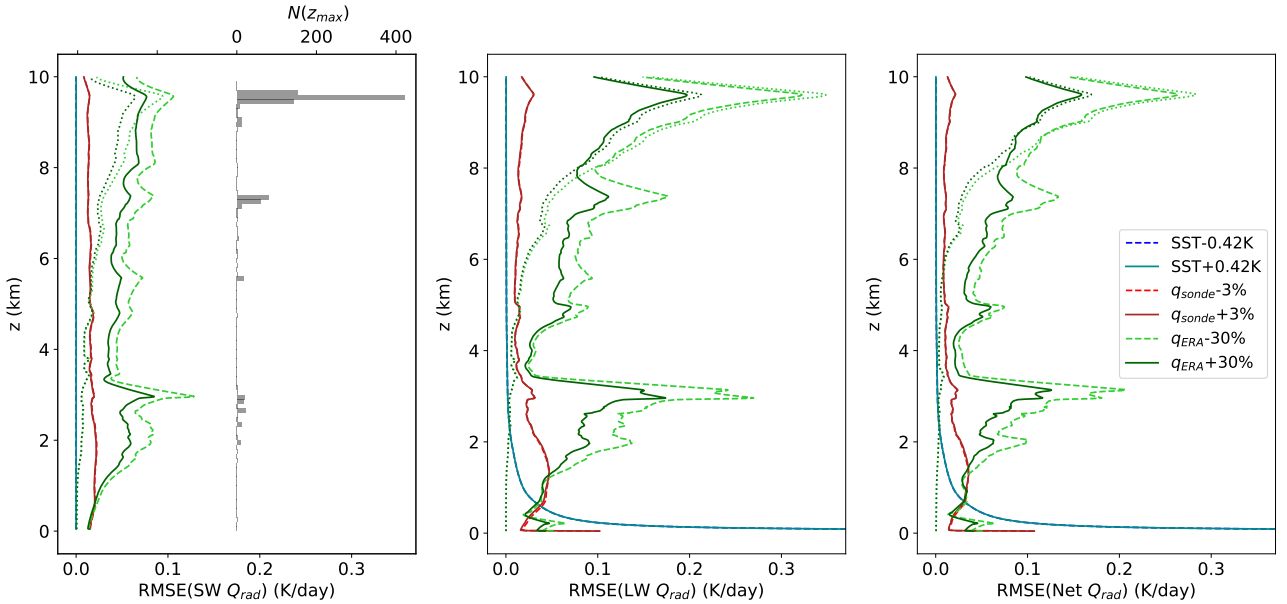


Figure C.7: Root-mean-square error estimates in shortwave (left), longwave (center) and net heating rates (right) for perturbations in SST_{skin} (blue), ERA5 humidity profiles (green) and sonde humidity measurements (red) for the 1314 daytime profiles that have valid data starting at 40 m. Dashed curves show negative perturbations, solid curves show positive perturbations and dotted green curves show ERA5 humidity perturbations restrained to the 1117 daytime profiles that have valid data at all levels between 40 m and 8 km. The horizontal grey bars on the left panel show the frequency distribution in the maximum level measured (z_{max}).

We then investigate the sensitivity to the uncertainty of sounding measurements by perturbing all soundings by a vertically-uniform relative error and redoing all radiative transfer calculations. The manufacturer predicts an uncertainty of ± 0.1 K for the temperature and ± 3

% for specific humidity (Vaisala, 2020). The temperature uncertainty has virtually no effect on radiative profiles (not shown). The effect of ± 3 % uncertainty on the specific humidity profiles is shown in Fig. C.7 in red. The highest RMSE for this specific humidity perturbation occurs in the cloud layer, between 800 m and 2 km, with a magnitude of 0.05 K/day for net radiative heating. A secondary peak with a magnitude of 0.03 K/day is also evident near the inversion, at about 3 km. Given a median radiative heating value of -1K/day throughout the lower troposphere (Sec. C.5.1), these errors are roughly 3-5% for the net radiative heating. These maxima likely correspond to the cumulative errors at the altitude of large vertical humidity gradients, which lead to peaks in longwave, and to a lesser extent shortwave heating rates for individual profiles.

Finally, we explore the uncertainty associated with ERA5 temperature and humidity data employed as an upper boundary condition. Similarly to the uncertainty analysis for the sounding data, we perturb ERA5 3D fields – used as input to the radiative transfer code – by a uniform relative error. Previous studies have shown that ERA5 reanalyses can present biases of various kinds (Dyroff et al., 2015; Nagarajan & Aiyer, 2004). We compare ERA5 humidity and temperature data with coincident radiosonde measures to obtain an estimate of ERA5 biases up to 100 hPa. From the surface to 100 hPa, the RMSE in temperature between co-located radiosonde soundings and ERA5 is between 0.3 and 0.7 K, with a mean of 0.5 K, and between 5% (at the surface) and 70 % (near the inversion) for the specific humidity, with a mean around 30 %.

Fig. C.7 only shows the effect of the ERA5 specific humidity uncertainty, taken at ± 30 %, on radiative profiles, as the temperature has once again a negligible influence. The corresponding green curves (respectively dashed and solid) reveal local maxima in the longwave and net heating rates around 3, 7 and 9.5 km. Again given a median radiative heating value of -1K/day throughout the lower troposphere (Sec. C.5.1), the errors at these local peaks are between 10-30%. These maxima coincide with the modes in the frequency distribution of the highest level z_{max} measured by the soundings, indicated in grey in the left panel. These peaks suggest that the uncertainty arises from the large discontinuities emerging at the ERA5-sounding junction level when perturbing ERA5 humidity profiles. The results suggest that the corresponding uncertainty mainly occurs in the vicinity of the junction levels. This notion is further confirmed by calculating the RMSE only on profiles which have data between 40 m and 8 km (ie. $z_{min} \leq 40$ m and $z_{max} \geq 8$ km, dotted green curve): the remaining 1117 profiles left do not contain vertical discontinuities in humidity in this range, and we see that the remaining upper-tropospheric discontinuities do not affect heating rates in the lowest troposphere.

Overall, the small uncertainty values given with these tests support the robustness of this data set and gives confidence regarding its use for more detailed investigations in the lower troposphere. The uncertainty from sea surface skin temperature is limited to the first few atmosphere layers, and uncertainty from merging with ERA5 specific humidity is largely contained to the sounding-reanalysis junction point. Uncertainty associated with observed specific humidity profiles produces localized errors in the cloud and inversion layers below 3km, though these errors are approximately 5% or less. We recommend that users carefully compare the magnitude of the signal they analyze with the magnitudes of the errors provided here.

C.7 Conclusions

The first objective of this work is to present the method used to calculate clear-sky, aerosol-free radiative profiles from 2504 radiosonde and dropsonde soundings launched during the EUREC⁴A field campaign. These radiative profiles are calculated using a state-of-the-art correlated- k model, RRTMGP, in which ERA5 reanalyses provide lower and upper boundary conditions. We then aggregate the radiative heating profiles at multiple scales to examine temporal and spatial variability in trade wind regimes. We find that radiative heating rates in the wintertime trade-wind environment display significant diurnal and day-to-day variability, and we observe hints that this variability may be associated with different types of mesoscale organization. An uncertainty assessment is further conducted to demonstrate that the influence of uncertainties in the sounding data, and upper and lower boundary conditions, is small relative to the magnitude of estimated radiative heating.

These results present a first overview of how this data set could help answer existing research questions, in particular: 1) What is the role played by radiation in the mesoscale organization of shallow convection? (e.g. C. S. Bretherton & Blossey, 2017; Seifert & Heus, 2013b) 2) what is the interplay between the diurnal variability in radiative heating, convection, and cloudiness? (e.g., W. M. Gray & Jacobson Jr, 1977; Ruppert Jr & O'Neill, 2019; Vial et al., 2019b), and 3) what is the influence of clear-sky radiative cooling gradients on atmospheric circulations? (e.g. K. Emanuel, Wing, & Vincent, 2014; W. Gray & Jacobson, 1977; Mapes, 2001; Naumann et al., 2019; Thompson, Bony, & Li, 2017). Such questions regarding the coupling of clouds, convection, and circulations in trade-wind regimes are at the heart of the EUREC⁴A field campaign, and the radiative profiles presented here complement other EUREC⁴A observations and data products in forming a toolbox for these investigations.

C.8 Code and data availability

All data are archived and freely available for public access on AERIS (Albright et al., 2020, <https://doi.org/10.25326/78>). The code used to compute the radiative profiles and python scripts used to generate the figures of the present paper are publicly released on Zenodo (<https://doi.org/10.5281/zenodo.4010195>) and Github (<https://github.com/bfildier/Albright2020>).

Appendix D

Résumé long en français

Chapitre 1 – Introduction

Historiquement, les modèles de circulation générale (MCG) ont joué un rôle crucial pour avertir les décideurs et le grand public des changements climatiques à venir. Ces modèles, au départ uniquement atmosphériques, ont été couplés à partir des années 80 à des modèles océaniques et utilisés pour comprendre le climat actuel et prédire le climat futur. Les MCG couplés ont joué un rôle clef dans les rapports du GIEC, le Groupe d’experts Intergouvernemental sur l’Evolution du Climat, qui synthétisent l’état des connaissances sur les changements climatiques et leurs conséquences sur les sociétés humaines.

Les MCG atmosphériques ont une structure duale : d’une part, ils sont fondés sur une forme particulière des équations de la mécanique des fluides, dites équations primitives, qui sont discrétisées sur une grille couvrant l’ensemble de l’atmosphère. Dans les MCG utilisés aujourd’hui, la résolution horizontale typique est comprise entre 50 et 300 km, avec une résolution verticale variant de quelques dizaines de mètres à quelques kilomètres selon l’altitude. Le pas de temps utilisé dans les MCG est typiquement de l’ordre de dix minutes, ce qui signifie que toutes les dix minutes, les variables d’état du modèle – la température, l’humidité spécifique, la pression et les vents horizontaux et verticaux notamment – sont calculées à chaque point de la grille du modèle.

Cette discrétisation permet de former la “dynamique” des MCG, qui est utilisée pour calculer les mouvements à grande échelle des masses d’air dans l’atmosphère. Cependant, de nombreux phénomènes se produisent à une échelle inférieure à celle de la grille des MCG et sont ignorés par la dynamique. Des paramétrisations sont introduites pour tenir compte malgré tout de ces processus et constituent la “physique” des MCG. En particulier, les MCG ont une grille trop grossière pour résoudre les nuages et la convection atmosphérique et utilisent des paramétrisations pour prendre en compte leur effet moyen à l’échelle de la grille du modèle. Au cours des cinquante dernières années, différentes paramétrisations de la convection ont été proposées (ex : Arakawa & Schubert, 1974; DelGenio & Yao, 1993; K. A. Emanuel & Živković-Rothman, 1999). En raison de l’importance de la convection atmosphérique pour le système climatique, ces paramétrisations de la convection sont un élément clé de tout MCG.

Cependant, au cours des 20 dernières années, des critiques ont été formulées à l’encontre des paramétrisations : en particulier, les progrès permis par les paramétrisations de la convection ont été parfois décrits comme trop lents, et certains modélisateurs ont vu dans les paramétrisations de la convection une “impasse” (D. Randall et al., 2003; D. A. Randall, 2013).

Des alternatives ont été proposées, fondées sur des modèles à plus haute résolution que les MCG ou sur du *machine learning*. Une discussion sur les fondements des paramétrisations et le type de compréhension qu’elles nous offrent pourrait aider à évaluer ce qui serait perdu ou gagné si les paramétrisations étaient remplacées par d’autres méthodes. Plus précisément, des questions épistémologiques comme : “Qu’essayons-nous de représenter lorsque nous paramétrisons la convection atmosphérique ? Dans quel sens les paramétrisations sont-elles en partie ad hoc, en partie déterminées par la théorie ? Quelles interprétations peut-on donner au formalisme employé dans les paramétrisations ? Comment ces interprétations aident-elles à mieux comprendre le rôle de l’ajustement des paramètres, ou *tuning*, dans les MCG ?” nous semblent tout à fait pertinentes dans le contexte actuel.

Pour répondre à ces questions, j’ai choisi dans ma thèse une approche originale, à l’interface entre les sciences du climat et la philosophie des sciences. J’ai eu la chance, pendant ma thèse de doctorat, de participer à plusieurs travaux scientifiques qui m’ont permis de découvrir différentes facettes de l’étude de la convection atmosphérique. En particulier, j’ai pris part à la campagne de terrain EUREC⁴A, qui s’est déroulée en janvier-février 2020 à l’ouest de l’océan Atlantique, et j’ai été impliqué dans plusieurs études qui en ont découlé (Albright et al., 2020; Stevens et al., 2021; Touzé-Peiffer et al., 2021). J’ai également implémenté une paramétrisation dans le MCG atmosphérique LMDZ et mené un travail important de re-*tuning* du modèle suite à l’ajout de cette nouvelle paramétrisation. Grâce à ce travail, j’ai acquis une connaissance pratique des défis épistémologiques posés par le *tuning* des modèles et des nouvelles perspectives offertes par les méthodes de *tuning* automatique. Enfin, j’ai été confronté au monde des simulations avec une résolution hectométrique (*large-eddy simulations*, LES) et kilométrique (*cloud-resolving models*, CRM) sur de grands domaines en contribuant à une étude définissant une méthode de détection des poches froides dans de telles simulations (Rochetin et al., 2021). Ces différents travaux sont utilisés tout au long de ma thèse comme exemples, études de cas ou prétextes pour mener une réflexion épistémologique sur les paramétrisations de la convection, leur utilisation dans les modèles, et leurs fondements. Dans la suite, je donne un résumé des différents chapitres qui composent le manuscrit.

Chapitre 2 – Nature et rôle des objets dans l’étude de la convection atmosphérique

Derrière la complexité des écoulements atmosphériques, certaines structures cohérentes jouant un rôle important pour la convection atmosphérique peuvent être distinguées et isolées d’un environnement – j’appelle de telles structures “objets”. Les objets les plus évidents sont ceux qui sont visibles à l’œil nu : un nuage ou une averse par exemple. D’autres objets sont invisibles, mais jouent également un rôle important dans la convection atmosphérique. La présence de vitesses verticales positives élevées dans une certaine région de l’espace peut par exemple être utilisée pour isoler des “ascendances” (parfois appelées “thermiques” ou “panaches”) d’un environnement avec des vitesses verticales plus faibles.

Que l’on étudie l’atmosphère à l’œil nu, avec des instruments sophistiqués ou avec des modèles numériques, chaque scène observée est unique. Les objets aident à nommer et à reconnaître certaines similitudes dans cette diversité apparente. Ils nous permettent d’identifier certains phénomènes récurrents derrière l’aspect désordonné et chaotique des écoulements atmosphériques. J’explique que les objets ne sont cependant pas donnés par la nature, mais résultent d’une construction. Trouver des objets pertinents pour décrire la convection atmosphérique est donc essentiel pour la comprendre. Les objets orientent la recherche atmosphérique vers certaines questions spécifiques, aident à formuler et à tester des hypothèses, font le lien entre des

isolées. Dans l'introduction de nombreuses études sur les poches froides, de telles questions partagées au niveau de la communauté scientifiques sont souvent présentées comme une motivation principale pour le travail proposé (voir par exemple l'introduction de Drager & van den Heever, 2017; Gentine et al., 2016; Torri et al., 2015).

En orientant la recherche atmosphérique vers certaines questions spécifiques, les objets permettent d'étudier une hypothèse donnée à l'aide d'une variété d'outils, et donc d'évaluer sa robustesse. Par exemple, au cours de la dernière décennie, différentes études ont porté sur le rôle des poches froides dans le cycle diurne de la convection sur continent. En particulier, lorsque la paramétrisation de Grandpeix and Lafore (2010) a été utilisée dans le MCG atmosphérique LMDZ, l'étude de Rio et al. (2009) a montré que le cycle diurne de la convection sur continent prédit par le modèle était décalé à la fin de l'après-midi. Cela a corrigé un biais présent dans LMDZ et partagé par de nombreux MCG, où la convection a tendance à être en phase avec l'insolation, donc à atteindre son intensité maximale aux alentours de midi, contrairement aux observations où l'intensité maximale de la convection sur continent est généralement atteinte quelques heures plus tard. Ce résultat a conduit à l'hypothèse que les poches froides jouent un rôle fondamental dans le cycle diurne de la convection sur continent. Par la suite, cette hypothèse a été testée par Haerter et al. (2019) dans une simulation LES idéalisée, puis par Hirt and Craig (2021) dans une simulation CRM réaliste sur l'Allemagne. Haerter (2019) a ensuite proposé d'expliquer le lien entre les poches froides et le cycle diurne de la convection à l'aide d'un modèle conceptuel. Une hypothèse donnée – le rôle des poches froides dans le cycle diurne de la convection – a ainsi pu être testée et confirmée en utilisant une variété d'outils. Cela illustre comment les objets peuvent servir de catalyseur pour guider les recherches autour de certaines hypothèses bien définies.

Les objets aident aussi à **résumer notre compréhension d'un certain phénomène**. Ils donnent en effet des mots auxquels nous pouvons attacher notre compréhension de certains phénomènes qui partagent des caractéristiques similaires. Par exemple, si deux études A et B prouvent avec succès deux hypothèses différentes sur le même phénomène, nous pourrions relier les résultats de A et B uniquement parce que nous avons un mot pour nommer ce phénomène, donc pour le reconnaître à la fois dans A et dans B. Plus généralement, les objets nous aident à relier des études qui donnent toutes une vision partielle d'un phénomène afin de comprendre ce phénomène dans son intégralité. En résumant ce que nous savons, les objets aident aussi à évaluer les limites de nos connaissances et donc à repousser ces limites. Nous montrons ce pouvoir des objets à travers le schéma conceptuel proposé par Rochetin et al. (2021) (voir 2.6), qui fait le lien entre différentes études et résume ainsi à la fois notre compréhension actuelle des poches froides et les questions en suspens.

Enfin, **les objets donnent des principes pour analyser observations et simulations à haute résolution**. Comme l'écrit Auguste Comte: "Si en contemplant les phénomènes, nous ne les rattachions point immédiatement à quelques principes, non seulement il nous serait impossible de combiner ces observations isolées, et, par conséquent, d'en tirer aucun fruit, mais nous serions même entièrement incapables de les retenir" (Comte, 1835). Selon le philosophe, nous avons besoin de certains principes pour étudier les phénomènes. Ces principes déterminent ce que nous pouvons apprendre des observations. Si l'on en croit cette théorie, nos connaissances accumulées sur les poches froides nous permettraient donc en retour de mieux les observer. Je l'illustre par deux études récentes sur les poches froides dans lesquelles j'ai été impliqué : 1) Rochetin et al. (2021), qui propose une méthode pour détecter les poches froides dans des

simulations CRM et LES et 2) Touzé-Peiffer et al. (2021), qui analyse les poches froides dans des régimes d'alizés à partir de sondages collectées pendant la campagne EUREC⁴A. Dans chacun de ces deux exemples, la compréhension que les auteurs ont des poches froides a été cruciale pour construire une méthode de détection la plus pertinente possible en fonction des données disponibles. En retour, les méthodes de détection utilisées ont permis de mieux comprendre les poches froides. Pour l'illustrer, je cite des exemples de nouveaux résultats qu'ont permis de montrer ces deux études.

Notre compréhension des objets est donc en grande partie incrémentale: ce que nous comprenons d'un objet nous donne un angle d'attaque pour l'étudier dans des observations ou dans des simulations à haute résolution, qui nous permettent de le comprendre encore davantage.

Chapitre 3 – Le monde semi-empirique des paramétrisations de la convection

De nombreux biais des MCG ont été attribués à la mauvaise représentation des processus sous-maille dans les paramétrisations, qui ont parfois été qualifiées de “talon d'Achille” des modèles climatiques (Stone & Risbey, 1990). En particulier, les paramétrisations de la convection sont souvent décrites comme une source majeure d'incertitude pour les projections climatiques (Stevens & Bony, 2013). De plus, bien que des progrès significatifs aient été réalisés depuis l'introduction des premières paramétrisations de la convection dans les MCG il y a 50 ans (Rio et al., 2019), certains ont critiqué la lenteur de ces progrès et décrit les paramétrisations de la convection comme une “impasse” (T. Palmer & Stevens, 2019; D. Randall et al., 2003).

Les critiques contre les paramétrisations ne concernent pas seulement leur effet dans les MCG, mais aussi leur fondement scientifique. En particulier, la **nature semi-empirique des paramétrisations est souvent critiquée**. Les paramétrisations ne sont en effet pas entièrement dérivées de principes fondamentaux, mais incluent des hypothèses ad hoc et des paramètres ajustés pour garantir une certaine cohérence avec les observations. Edwards (2001) soutient que la “relation floue” des paramétrisations avec les données d'observation est en contradiction avec l'impératif réductionniste des sciences physiques, qui tentent “d'expliquer les phénomènes à grande échelle comme un résultat de processus à plus petite échelle”. Cependant, Edwards (2001) n'explique pas pourquoi les paramétrisations devraient répondre aux normes réductionnistes. Plus généralement, alors que de nombreux philosophes des sciences ont mentionné et parfois critiqué le caractère semi-empirique des paramétrisations (ex : Lloyd, 2009; Petersen, 2012), peu d'études ont défini précisément dans quel sens les paramétrisations sont en partie déterminées par la théorie, en partie déterminées par les observations, et donc si et pourquoi leur nature semi-empirique est légitime, ou non.

Je tente de répondre à ces questions en adoptant une approche partiellement historique pour expliquer le contexte dans lequel les paramétrisations ont émergé et comment le problème de la paramétrisation de la convection a initialement été formulé et a évolué par la suite. La genèse des paramétrisations de la convection tient dans le fait que, lorsque les premiers MCG atmosphériques ont été développés dans les années 60, un problème majeur est apparu : les profils atmosphériques prédits par les MCG dans les tropiques devenaient rapidement instables, ce qui entraînait des mouvements verticaux intenses à l'échelle de la grille du modèle. Cela était dû au fait que la grille utilisée dans les MCG, avec une résolution horizontale de l'ordre de quelques centaines de kilomètres, était trop grossière pour résoudre la convection atmosphérique. La convection agit comme une rétroaction négative sur les instabilités de grande échelle, mais elle a lieu à une échelle très inférieure à celle de la grille des MCG : pour résoudre la convection

atmosphérique, il faut utiliser des modèles numériques avec une résolution horizontale d'un kilomètre ou moins. La présence de convection atmosphérique à l'échelle de la grille dans les premiers MCG était donc très peu réaliste : lorsqu'elle se développait, les erreurs augmentaient de façon exponentielle dans ces modèles (Kasahara, 1961; Lilly, 1962).

Pour obtenir une circulation atmosphérique à grande échelle réaliste à l'aide des MCG, il était nécessaire de trouver un moyen de tenir compte de la présence de la convection atmosphérique dans chaque maille d'un MCG, et donc d'empêcher les profils atmosphériques simulés de devenir instables. Cela a motivé l'apparition des premières paramétrisations de la convection atmosphérique, qui étaient au départ un simple ajustement des gradients de température dans l'atmosphère lorsque les profils correspondants devenaient instables – une procédure appelée “ajustement convectif humide” (Manabe & Strickler, 1964). Cette stratégie a aidé les premiers MCG à surmonter leurs difficultés à prédire des profils thermodynamiques verticaux réalistes.

Les auteurs mêmes de ces ajustements les décrivaient cependant comme ad hoc, imparfaits, et souhaitaient qu'ils soient rapidement remplacés par des paramétrisations davantage ancrées dans la théorie (Manabe et al., 1965; Smagorinsky et al., 1965). Rapidement, des théories ont ainsi été proposées pour formuler le problème de la convection atmosphérique. La plupart de ces théories reposaient sur un formalisme appelé le **formalisme en flux de masse**, qui tient son inspiration dans l'hypothèse des “tours chaudes”, initialement formulée par Riehl and Malkus (1958). Cette hypothèse explique le bilan thermodynamique associé à la circulation de Hadley dans les tropiques en considérant que l'apparente ascension à grande échelle près de l'équateur est en fait la somme d'une ascension vigoureuse dans quelques “tours chaudes” isolées et d'une descente dans leur environnement. Selon cette hypothèse, les tours chaudes ne couvrent qu'une petite fraction des Tropiques et ne se mélangent à leur environnement qu'à leur base et à leur sommet. Cela leur permet de transporter la chaleur directement de la surface jusqu'au sommet de la troposphère.

L'hypothèse des tours chaudes a considérablement influencé le développement des premières paramétrisations de la convection. Reprenant à son compte les idées de Riehl, Ooyama (1964) a notamment formulé un modèle dynamique pour l'étude du développement des cyclones tropicaux dans lequel les effets de chauffage des nuages étaient représentés suivant une forme de flux de masse. Quelques années plus tard, Arakawa (1969) a adapté l'approche en flux de masse aux MCG.

Cette approche deviendra rapidement la manière courante de représenter la convection atmosphérique dans les paramétrisations. La forme du flux de masse suppose que la convection atmosphérique se produit dans des “panaches convectifs” couvrant une petite partie de la maille, conformément à l'hypothèse des tours chaudes de Riehl. Dans cette approche, le transport vertical dans les panaches convectifs est représenté par un flux de masse défini par $f = \rho \alpha w_u$, avec une vitesse verticale w_u supposée homogène horizontalement. ρ désigne la densité et α la fraction occupée par les panaches dans la maille. De même, le transport vertical dans l'environnement – la subsidence compensatoire – est supposé avoir une vitesse horizontalement homogène w_e et un flux de masse correspondant $f_e = \rho(1 - \alpha)w_e = -f$. Dans les premières paramétrisations telles que celle proposée par Ooyama (1964), les flux de masse convectifs et environnementaux (et donc w_u et w_e) sont supposés constants. Par conséquent, l'air n'est entraîné dans le flux de masse qu'à la base des nuages et n'est détaché qu'à leur sommet. Dans ce schéma simple, le nombre d'inconnues est réduit à une seule, le flux de masse à la

base du nuage f , qui doit être spécifié en fonction de l'état à grande échelle. Le problème de la spécification du flux de masse convectif - et donc de l'intensité de la convection - en fonction de l'état à grande échelle de l'atmosphère, est appelé "fermeture".

A partir de ce formalisme ou de ses variations, de nombreuses paramétrisations de la convection ont été proposées, qui diffèrent notamment par leur fermeture. Toutes produisent une rétroaction négative sur les instabilités de grande échelle et ont donc globalement le même effet que les schémas d'ajustement convectifs humides. Cependant, alors que ces derniers imposent la solution à partir des profils typiques observés dans les observations, les paramétrisations en flux de masse donnent un cadre théorique pour expliquer pourquoi et comment l'atmosphère est stabilisé par la convection.

Différentes interprétations ont été données au formalisme en flux de masse. Ces interprétations ne sont pas des justifications, mais proposent des analogies pour interpréter le cadre théorique des paramétrisations en flux de masse. Nous nous attachons à trois types d'analogies utilisées pour interpréter le formalisme en flux de masse : 1) la paramétrisation de Arakawa and Schubert (1974), qui repose sur la notion de quasi-équilibre convectif que l'on peut comprendre par analogie avec la théorie quasi-géostrophique. Comme l'explique Schubert (2000), le quasi-équilibre filtre l'ajustement transitoire d'un ensemble de nuages de la même façon que la théorie quasi-géostrophique filtre les ondes de fréquences élevées. 2) L'interprétation statistique de la convection, qui a inspiré la paramétrisation stochastique de la convection de Craig and Cohen (2006). Par analogie avec la physique statistique, Craig and Cohen (2006) font l'hypothèse qu'un certain équilibre est satisfait à grande échelle, mais que l'effet de la convection dans une maille donnée est un sous-ensemble de l'équilibre de grande échelle, qui doit être tiré d'une distribution déterminée par les conditions de grande échelle. 3) Enfin, nous examinons l'interprétation donnée par Yano (2014) qui s'intéresse au flux de masse en lui-même plutôt qu'à sa fermeture. Il interprète le formalisme en flux de masse comme une approche en volume fini appliqué à la grille du modèle et, à partir de cette analogie, dérive les équations du flux de masse par approximations successives.

Les analogies avec les théories sont intéressantes et fournissent une base pour construire des paramétrisations, mais dans de nombreux cas, elles ne sont pas suffisantes. Par conséquent, nous ne devrions pas nous fier exclusivement à la théorie et accepter d'utiliser des hypothèses ad hoc pour rendre nos paramétrisations plus réalistes. Les schémas d'ajustement convectifs, directement fondés sur des observations, et les trois paramétrisations décrites ci-dessus, qui reposent sur des analogies avec certaines théories, doivent donc être considérés comme des positions extrêmes. Les paramétrisations utilisées dans le modèle atmosphérique LMDZ – et dans de nombreux autres MCG – se fondent sur des objets et sont à cheval entre ces deux positions antagonistes.

Les objets fournissent en effet un cadre flexible pour combiner les informations provenant des observations et de la théorie dans les paramétrisations de la convection. Une paramétrisation fondée sur un objet n'est ni entièrement empirique comme les profils de référence utilisés dans les schémas d'ajustement convectifs humide, ni entièrement dérivée de principes fondamentaux : elle repose en partie sur des observations et en partie sur certaines considérations théoriques. La théorie fournit un cadre pour décrire l'objet à partir de certains principes physiques. Des hypothèses plus ad hoc, fondées sur des observations, sont utilisées pour garantir que cette base théorique soit suffisamment proche du monde réel. Les paramétrisations basées sur les objets

sont donc semi-empiriques : les hypothèses qu’elles utilisent sont motivées en partie par des raisons théoriques, en partie par des analyses empiriques.

Les objets constituent un moyen naturel et pratique d’analyser à la fois les observations et les simulations à haute résolution. Lorsque nous observons une scène nuageuse, nous voyons des objets qui interagissent les uns avec les autres. L’analyse de ces interactions permet d’abord de caractériser les objets, puis de les comprendre. Penser les paramétrisations avec les objets aide à utiliser cette compréhension à la fois pour construire et pour tester les paramétrisations. Par exemple, la vitesse de propagation des poches froides dans la paramétrisation de Grandpeix and Lafore (2010) peut être directement comparée à la vitesse de propagation des poches froides dans les observations ou les simulations à haute résolution. S’il y a un décalage entre ce qui est prédit dans une paramétrisation et les observations, certaines hypothèses de la paramétrisation peuvent être ajustées en fonction des données des observations. Cela permet de s’assurer que les objets représentés dans les paramétrisations soient aussi proches que possible du monde réel.

A l’inverse, une paramétrisation fondée sur une certaine théorie prend plus de distance par rapport aux phénomènes observés. Elle tente de déduire une représentation cohérente de la convection à partir de principes fondamentaux. Cette déduction logique donne un cadre strict pour représenter les phénomènes réels. Si ce cadre ne donne pas de résultats réalistes, il est difficile de tirer des enseignements des observations pour améliorer la théorie employée. Une nouvelle théorie logiquement cohérente doit être proposée pour résoudre le problème identifié. Il est bien sûr impossible de formuler une théorie entièrement nouvelle pour chaque biais identifié dans le modèle. Par conséquent, la cohérence n’est pas un gage de fiabilité : une paramétrisation peut être parfaitement cohérente et en même temps très irréaliste, c’est-à-dire éloignée des phénomènes atmosphériques réels. Pour les paramétrisations, la cohérence n’est donc pas une panacée. Ce n’est qu’un critère, parmi d’autres, à prendre en compte pour évaluer leur qualité.

Le monde semi-empirique des paramétrisations fondées sur des objets n’est donc pas une faiblesse, mais une force. Le cadre qu’il donne est plus flexible que le cadre strict et logiquement cohérent imposé par des paramétrisations fondées sur une certaine théorie. Les objets nous permettent de reconnaître que les représentations utilisées dans les paramétrisations sont imparfaites, de sorte qu’une seule représentation cohérente de la convection n’est pas suffisante. Chaque objet a un pouvoir explicatif et prédictif limité. Cependant, penser avec des objets aide à représenter dans les paramétrisations des images physiques qui sont certes en partie contradictoires, mais aussi complémentaires pour donner une image réaliste de la convection atmosphérique.

Chapitre 4 – Epistémologie du *tuning* des modèles de climat

Comme leur nom l’indique, les paramétrisations contiennent des paramètres plus ou moins bien contraints par les observations. Lorsqu’une paramétrisation est développée, la valeur de chacun de ses paramètres est fixée de façon plus ou moins arbitraire, en tenant compte de certains diagnostics dérivés d’observations, de l’expertise du modélisateur, ou des résultats de modèles simplifiés comme des modèles 1D utilisés dans le développement de la paramétrisation. Lorsque des simulations sont faites à partir d’un MCG 3D (atmosphérique ou couplé), certains paramètres sont à nouveau réglés pour améliorer la qualité des résultats des simulations et garantir notamment que les températures de surface prédites par le MCG couplé soient suffisamment proches de celles du climat actuel. Ce second réglage des paramètres est généralement

appelé *tuning* (Mauritsen et al., 2012; D. A. Randall & Wielicki, 1997).

La plupart des MCG couplés sont réglés en particulier avant chaque phase du projet d'inter-comparaison de modèles couplés (CMIP, pour *Coupled Model Intercomparison Project*). Dans le cadre de CMIP, tous les 5 ou 6 ans, les principaux MCG couplés utilisés dans le monde sont soumis à des protocoles de simulation standardisés. Cet effort représente une opportunité de comparer les dernières versions des MCG couplés dans des conditions similaires, et notamment de comparer leurs prédictions climatiques afin de fournir des déclarations pour les rapports du GIEC. Les résultats des MCG couplés dépendent dans une large mesure du *tuning* de leurs paramètres. Pourtant, jusqu'à récemment, les procédures de *tuning* employées n'étaient pas bien documentées, et donc pas prises en compte dans les nombreuses analyses fondées sur les résultats de CMIP. Contrairement à la calibration des instruments de laboratoire, qui suit généralement un protocole strict (ex : Manrique et al., 2020), le *tuning* des MCG n'est pas standardisé au niveau de la communauté de modélisation du climat. Les pratiques de *tuning* varient d'un centre de modélisation à l'autre, et les critères utilisés pour le réglage des paramètres sont en partie subjectifs et dépendent de l'expertise personnelle des modélisateurs climatiques. Ce n'est que récemment que les modélisateurs climatiques ont ressenti le besoin de documenter et de comparer systématiquement les procédures de *tuning* qu'ils utilisaient (Hourdin et al., 2017a; Mauritsen & Roeckner, 2020; Mignot et al., 2021).

À l'inverse, au cours des vingt dernières années, le *tuning* des modèles climatiques a attiré l'attention de nombreux philosophes des sciences. Ces derniers se sont intéressés notamment à l'interprétation de l'accord entre les résultats des MCG et les observations : si les observations sont utilisées pour ajuster les paramètres du modèle, que signifie l'accord entre les résultats des MCG et les observations ? Doit-il renforcer notre confiance dans les modèles, donc dans les projections climatiques futures ? Ou au contraire, n'est-il que la conséquence du processus de *tuning* ? De nombreux philosophes ont en effet noté le risque qu'il y a à utiliser des observations parfois similaires pour régler les paramètres d'un modèle et évaluer ses résultats (W. S. Parker, 2011; Stainforth et al., 2007). D'autres, au contraire, sont plus optimistes et soulignent la diversité des preuves utilisées pour tester les modèles climatiques comme un argument en faveur de leur adéquation empirique (Lloyd, 2009, 2010).

Les philosophes des sciences se sont cependant davantage intéressés aux conséquences du *tuning* sur l'interprétabilité des MCG, qu'au processus de *tuning* en lui-même. En particulier, la variété des stratégies utilisées dans le processus de *tuning* est peu abordée dans la littérature philosophique. Les paramètres utilisés pour le *tuning* sont rarement spécifiés, tout comme les besoins auxquels le *tuning* répond. De plus, nous n'avons trouvé aucune tentative dans la littérature de définir ce qui rend le *tuning* des modèles climatiques différent des procédures de calibration utilisées dans d'autres domaines scientifiques. Une telle comparaison serait précieuse pour étudier dans quelle mesure les arguments utilisés pour justifier la calibration s'appliquent au *tuning* des MCG.

Je tente de répondre à ces questions en proposant tout d'abord une typologie des paramètres utilisés dans les MCG. Aujourd'hui, les MCG couplés contiennent typiquement plus d'une centaine de paramètres. Je propose quatre critères pour distinguer à différents niveaux le rôle de ces paramètres dans un MCG : 1) Que représente le paramètre ? 2) Est-il observable, c'est-à-dire a-t-il un équivalent empirique ? 3) Comment s'exprime-t-il mathématiquement ? 4) Quel effet le paramètre a-t-il sur les résultats du modèle ? À partir de ces quatre critères, je

montre que tous les paramètres n'ont pas la même importance pour le *tuning* des modèles de climat. Les paramètres les plus importants sont ceux qui ne sont pas bien contraints par les observations et ont, en même temps, un impact important sur certains aspects essentiels des résultats du modèle – typiquement, les paramètres impliqués dans la convection atmosphérique (Hourdin et al., 2017a).

Je donne ensuite trois raisons fondamentales pour justifier le *tuning* des modèles de climat. La première raison est que **la valeur de certains paramètres n'est pas connue avec précision**. En effet, certains paramètres n'ont pas d'équivalent dans le monde réel. D'autres ont un équivalent, mais ce dernier n'est pas facile à calculer en pratique. Plutôt que de donner une valeur arbitraire à de tels paramètres, il semble légitime d'utiliser une certaine procédure de *tuning* pour les ajuster. Cet argument s'applique en fait également aux paramètres qui ont un équivalent facilement mesurable dans le monde réel, car ce dernier ne fournit qu'une contrainte assez faible sur la valeur que le paramètre correspondant doit prendre dans le modèle. J'explique que cela est lié à la non-exactitude des représentations utilisées dans les paramétrisations. Ces dernières s'inspirent du monde réel, mais contiennent de nombreuses simplifications et idéalizations. En raison de ces approximations, les paramètres qu'elles contiennent ne sont pas censés correspondre parfaitement aux paramètres mesurés dans le monde réel.

Une deuxième raison d'ajuster les paramètres des MCG est la nécessité de **simuler un bilan énergétique de la Terre correct**. Ce dernier représente l'équilibre entre l'énergie que la Terre reçoit du soleil et celle qu'elle renvoie vers l'espace. Si le bilan énergétique de la Terre n'est pas respecté dans un MCG couplé, le climat simulé par le modèle dérivera vers un état climatique très différent du climat actuel (Marotzke & Stone, 1995; Murphy, 1995). Concrètement cela se manifestera par une augmentation ou une diminution progressive des températures de surface de la mer (SSTs pour *sea surface temperatures*) dans le modèle utilisé dans les conditions actuelles. Cette réponse sera potentiellement plus forte que la réponse transitoire simulée par le modèle à une augmentation de la concentration en CO₂ dans l'atmosphère et rendra donc difficile l'utilisation du modèle pour étudier les changements climatiques.

Enfin, le *tuning* peut être utilisé pour **rendre d'autres aspects des résultats des MCG plus réalistes**, c'est-à-dire plus conformes aux observations. Certains groupes de modélisation considèrent en effet qu'il est nécessaire de régler certains aspects clés du climat afin d'avoir des MCG couplés suffisamment réalistes pour être utiles à l'étude des changements climatiques. En conséquence, ils ont recours au *tuning* pour garantir, par exemple, le réalisme du volume ou de l'étendue de la banquise, celui de la circulation atmosphérique ou encore celui de l'augmentation des températures de surface au cours du XXe siècle (Hourdin et al., 2017b). Néanmoins, une telle utilisation du *tuning* est plus controversée. Certains modélisateurs soulignent en effet qu'un *tuning* trop important pourrait introduire des erreurs qui se compensent dans les résultats des MCG et conduire à une confiance excessive dans l'interprétation de ces résultats.

J'étudie ensuite comment le *tuning* est réalisé en pratique, c'est-à-dire quelles stratégies sont utilisées pour régler les paramètres des MCG. Il est tout d'abord possible d'ajuster les paramètres d'un MCG en utilisant **la version 1D du modèle**, qui contient une seule colonne atmosphérique du modèle et peut être forcée avec des conditions représentatives de diverses situations atmosphériques. Les modèles 1D constituent un cadre utile pour comprendre l'effet des paramétrisations et régler la valeur des paramètres sans rentrer dans la complexité des

interactions entre la physique (les paramétrisations) et la dynamique (les équations primitives résolues à l'échelle de la grille) d'un MCG. Cependant, les propriétés globales d'un MCG doivent elles aussi être ajustées, notamment les flux radiatifs au sommet de l'atmosphère, pour simuler correctement le bilan énergétique de la Terre. Ces propriétés peuvent être réglées d'abord **dans un MCG uniquement atmosphérique** avec des conditions aux limites bien définies, puis **dans un MCG couplé**. Le réglage d'un MCG couplé complet est un long processus – trois ans pour la version CMIP6 du MCG de l'IPSL (Mignot et al., 2021) – avec de nombreux allers-retours entre du *tuning* dans des modèles 1D, dans des versions uniquement atmosphériques, et dans des versions couplées.

Je me demande ensuite dans quelle mesure le *tuning* est similaire à une procédure de calibration utilisée pour calibrer des instruments de mesure plus classiques. Calibrer un instrument, c'est "utiliser un signal de substitution pour le standardiser" selon le philosophe des sciences Allan Franklin (Franklin, 1999). Le signal de substitution est en général **fiable** et **adapté** à l'utilisation qu'on souhaite faire de l'instrument. Il y a d'ailleurs souvent un consensus à l'échelle d'une communauté de spécialistes sur les références ou étalons qu'il faut utiliser pour calibrer un certain instrument. La procédure de calibration elle-même est généralement **standardisée** et **bien documentée**. Cette standardisation permet la **reproductibilité** de toute mesure. Une mesure est reproductible s'il est possible de la reproduire en laboratoire et d'obtenir les mêmes résultats avec un bon degré d'accord. Comme les mesures d'un instrument dépendent de sa calibration, la calibration utilisée doit être standardisée pour ne pas introduire de différences dans les résultats de l'expérience.

A première vue, le *tuning* des modèles de climat semble partager de nombreuses caractéristiques avec une procédure traditionnelle de calibration. Lorsque les paramètres d'un modèle climatique sont réglés, ils sont comparés à certaines références bien choisies, et ajustés pour répondre à certains critères en fonction de l'utilisation prévue pour le modèle. Cependant, le *tuning* des modèles de climat est à bien des égards plus complexe qu'une procédure de calibration. **Une première source de complexité provient des références utilisées.** Comme dans toute procédure de calibration, la précision et l'adéquation des références choisies pour le réglage des modèles climatiques sont cruciales. Or, certaines observations utilisées comme références, les SSTs par exemple, sont relativement bien connues, tandis que d'autres souffrent de grandes incertitudes – les observations de précipitations, notamment (Stephens et al., 2012). L'adéquation des références utilisées soulève encore plus de questions. Les modèles climatiques sont principalement utilisés pour simuler le climat futur, potentiellement très différent du climat actuel. Il n'est donc pas certain que les observations du climat actuel fournissent des contraintes pertinentes pour la prédiction des changements climatiques. Mais, par définition, il n'y a pas d'observations du climat futur (hypothétique) qui puissent être utilisées comme références pour le *tuning*.

La deuxième source de complexité provient des modèles climatiques eux-mêmes. **Les modèles climatiques ont tous des limites structurelles** : même les modèles climatiques finement réglés ne parviennent pas à reproduire certaines caractéristiques importantes du climat, par exemple la propagation de l'oscillation de Madden-Julian (Ahn et al., 2020; Jiang et al., 2015; Lin et al., 2006), ou la structure de la zone de convergence intertropicale (Lin, 2007; Oueslati & Bellon, 2015). En raison de ces limitations, et parce que les références utilisées ne sont ni très précises ni parfaitement adéquates, le *tuning* ne vise pas à garantir une correspondance exacte entre les résultats des modèles climatiques et les observations utilisées comme

références. Il s'agit d'une autre différence importante avec les procédures de calibration traditionnelles. Puisque les MCG souffrent de différents biais, et puisque les références utilisées pour le *tuning* ne sont pas parfaites, un désaccord parfois important entre les résultats du modèle et les références est acceptable.

Le réglage d'un modèle climatique s'apparente donc à la recherche d'un certain optimum dans l'espace des paramètres en tenant compte des diverses références choisies. Cette recherche est rendue complexe en raison du **grand nombre de paramètres et des divers modules (les paramétrisations) en interaction** que contient un MCG. En principe, la modularité d'un instrument rend sa calibration plus simple. En effet, la calibration d'un instrument contenant plusieurs modules se réduit généralement à une série de calibrations de ses différents modules. Il n'est pas possible de procéder de cette façon pour le *tuning* des modèles climatiques en raison de ce que Lenhard and Winsberg (2010) appellent la "modularité floue" des MCG : les paramétrisations sont trop interdépendantes pour être réglées séparément. Les diverses interactions entre les paramétrisations doivent être prises en compte dans le *tuning* et les paramètres ne peuvent pas être ajustés un par un : l'espace complet des paramètres doit être considéré. En raison du nombre de paramètres impliqués dans les modèles climatiques et du coût des simulations numériques, il est cependant impossible d'explorer l'espace complet des paramètres. La recherche d'un optimum est donc dépendante du chemin suivi et il existe un risque d'être piégé dans un optimum local et de manquer l'optimum global de l'espace des paramètres.

Ces difficultés permettent de comprendre pourquoi, contrairement au réglage de la plupart des instruments de laboratoire ou du quotidien, le *tuning* des modèles de climat n'est pas encore standardisé. Il n'y a pas de consensus sur la meilleure approche pour régler un modèle climatique, et les stratégies de *tuning* utilisées varient beaucoup d'un centre de modélisation à l'autre.

Il existe par ailleurs une peur d'aller trop loin dans le processus de *tuning*. Selon D. A. Randall and Wielicki (1997), par exemple, le *tuning* pourrait "empêcher artificiellement un modèle de produire un mauvais résultat". Randall et Wielicki reconnaissent que, dans certaines situations, le *tuning* est une nécessité, mais soutiennent qu'il devrait se limiter aux processus qui sont à la fois importants pour les résultats du modèle et mal compris. Par conséquent, les paramètres qui ont un équivalent empirique direct et qui peuvent être mesurés devraient être fixés une fois pour toutes avant l'exécution du modèle.

W. S. Parker (2011) nuance ce point de vue en soutenant qu'il ne s'applique qu'aux modèles qui sont censés "fournir une représentation très précise des aspects (pertinents) d'un système cible, avec des paramètres qui ont des corrélations physiques claires", ce qui n'est pas le cas de la modélisation du climat aujourd'hui. Le choix de certaines valeurs de paramètres différentes de celles trouvées dans les observations peut en effet être justifié en raison de la non-exactitude des représentations utilisées dans les modèles climatiques. Lorsqu'ils développent une nouvelle paramétrisation, les modélisateurs climatiques ont certainement une image physique en tête : des courants ascendants transportant la chaleur et l'humidité de la surface vers la couche nuageuse, des poches froides s'étalant à la surface, etc. Ces images physiques sont des sources d'inspiration pour développer des paramétrisations. Elles fournissent un cadre conceptuel pour établir des relations entre les différentes variables du modèle. Cependant, il est important de faire une distinction dans une paramétrisation entre ce que l'on cherche à reproduire et ce que l'on cherche à représenter. Le but des paramétrisations est de *reproduire* certaines *variables*

cibles à l'échelle de la grille – les tendances de température et d'humidité, la fraction nuageuse moyenne, les précipitations de surface, etc. – grâce à une *représentation* des processus sous-maille. En tant que représentations, *les variables et paramètres intermédiaires* ne sont donc pas censés correspondre parfaitement au monde réel.

Le *tuning* est parfois accusé d'introduire des **erreurs qui se compensent** dans les modèles de climat. Ce que nous venons de dire sur la non-exactitude des paramétrisations montre qu'il n'y a pas à proprement parler d'erreurs qui se compensent dans les paramétrisations. Les paramétrisations peuvent contenir des bugs et des incohérences, mais le terme "erreur" prête à confusion parce qu'il implique que les représentations utilisées dans les paramétrisations pourraient en théorie être justes. Au contraire, en tant que représentations idéalisées de la réalité, les paramétrisations ne sont ni vraies, ni fausses. Elles peuvent simplement conduire à des résultats plus ou moins réalistes.

Cela ne signifie pas qu'un modèle climatique dans son ensemble n'est pas testable à partir d'observations. Nous pourrions trouver des erreurs – et même des erreurs qui se compensent – dans les *variables cibles* des modèles climatiques. Par exemple, Winsberg (2018b) explique qu'un modèle climatique peut décrire avec précision l'effet des rétroactions nuageuses à l'échelle mondiale, en surestimant le degré de formation des nuages de basse altitude dans les régions tropicales, et en sous-estimant le degré de formation des nuages de haute altitude aux latitudes moyennes. On peut parler ici d'erreurs qui se compensent car ces erreurs concernent des variables cibles du modèle, que l'on cherche à reproduire plutôt qu'à représenter, et non des *variables intermédiaires* utilisées dans les paramétrisations qui ne sont pas censées être reproduites avec exactitude.

Le terme d'**erreur structurelle** est parfois utilisé pour désigner les erreurs qui résistent au *tuning*, c'est-à-dire qui sont associés à la structure même du modèle, indépendamment du choix des paramètres. Ce terme suggère la présence d'erreurs dans la formulation du modèle, donc dans les paramétrisations. Or, avec ce que nous avons vu, les paramétrisations ne sont ni correctes, ni erronées. Il y a seulement des choses qu'elles permettent de faire et d'autres qu'elles ne permettent pas. Je défends donc qu'il faudrait plutôt parler des "limites structurelles" d'un modèle de climat que de son erreur structurelle.

Définir les limites structurelles d'un modèle me permet aussi de préciser le sens du "sur-ajustement" des paramètres, ou **overtuning**. Nous définissons un modèle sur-ajusté comme un modèle pour lequel on peut trouver un meilleur équilibre dans les résultats du modèle compte tenu de ses limites structurelles et des préférences personnelles du modélisateur. Il y a donc une composante subjective dans la définition de l'*overtuning* : un modèle qui semble bien équilibré pour certains modélisateurs climatiques compte tenu de ses limites structurelles peut sembler sur-ajusté pour d'autres. Des modélisateurs climatiques différents peuvent en effet avoir des priorités différentes et, en fonction de ces priorités, tolérer plus ou moins certaines erreurs.

Je termine mon étude épistémologique du *tuning* en m'intéressant au changement de paradigme apporté par les méthodes de *tuning* automatique. Traditionnellement, le *tuning* des MCG est effectué à la main, le plus souvent en suivant une méthode d'essai-erreur. Comme l'espace des paramètres – l'ensemble de toutes les combinaisons entre les valeurs des paramètres – est extrêmement vaste, toutes les valeurs des paramètres ne peuvent pas être testées. Il n'est pas possible de connaître de manière exhaustive les limites structurelles d'un modèle climatique.

Ajuster les paramètres d'un modèle est donc une tâche délicate, qui fait appel non seulement à des compétences scientifiques, mais aussi à une grande part d'intuition et de jugement d'expert (Hourdin et al., 2017a). Le réglage complet d'un MCG couplé peut prendre plusieurs mois à plusieurs années et contrairement à une procédure de calibration traditionnelle est en partie subjectif et difficile à reproduire. De plus, le manque de standardisation du *tuning* complexifie considérablement la comparaison de deux MCG, ou de deux versions d'un GCM. En effet, sans un protocole de *tuning* standardisé, il est difficile de distinguer les différences de résultats qui sont dues à des choix de *tuning* différents, et celles qui sont dues à des différences dans la formulation des paramétrisations.

Des progrès récents dans les méthodes de *tuning* automatique pourraient rendre le processus de *tuning* plus rapide et plus standardisé, donc plus similaire à une procédure de calibration traditionnelle. En particulier, l'explorateur *htexplo* (pour HIGH-TUNE explorer) est un outil développé récemment par deux groupes français de modélisation climatique (IPSL et CNRM) en collaboration avec l'Université d'Exeter (Couvreur et al., 2021; Hourdin, Williamson, et al., 2020). Pour utiliser cet outil, les modélisateurs climatiques doivent d'abord choisir explicitement les paramètres de réglage et la plage de valeurs acceptables pour chacun de ces paramètres, puis les références et l'incertitude qui les entoure, les "métriques", qui définissent comment les résultats du modèle sont comparés aux références, et enfin les "tolérances à l'erreur", c'est-à-dire le degré d'accord entre les référence et le modèle qui semble acceptable compte tenu des limites structurelles du modèle.

Ensuite, *htexplo* recherche des optima globaux dans l'espace des paramètres. Cela représente un avantage majeur par rapport aux méthodes de réglage traditionnelles. Lorsque l'accord est effectué à la main, en déplaçant les paramètres les uns après les autres, il y a un risque d'être bloqué dans un optimum local dans l'espace des paramètres. Au contraire, comme *htexplo* explore un échantillon beaucoup plus large de l'espace des paramètres, il aide les modélisateurs à comprendre l'ensemble des limites structurelles d'un MCG donné. Il permet ainsi de rendre le processus de *tuning* à la fois plus robuste et plus standardisé. En rendant le processus de *tuning* plus rapide, il facilite également le développement et le test de nouvelles paramétrisations.

Chapitre 5 – Développement de paramétrisation et *tuning* en pratique : Recouvrement des nuages et des précipitations dans LMDZ

J'étudie dans ce chapitre comment les différentes questions épistémologiques concernant les paramétrisations et le *tuning* apparaissent en pratique en implémentant une paramétrisation de recouvrement des nuages et des précipitations dans le modèle atmosphérique LMDZ. En particulier, je m'intéresse à un problème méthodologique qui se pose systématiquement lors du développement d'une nouvelle paramétrisation : la paramétrisation améliore-t-elle le MCG ou non ? Il semble nécessaire de répondre à cette question pour déterminer si une certaine paramétrisation doit être ajoutée ou non à un MCG donné.

Cette question nécessite de définir le sens du mot amélioration, ou progrès pour un MCG. Pour cela, je pars de la définition du progrès scientifique donnée par Kuhn dans son livre *La structure des révolutions scientifiques* : selon le philosophe et historien des sciences, le progrès de la science se produit non seulement pas une accumulation de connaissances, mais aussi par le remplacement de certaines idées par d'autres. Lorsque deux idées sont en concurrence, leur capacité à résoudre certains *mystères* est cruciale pour déterminer quelle idée sera retenue. Une nouvelle idée sera retenue si elle préserve une grande partie de la capacité de celle qu'elle rem-

place à résoudre des problèmes, mais aborde également des problèmes importants et jusqu'alors inexpliqués (Kuhn, 1962).

Si nous essayons d'appliquer la théorie du progrès scientifique de Kuhn aux paramétrisations, nous pourrions dire qu'une nouvelle paramétrisation améliore un modèle climatique si les résultats entre la nouvelle version et la version originale sont comparables, sauf pour une ou plusieurs variables cibles pour lesquelles la nouvelle version montre des améliorations significatives. Néanmoins, cette définition du progrès scientifique pour les paramétrisations semble trop simpliste. Tout d'abord, les effets d'une paramétrisation dans un MCG donné dépendent des valeurs des paramètres du MCG, donc de son *tuning*. Sans un réglage fin des paramètres du modèle, l'ajout d'une nouvelle paramétrisation peut conduire à des résultats très irréalistes. De plus, même après ré-ajustement des paramètres, une nouvelle paramétrisation détériore souvent certains résultats du modèle. La critère donné par Kuhn doit donc être assoupli : dans la pratique, il est très rare qu'une nouvelle paramétrisation préserve la plupart des résultats d'un MCG tout en améliorant certains de ses biais. Enfin, les progrès conceptuels apportés par une nouvelle paramétrisation doivent être pris en compte. Il ne suffit pas que les simulations produisent des résultats réalistes – il faut aussi pouvoir les interpréter. Une nouvelle paramétrisation peut ou non faciliter cette interprétation.

Il faut donc donner de nouveaux critères pour préciser la notion de progrès pour les paramétrisations. Pour cela, je distingue deux types de progrès : les progrès conceptuels, et les progrès apportés par une nouvelle paramétrisation sur les résultats du modèle. Je propose quatre critères pour évaluer les progrès conceptuels apportés par une paramétrisation : 1) la **cohérence** – est-ce que la paramétrisation rend le modèle plus cohérent, en corrigeant par exemple certaines erreurs ou unifiant le modèle ? 2) l'**interprétabilité** – est-ce que la paramétrisation est fondée sur des concepts bien compris et qui peuvent être facilement interprétés physiquement ? 3) la **simplicité** – la paramétrisation complexifie-t-elle ou au contraire simplifie-t-elle le modèle ? 4) l'**exhaustivité** – dans quelle mesure la paramétrisation permet-elle de prendre en compte un processus important jusqu'alors ignoré dans le modèle ? Ces critères, appelés critères CISC (pour *Coherence, Interpretability, Simplicity, Comprehensiveness* en anglais) donnent un cadre pour évaluer les progrès conceptuels apportés par une paramétrisation dans un modèle de climat donné.

Un cadre doit également être défini pour les progrès sur les résultats du modèle. Pour cela, il faut tout d'abord préciser **les conditions dans lesquelles les simulations sont réalisées** : des progrès observés dans des simulations 1D ne se traduisent en effet pas nécessairement en progrès dans des simulations atmosphériques ou couplées. Ensuite, il faut définir un **protocole de tuning** aussi standardisé que possible de manière à régler à la fois la version standard et la nouvelle version du modèle. Sans un tel protocole, il sera difficile de diagnostiquer si les différences observées sont dues à la paramétrisation elle-même ou à un *tuning* différent. Enfin, il est nécessaire de spécifier comment la comparaison entre la version originale du modèle et la version contenant la nouvelle paramétrisation est effectuée, c'est-à-dire **sur quelles variables cibles** du modèle elle porte, et **quelles sont les références et métriques associées**.

J'applique ces critères pour évaluer les progrès apportés par la paramétrisation de recouvrement entre les nuages et les précipitations de Jakob and Klein (2000) que j'ai implémentée dans LMDZ. Cette dernière repose sur une séparation du flux de précipitations en un flux de précipitations ciel-clair et un flux de précipitations nuageuses. Alors que dans la version

standard du modèle, les précipitations étaient évaporées sur toute la maille comme s'il n'y avait pas de nuages, la nouvelle paramétrisation tient compte du fait qu'une partie du flux de précipitations tombe dans de l'air nuageux saturé, donc ne s'évapore pas. Cela permet de résoudre certaines incohérences présentes dans la version standard du modèle (*critère de cohérence*), remplacer certaines hypothèses arbitraires par des hypothèses plus faciles à interpréter physiquement (*critère d'interprétabilité*), et de prendre en compte un processus important – le fait qu'une partie des précipitations ne s'évapore pas avant la base du nuage – qui n'était pas pris en compte dans la version standard du modèle (*critère d'exhaustivité*). De plus, la paramétrisation n'introduit que quatre nouvelles variables dans le modèle (les flux de précipitations ciel-clair et nuageuses, et les fractions précipitantes associées) et se fonde sur des considérations géométriques assez simples : elle ne complique donc pas beaucoup le modèle LMDZ (*critère de simplicité*). D'après les critères CISC, nous en concluons donc que la nouvelle paramétrisation améliore conceptuellement le modèle.

Pour évaluer les progrès apportés par la nouvelle paramétrisation sur les résultats du modèle, nous devons ré-ajuster les paramètres de la nouvelle version du modèle suite à l'ajout de la nouvelle paramétrisation. Nous utilisons pour cela l'explorateur HIGH-TUNE *htexplo*. Dans un premier temps, nous ajustons les paramètres à partir de trois cas d'étude 1D : le cas ARMCU, qui est un cas de cycle diurne de convection sur continent (Brown et al., 2002), le cas RICO, centré sur la précipitation sous les petits cumulus d'alizés (Raubert et al., 2007), et le cas SANDU qui est un cas de transition entre des stratocumulus et des cumulus au-dessus des océans subtropicaux (Sandu & Stevens, 2011). Des simulations LES sont utilisées comme référence pour chacun des cas. Différents types de paramètres sont choisis pour le *tuning* du modèle : 1) les paramètres **EVAP**, **CLC**, **CLTAU** et **RI** utilisés dans la nouvelle paramétrisation ou dans les équations gouvernant la formation ou l'évaporation des précipitations 2) les paramètres **A1**, **A2**, **B1**, **BG1**, **DZ** impliqués dans la paramétrisation de la convection peu profonde. 3) Les paramètres **FALLV**, **OMEPMX**, **REI**, **RSQPO**, **RQSDP**, **RQSH** utilisés dans la paramétrisation de la convection profonde, qui n'ont pas d'influence sur les cas 1D considérés mais sont gardés pour la phase de *tuning* 3D. Pour chacun de ces paramètres, les valeurs minimale et maximale explorées par *htexplo* sont définies. Les métriques utilisées ciblent principalement la couverture nuageuse, l'altitude des nuages, ainsi que les propriétés thermodynamiques de la couche sous-nuageuse.

Nous avons réalisé 45 vagues de *tuning* 1D, comprenant 90 simulations chacune, pour la version standard et la nouvelle version de LMDZ. Après 45 vagues, les résultats sont comparables entre la nouvelle version et la version standard du modèle pour les métriques choisies. De plus, les meilleures simulations de la nouvelle version du modèle ont des taux de pluie en surface plus réalistes que celles de la version standard, pour des fractions nuageuses comparables. Dans ces trois cas d'étude, le critère donné par Kuhn semble s'appliquer : après *tuning*, la plupart des capacités précédentes du modèle sont conservées, mais la nouvelle paramétrisation permet en plus de réduire certains biais sur les taux de précipitations qui ne pouvaient être corrigés avec du *tuning* dans la version standard du modèle.

Nous comparons ensuite les résultats de la version standard du modèle et de la nouvelle version dans des simulations atmosphériques globales. Nous utilisons pour cela le MCG atmosphérique 3D forcé avec des valeurs standard de température de surface sur océans et avec le cycle saisonnier de la banquise. Les paramètres choisis sont les mêmes que ceux utilisés dans le réglage du modèle 1D et l'espace des paramètres utilisé est celui obtenu après 45 vagues de

réglage 1D. Nous effectuons deux vagues supplémentaires en 3D, de 180 simulations chacune en utilisant les mêmes métriques SCM que celles considérées pour les 45 premières vagues, ainsi que des métriques 3D supplémentaires, ciblées sur les flux radiatifs au sommet de l’atmosphère et les taux de précipitations en surface.

Le protocole de *tuning* utilisé permet au bout des deux vagues de *tuning* 3D d’obtenir des flux radiatifs au sommet de l’atmosphère réalistes avec la nouvelle version du modèle. Cependant, il est difficile d’évaluer si la paramétrisation améliore ou non les résultats du modèle. Même en considérant simplement deux variables – la température de surface sur continents et les précipitations moyennées sur une année – nous observons que la nouvelle version du modèle conduit à certaines améliorations, mais aussi à certaines dégradations. Contrairement à ce qu’il se passe en 1D, le critère de Kuhn ne s’applique donc pas : dans les simulations globales, les capacités antérieures de la version standard du modèle ne sont pas entièrement préservées. Des améliorations dans certaines régions s’accompagnent de détériorations dans d’autres, ce qui rend l’évaluation des simulations des MCG difficile. La comparaison de deux simulations implique donc des choix sur les aspects des résultats des simulations qui sont considérés comme les plus importants.

Pour rendre ces choix explicites, il est possible d’utiliser des métriques. Nous donnons la proposition suivante (**Proposition 1**) : avec un ensemble n de métriques $(p_i)_{i \leq n}$ et de poids associés $(m_i)_{i \leq n}$, nous considérons qu’une simulation S_1 est meilleure qu’une autre S_2 si et seulement si la moyenne pondérée des erreurs normalisées de S_1 selon les métriques définies est plus petite que la moyenne pondérée des erreurs normalisées de S_2 selon les mêmes métriques. Cette proposition donne une méthode quantitative pour comparer deux simulations. Elle ne permet cependant pas de comparer deux versions d’un modèle. En effet, le fait qu’une simulation particulière d’une version d’un modèle V_1 soit meilleure qu’une simulation particulière d’une autre version V_2 de ce modèle ne prouve pas que V_1 est meilleure que V_2 . Certaines simulations d’une même version d’un modèle peuvent en effet conduire à des résultats plus ou moins bons en fonction du *tuning* de leurs paramètres. Pour comparer deux versions d’un modèle indépendamment de leur *tuning*, nous proposons donc la deuxième proposition suivante (**Proposition 2**) : une version V_1 est meilleure qu’une autre version V_2 si et seulement s’il existe un jeu de paramètres $P_{1,0}$ dans l’espace de l’ensemble des paramètres de V_1 , Ω_1 , tel que la simulation de V_1 avec le jeu de paramètres $P_{1,0}$, $S_{1,0} = V_1(P_{1,0})$, est meilleure que toute simulation $S_2 = V_2(P_2)$, avec P_2 appartenant à l’espace de paramètres de V_2 , Ω_2 .

En pratique, cette proposition est difficile à utiliser parce qu’elle nécessite un protocole de *tuning* qui explore de façon exhaustive l’espace des paramètres. A cause du nombre de paramètres que contient un MCG et de la gamme de valeurs possibles pour chaque paramètre, cela est impossible en pratique. Le choix des paramètres explorés et de leur valeur sera toujours en partie arbitraire. De même, les métriques et les poids donnés à chaque métrique sont fixées en partie arbitrairement. Il n’y a pas encore de consensus sur les métriques qui sont les plus pertinentes pour évaluer la performance d’un modèle de climat. Il faut donc garder à l’esprit que le résultat de la comparaison dépend des métriques choisies et de leurs poids respectifs.

Nous nous inspirons cependant de la proposition 2 pour comparer la nouvelle version de LMDZ à la version standard, en identifiant certaines différences entre les deux versions qui semblent indépendantes du *tuning*. Nous identifions notamment deux biais dans la version standard du modèle, résistants au *tuning*, qui sont largement atténués dans la nouvelle version :

les pluies excessives au-dessus de l'Indonésie et sur la côte nord-est de l'Amérique du Sud, et le biais chaud estival des températures de surface sur les continents des latitudes moyennes. Ces résultats suggèrent que la nouvelle paramétrisation implémentée dans LMDZ conduit à des améliorations non seulement dans le modèle 1D, mais aussi dans le modèle 3D et illustrent le potentiel des méthodes de *tuning* automatique comme *htexplo* pour guider le développement et le test de nouvelles paramétrisations dans les MCG.

Chapitre 6 – Le projet d'intercomparaison de modèles couplés CMIP : Histoire, utilisations, et effets structurels sur les recherches en sciences du climat

Le principe des projets d'intercomparaison de modèles est de comparer des simulations MCG réalisées dans les mêmes conditions. Le projet d'intercomparaison de modèles atmosphériques AMIP (pour *Atmospheric Model Intercomparison Project* en anglais), qui a débuté en 1990, a été la première tentative de coordination d'une telle activité. Différents projets d'intercomparaison ont suivi : en particulier, le projet d'intercomparaison de modèles couplés CMIP (pour *Coupled Model Intercomparison Project* en anglais) est maintenant considéré comme "l'un des éléments fondateurs des sciences du climat" (Eyring et al., 2016, p. 1937). Depuis sa création au milieu des années 1990, il a évolué en cinq phases, auxquelles ont participé les principaux groupes de modélisation du climat au monde. En plus de leur rôle pour les recherches sur le climat, ces phases ont occupé une place centrale dans les rapports du GIEC, dont l'objectif est de synthétiser l'état actuel des connaissances sur le climat et les changements climatiques. Les ensembles multi-modèles de CMIP permettent en effet de résumer les prédictions des modèles de climat les plus réalistes, les MCG couplés, et ainsi de fournir des déclarations pour les rapports du GIEC.

Si CMIP a joué un rôle incontestable pour les rapports du GIEC, son rôle pour la recherche climatique est plus controversée. En analysant un ensemble de 280 papiers fondés sur les résultats de CMIP publiés entre 2012 et 2018 dans six journaux renommés en sciences du climat, je distingue trois utilisations principales des résultats de CMIP : l'exploration des changements climatiques futurs et des incertitudes associés, la comparaison des simulations de CMIP et des observations et l'interprétation des résultats des modèles. Je montre que chacune de ces utilisations pose des problèmes épistémologiques majeurs.

Lorsque les ensembles multi-modèles de CMIP sont utilisés pour **explorer les changements climatiques futurs**, l'accord, ou le désaccord, entre les modèles est souvent utilisé pour estimer l'incertitude sur différentes prédictions climatiques. Pourtant, comme la participation à CMIP se fait sur la base du volontariat, les modèles participant à CMIP n'ont pas été choisis pour explorer de façon systématique l'incertitude des prédictions climatiques futures. L'écart entre les modèles ne peut donc être vu que comme une borne inférieure de l'incertitude. Par conséquent, si les modèles de CMIP sont d'accord sur un certain résultat – par exemple, s'ils prédisent tous une hausse de température similaire d'ici la fin du XXIème siècle selon un certain scénario d'émission – cela ne suffit pas à prouver que ce résultat est entouré de peu d'incertitudes, contrairement à ce qui est suggéré par certaines études (ex : Jayasankar et al., 2015).

Une autre utilisation courante des résultats de CMIP consiste à les **comparer avec des observations** afin d'évaluer la performance des MCG correspondants. Cependant, le *tuning* complique l'interprétation que l'on peut donner à l'accord entre les MCG et les observations : l'accord avec les observations peut en effet être amélioré en modifiant la valeur des paramètres

du modèle de façon ad hoc. Il est donc possible d'obtenir de bons résultats pour de mauvaises raisons : l'accord avec les observations peut résulter d'erreurs qui se compensent et ne prouve pas nécessairement que les processus sont bien représentés dans le modèle. Ce risque est particulièrement fort si les mêmes ensembles de données sont utilisées pour régler les paramètres d'un modèle et évaluer ses performances (Frisch, 2019; Tebaldi & Knutti, 2007).

Un autre problème concerne les observations elles-mêmes. La majorité des études que nous avons consultées ne comparent pas les simulations de CMIP avec des observations indépendantes du modèle, mais avec des réanalyses. Or, les modèles utilisés dans les réanalyses se fondent sur des méthodes numériques, des hypothèses et des paramétrisations similaires à celles des modèles de CMIP et ont donc des biais en commun avec les MCG (Edwards, 2010). Même si l'analyse est continuellement corrigée avec les données d'observation disponibles, les modèles de réanalyse transmettent une partie de leurs biais aux produits de réanalyse. Par conséquent, un accord entre les modèles climatiques et les données de réanalyse illustre dans certains cas simplement des biais communs entre les MCG et la réanalyse utilisée comme référence. Le problème est particulièrement critique pour les phénomènes pour lesquels nous disposons de peu d'observations, le cycle de l'eau notamment. La réanalyse est alors faiblement contrainte par les observations, donc principalement dérivée du modèle et peu fiable.

La dernière utilisation de CMIP consiste à **interpréter les résultats des modèles**, mais dans un sens très spécifique. Les motivations originelles d'AMIP, un des premiers projets d'intercomparaison de modèles, étaient de "diagnostiquer et interpréter en profondeur" les résultats des modèles (Gates, 1992). L'idée derrière AMIP était de comprendre les causes des succès et des échecs des MCG atmosphériques, c'est-à-dire de les rattacher à certaines hypothèses ou formulations utilisées dans les paramétrisations de ces modèles. Cette vocation est en contraste avec l'utilisation qui est faite de CMIP aujourd'hui : la plupart des articles que nous avons consultés s'intéressent en effet uniquement aux sorties des modèles. Le lien avec les paramétrisations sous-jacentes n'est quasiment jamais effectué.

Selon Lenhard and Winsberg (2010), cela manifeste une certaine forme d'*holisme de confirmation* caractéristique des MCG. Ce concept est traditionnellement défendu en philosophie des sciences comme l'idée selon laquelle une certaine hypothèse ne peut jamais être testée de façon isolée, mais que tout test expérimental dépend inévitablement d'autres théories ou hypothèses. Comme l'expliquent Lenhard and Winsberg (2010), les modèles climatiques sont confrontés à une forme particulièrement forte d'holisme de confirmation car il y a à la fois une grande modularité dans leur développement et beaucoup d'interactions entre leurs différents modules, i.e. entre les paramétrisations. Selon Lenhard and Winsberg (2010), les interactions entre les paramétrisations sont si fortes qu'il n'est pas possible de démêler leur rôle respectif dans les succès et les échecs des résultats des modèles.

Je m'oppose à cette vision en expliquant qu'un modèle climatique n'est pas seulement la somme de son code (et des hypothèses qui le sous-tendent) et des résultats qu'il donne pour un ensemble particulier de simulations. Au contraire, il s'agit d'une entité dynamique avec laquelle il est possible d'interagir. Il est possible d'initialiser un modèle avec différentes conditions, de modifier ses paramètres, d'utiliser des versions simplifiées ou idéalisées du modèle, etc. Grâce à ces différentes interactions, les modélisateurs peuvent acquérir une connaissance du comportement d'un modèle climatique, de ce qu'il fait et pourquoi. Dans la littérature, il existe ainsi de nombreuses études qui font le lien entre les résultats d'un modèle et ses paramétrisations (e.g.

Hourdin et al., 2013; Notz et al., 2013). Mais la plupart de ces études sont faites pour un seul modèle – quand de nombreux modèles sont étudiés, comme dans les ensembles multi-modèles du CMIP, les détails des paramétrisations impliquées ne sont presque jamais pris en compte. J’en conclus que ce ne sont pas les MCG qui font face à une forme forte de holisme de confirmation, mais CMIP. Les études fondées sur les résultats multi-modèles de CMIP considèrent en effet les MCG comme des boîtes noires, et ne s’intéressent qu’aux résultats des modèles et non à leur contenu.

J’explique enfin que **CMIP a eu des effets structurels sur les sciences du climat**, qui font de lui une véritable “infrastructure” pour la recherche climatique (Pitt, 2000). Tout d’abord, CMIP a focalisé l’attention des scientifiques sur les MCG. Chaque phase de CMIP est apparue comme un rendez-vous international pour les principaux groupes de modélisation mondiaux. Pour avoir voix au chapitre, chaque groupe de modélisation devait présenter la version la plus récente de son MCG. Cela a créé une pression autour du développement des MCG, au détriment des autres outils qui peuvent être utilisés pour étudier le climat.

La façon de travailler avec des MCG a également été transformée. Avant AMIP et CMIP, les différents groupes de modélisation travaillaient de manière relativement isolée : chaque MCG présente en effet ses particularités, son langage qui lui est propre, et l’expertise acquise sur un MCG donné n’est pas facilement transférable à un autre modèle. Comme l’a écrit Edwards (2010), les projets d’intercomparaison de modèles ont lié ces systèmes isolés en un réseau. Ils ont permis non seulement de rendre les MCG comparables en les utilisant dans des conditions similaires, mais les ont également aidés à partager certaines références communes et à utiliser des approches de modélisation ou d’analyse similaires.

Le libre accès aux données de CMIP a par ailleurs considérablement facilité les analyses des résultats des MCG. Cela a créé un fossé grandissant entre les développeurs de modèles et leur utilisateurs. Avant AMIP et CMIP, les résultats des simulations des MCG étaient généralement analysés par les quelques personnes qui avaient participé au développement du MCG correspondant (ex : Manabe & Wetherald, 1975; Washington et al., 1980). Les utilisateurs des modèles avaient donc une vision critique des forces et des faiblesses du modèle climatique qu’ils analysaient, car ils avaient contribué au développement de certains de ses aspects. Lorsque les données des projets d’intercomparaison ont été rendues librement accessibles, les MCG ont commencé à être analysés par des personnes qui n’avaient pas participé à leur développement. Il en a résulté une perte de compréhension des résultats des modèles climatiques, et une tendance accrue à utiliser les MCG uniquement comme des boîtes noires.

Chapitre 7 – Conclusion

En conclusion, je commence par résumer la démarche que j’ai adoptée dans ma thèse et les principales questions que j’ai abordées tout au long des chapitres. Je fais ensuite le lien entre ma thèse et des questions plus générales sur l’avenir des modèles de climat. Je m’intéresse d’abord à l’avenir de CMIP. Face aux peu de progrès des GCMs d’une phase de CMIP à l’autre, je propose d’abandonner le concept de “phases de CMIP”. Je suggère que les modélisateurs puissent soumettre une nouvelle version de leurs modèles aux simulations CMIP quand ils le souhaitent, c’est-à-dire chaque fois qu’ils estiment qu’une nouvelle version est suffisamment différente de la précédente et suffisamment mature pour justifier une nouvelle série de simulations. Cela créerait moins de pression sur les développeurs de modèles et leur permettrait de développer leurs modèles en fonction de leurs propres échéances et de leurs priorités, indépendamment du

calendrier imposé par le CMIP. Cela permettrait également d'éviter la "course à la publication" après chaque phase de CMIP.

Je me penche ensuite sur le projet DestinE (*Destination Earth*) qui a pour objectif de construire des "jumeaux numériques de la Terre" (Bauer et al., 2021). Les auteurs du projet expliquent que de tels jumeaux numériques de la Terre, contraints par les observations et les lois de la physique, pourraient être utilisés pour guider les décideurs politiques et les entreprises pour une variété d'applications et répondre à des questions comme : "Quel est l'impact économique de l'ajout de centrales éoliennes dans une région quand l'Europe sera neutre en carbone ? Quelle politique agricole est la plus durable compte tenu des changements de régimes météorologiques à venir ?" (Bauer et al., 2021, p. 81-82). Pour répondre à ces ambitions, DestinE a pour vocation de donner à n'importe quel décideur la possibilité d'utiliser les jumeaux numériques à ses propres fins. L'interprétation des résultats des jumeaux numériques ne serait donc pas confinée à une "couche d'experts". Avec cette ambition, les modèles sont vus comme des outils prédictifs dont les résultats ont une valeur en soi.

Au contraire, on pourrait soutenir que la valeur d'un modèle provient davantage de l'interprétation que l'on donne à ses résultats que des résultats eux-mêmes : dans ce cas, les modèles sont considérés comme des guides pour formuler et tester des hypothèses plutôt que comme la vérité. Une telle utilisation des modèles nécessite une bonne compréhension de leurs limites : une certaine expertise sur le fonctionnement et les limites de ces modèles semble donc nécessaire pour interpréter correctement leurs résultats. Permettre à n'importe qui d'utiliser un modèle comme une "boîte noire" – comme le propose DestinE avec ses jumeaux numériques – semblerait alors peu souhaitable, voire dangereux. En plus du débat technique sur la possibilité ou non de créer des jumeaux numériques de la Terre, il y a donc un débat épistémologique derrière DestinE : quels types de modèles voulons-nous à l'avenir et comment voulons-nous les utiliser ?

Cette question ne concerne pas seulement DestinE. La modélisation de la convection atmosphérique - et plus généralement du climat - est aujourd'hui à un tournant. Les paramétrisations de la convection atmosphérique, qui ont constitué le paradigme dominant sur lequel les MCG ont été construits pendant plus de 50 ans, sont remises en question. En particulier, certains scientifiques proposent de remplacer les paramétrisations par des simulations à haute résolution – sous la forme de MCG ou de super-paramétrisations – ou par du *machine learning*. Ces différentes options s'opposent pour des raisons techniques, mais aussi pour des raisons épistémologiques et reflètent des conceptions différentes sur le rôle des modèles. Clarifier le rôle et les limites des MCG actuels a été une motivation importante de ma thèse de doctorat. Pour poursuivre cette réflexion, je propose dans la conclusion trois conseils épistémologiques à l'intention des modélisateurs de climat.

Le premier est simplement que **tous les modèles ne sont pas équivalents**. J'insiste sur la nécessité de reconnaître la diversité des modèles a fortiori lorsque des modèles climatiques sont couplés à des modèles socio-économiques comme DestinE cherche à le faire. J'explique notamment que les modèles utilisés pour simuler la circulation générale de l'atmosphère et des océans présentent une particularité : ils sont fondés sur les équations fondamentales de la thermodynamique et de la mécanique des fluides. Ces équations sont un outils très précieux pour les physiciens de l'atmosphère et du climat, mais malheureusement, il n'existe pas d'équations similaires pour décrire les interactions sociales et économiques. Il existe donc un fossé épistémologique entre les modèles qui servent à simuler les flux atmosphériques et océaniques et ceux qui intègrent

les activités socio-économique. Les méthodes qui ont fait leurs preuves pour les premiers ne seront donc pas nécessairement aussi efficaces pour les seconds.

Le deuxième conseil est qu'il est important **d'évaluer et d'accepter les limites des modèles**. Tous les modèles ont une certaine utilité, mais aussi certaines limites. Ces limites doivent être soigneusement évaluées afin de ne pas donner de mauvaises interprétations des résultats du modèle. Il existe de nombreux exemples d'interprétations erronées de modèles dans la littérature, avec parfois des conséquences politiques directes. Pour éviter une confiance excessive dans un modèle, ses limites doivent être prises en compte dans l'interprétation de ses résultats. Il nous semble important d'insister sur ce point car, comme l'illustrent les ambitions de DestinE, les applications des modèles sont de plus en plus nombreuses. Cela peut donner l'impression trompeuse que les modèles peuvent fournir des réponses quantitatives à n'importe quel problème. Au contraire, je soutiens qu'un raisonnement qualitatif fondé sur une analyse approfondie de certaines hypothèses clés semble dans de nombreuses situations préférable aux résultats quantitatifs donnés par un modèle.

Enfin, le troisième conseil est qu'il **ne faut pas laisser les modèles devenir des boîtes noires**. En effet, sans une compréhension du contenu d'un modèle, ses utilisateurs ne peuvent pas évaluer les hypothèses sur lesquelles il se fonde. Or, certaines hypothèses justifiées pour une certaine utilisation peuvent sembler absurdes pour d'autres utilisations. Comprendre le contenu d'un modèle permet donc de déterminer à la fois les utilisations légitimes d'un modèle et ses limites. Face à l'utilisation de modèles de plus en plus complexes difficiles à interpréter, je conclus que des modèles idéalisés ont encore toute leur pertinence pour comprendre le climat.

References

- Adler, R. F., Sapiiano, M. R., Huffman, G. J., Wang, J.-J., Gu, G., Bolvin, D., . . . others (2018). The global precipitation climatology project (gpcp) monthly analysis (new version 2.3) and a review of 2017 global precipitation. *Atmosphere*, 9(4), 138.
Cited on pages 136, 139, and 143.
- Ahn, M.-S., Kim, D., Kang, D., Lee, J., Sperber, K. R., Gleckler, P. J., . . . Kim, H. (2020). Mjo propagation across the maritime continent: Are cmip6 models better than cmip5 models? *Geophysical Research Letters*, e2020GL087250.
Cited on pages 104 and 223.
- Albright, A. L., Fildier, B., & Touzé-Peiffer, L. (2020). Atmospheric radiative profiles during eurec4a. *Earth System Science Data*. doi: 10.25326/78
Cited on pages 29, 165, 196, 211, and 214.
- Allan, R. P., & Soden, B. J. (2007). Large discrepancy between observed and simulated precipitation trends in the ascending and descending branches of the tropical circulation. *Geophysical Research Letters*, 34(18).
Cited on page 15.
- Arakawa, A. (1969). Parameterization of cumulus convection. In *Proceedings of wmo/iugg symposium, numerical weather prediction, japan meteorological agency* (Vol. 8, pp. 1–6).
Cited on pages 72, 73, 74, 75, 76, 77, 83, and 218.
- Arakawa, A. (2004). The cumulus parameterization problem: Past, present, and future. *Journal of Climate*, 17(13), 2493–2525.
Cited on pages 22 and 73.
- Arakawa, A., & Jung, J.-H. (2011). Multiscale modeling of the moist-convective atmosphere—a review. *Atmospheric research*, 102(3), 263–285.
Cited on page 71.
- Arakawa, A., & Schubert, W. H. (1974). Interaction of a cumulus cloud ensemble with the large-scale environment, part i. *Journal of the Atmospheric Sciences*, 31(3), 674–701.
Cited on pages 18, 38, 46, 65, 70, 73, 75, 76, 79, 84, 85, 90, 213, and 219.
- Balaji, V. (2021). Climbing down charney’s ladder: machine learning and the post-dennard era of computational climate science. *Philosophical Transactions of the Royal Society A*, 379(2194), 20200085.
Cited on page 22.
- Bauer, P., Stevens, B., & Hazeleger, W. (2021). A digital twin of earth for the green transition. *Nature*

Climate Change, 11(2), 80–83.

Cited on pages 169, 170, 171, and 233.

Baumhefner, D., & Downey, P. (1978). Forecast intercomparisons from three numerical weather prediction models. *Monthly Weather Review*, 106(9), 1245–1279.

Cited on page 149.

Beadling, R., Russell, J., Stouffer, R., Mazloff, M., Talley, L., Goodman, P., ... Pandde, A. (2020). Representation of southern ocean properties across coupled model intercomparison project generations: Cmp3 to cmp6. *Journal of Climate*, 33(15), 6555–6581.

Cited on page 168.

Bechtel, W., & Abrahamsen, A. (2005). Explanation: A mechanist alternative. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 36(2), 421–441.

Cited on page 36.

Bechtold, P., Krueger, S., Lewellen, W., Van Meijgaard, E., Moeng, C.-H., Randall, D., ... Wang, S. (1996). Modeling a stratocumulus-topped pbl: Intercomparison among different one-dimensional codes and with large eddy simulation. *Bulletin of the American Meteorological Society*, 77(9), 2033–2042.

Cited on page 20.

Bechtold, P., Semane, N., Lopez, P., Chaboureaud, J.-P., Beljaars, A., & Bormann, N. (2014). Representing equilibrium and nonequilibrium convection in large-scale models. *Journal of Atmospheric Sciences*, 71(2), 734–753.

Cited on page 71.

Benjamin, T. B. (1968). Gravity currents and related phenomena. *Journal of Fluid Mechanics*, 31(2), 209–248.

Cited on page 53.

Berthelot, J.-M. (2018). *Épistémologie des sciences sociales*. Presses universitaires de France.

Cited on page 20.

Betts, A. (1973). Non-precipitating cumulus convection and its parameterization. *Quarterly Journal of the Royal Meteorological Society*, 99(419), 178–196.

Cited on page 70.

Betts, A. K. (1978). Convection in the tropics. *Meteorology over the Tropical Oceans*, 105–132.

Cited on page 41.

Betts, A. K. (1986). A new convective adjustment scheme. part i: Observational and theoretical basis. *Quarterly Journal of the Royal Meteorological Society*, 112(473), 677–691.

Cited on pages 38, 70, 71, and 74.

Bock, L., Lauer, A., Schlund, M., Barreiro, M., Bellouin, N., Jones, C., ... Eyring, V. (2020). Quantifying progress across different cmp phases with the esmvaltool. *Journal of Geophysical Research: Atmospheres*, 125(21), e2019JD032321.

Cited on page 167.

- Bogenschutz, P. A., & Krueger, S. K. (2013). A simplified pdf parameterization of subgrid-scale clouds and turbulence for cloud-resolving models. *Journal of Advances in Modeling Earth Systems*, 5(2), 195–211.
Cited on page 18.
- Bony, S., & Dufresne, J.-L. (2005). Marine boundary layer clouds at the heart of tropical cloud feedback uncertainties in climate models. *Geophysical Research Letters*, 32(20). doi: 10.1029/2005GL023851
Cited on page 196.
- Bony, S., Schulz, H., Vial, J., & Stevens, B. (2020a). Sugar, gravel, fish, and flowers: Dependence of mesoscale patterns of trade-wind clouds on environmental conditions. *Geophysical research letters*, 47(7), e2019GL085988.
Cited on pages 38 and 39.
- Bony, S., Schulz, H., Vial, J., & Stevens, B. (2020b). Sugar, gravel, fish, and flowers: Dependence of mesoscale patterns of trade-wind clouds on environmental conditions. *Geophysical Research Letters*, 47(7), e2019GL085988. (e2019GL085988 10.1029/2019GL085988) doi: 10.1029/2019GL085988
Cited on pages 197 and 205.
- Bony, S., & Stevens, B. (2019). Measuring area-averaged vertical motions with dropsondes. *Journal of the Atmospheric Sciences*, 76(3), 767–783. doi: 10.1175/JAS-D-18-0141.1
Cited on pages 186 and 187.
- Bony, S., Stevens, B., Ament, F., Bigorre, S., Chazette, P., Crewell, S., ... Wirth, M. (2017). Eurec4a: A field campaign to elucidate the couplings between clouds, convection and circulation. *Surveys in Geophysics*, 38(6), 1529–1568. doi: 10.1007/s10712-017-9428-0
Cited on pages 16, 59, 177, 187, 196, and 198.
- Bony, S., Stevens, B., Held, I., Mitchell, J., Dufresne, J.-L., Emmanuel, K., ... Senior, C. (2013). Carbon dioxide and climate : Perspectives on a scientific assessment. In G. Asrar & J. Hurrell (Eds.), *Climate science for serving society, research, modeling and prediction priorities* (pp. 391–414). Springer.
Cited on pages 21, 51, 160, and 163.
- Boucher, O., Randall, D., Artaxo, P., Bretherton, C., Feingold, G., Forster, P., ... others (2013). Clouds and aerosols. In *Climate change 2013: the physical science basis. contribution of working group i to the fifth assessment report of the intergovernmental panel on climate change* (pp. 571–657). Cambridge University Press.
Cited on page 14.
- Boucher, O., Servonnat, J., Albright, A. L., Aumont, O., Balkanski, Y., Bastrikov, V., ... others (2020). Presentation and evaluation of the ipsl-cm6a-lr climate model. *Journal of Advances in Modeling Earth Systems*, 12(7), e2019MS002010.
Cited on page 143.
- Box, G. E. (1976). Science and statistics. *Journal of the American Statistical Association*, 71(356), 791–799.
Cited on page 76.

- Brenowitz, N. D., Beucler, T., Pritchard, M., & Bretherton, C. S. (2020). Interpreting and stabilizing machine-learning parametrizations of convection. *Journal of the Atmospheric Sciences*, 77(12), 4357–4375.
Cited on page 23.
- Bretherton, C., & Blossey, P. (2017). Understanding mesoscale aggregation of shallow cumulus convection using large-eddy simulation. *Journal of Advances in Modeling Earth Systems*, 9(8), 2798–2821. doi: 10.1002/2017MS000981
Cited on pages 177 and 187.
- Bretherton, C. S., & Blossey, P. N. (2017). Understanding mesoscale aggregation of shallow cumulus convection using large-eddy simulation. *Journal of Advances in Modeling Earth Systems*, 9(8), 2798–2821. doi: 10.1002/2017MS000981
Cited on page 211.
- Bretherton, C. S., Blossey, P. N., Khairoutdinov, M., Bretherton, C. S., Blossey, P. N., & Khairoutdinov, M. (2005, dec). An Energy-Balance Analysis of Deep Convective Self-Aggregation above Uniform SST. *Journal of the Atmospheric Sciences*, 62(12), 4273–4292. doi: 10.1175/JAS3614.1
Cited on page 205.
- Bretherton, C. S., Krueger, S. K., Wyant, M. C., Bechtold, P., Van Meijgaard, E., Stevens, B., & Teixeira, J. (1999). A gcss boundary-layer cloud model intercomparison study of the first astex lagrangian experiment. *Boundary-Layer Meteorology*, 93(3), 341–380.
Cited on page 20.
- Brient, F., Couvreux, F., Villefranque, N., Rio, C., & Honnert, R. (2019). Object-oriented identification of coherent structures in large eddy simulations: Importance of downdrafts in stratocumulus. *Geophysical Research Letters*, 46(5), 2854–2864.
Cited on pages 32, 45, and 81.
- Brown, A., Cederwall, R., Chlond, A., Duynkerke, P., Golaz, J.-C., Khairoutdinov, M., ... others (2002). Large-eddy simulation of the diurnal cycle of shallow cumulus convection over land. *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography*, 128(582), 1075–1093.
Cited on pages 126 and 228.
- Browning, K. (1994). Survey of perceived priority issues in the parametrizations of cloud-related processes in gcms. *Quarterly Journal of the Royal Meteorological Society*, 120(516), 483–487.
Cited on page 19.
- Browning, K., Betts, A., Jonas, P., Kershaw, R., Manton, M., & Mason, P. (1993). The gewex cloud system study (gcss). *Bulletin of the American Meteorological Society*, 74(3), 387–400.
Cited on page 100.
- Bryan, G. H., & Parker, M. D. (2010). Observations of a squall line and its near environment using high-frequency rawinsonde launches during VORTEX2. *Monthly weather review*, 138(11), 4076–4097. doi: <https://doi.org/10.1175/2010MWR3359.1>
Cited on page 176.
- Byers, H. R., & Braham, R. R. (1949). *The thunderstorm: report of the thunderstorm project*. US

Government Printing Office.

Cited on page 52.

Canut, G., Couvreur, F., Lothon, M., Pino, D., & Saïd, F. (2012). Observations and large-eddy simulations of entrainment in the sheared Sahelian boundary layer. *Boundary-layer meteorology*, 142(1), 79–101. doi: <https://doi.org/10.1007/s10546-011-9661-x>

Cited on pages 59 and 178.

Cao, G., Giambelluca, T. W., Stevens, D. E., & Schroeder, T. A. (2007). Inversion variability in the hawaiian trade wind regime. *Journal of Climate*, 20(7), 1145–1160. doi: 10.1175/JCLI4033.1

Cited on page 202.

Charney, J. (1948). The dynamics of long waves in a baroclinic westerly current.” *j. meteor.*, 4, 135–162 (1947). *On the Scale of Atmospheric Motions.*” *Geofys. Publikasjoner*, 17, 17.

Cited on page 75.

Charney, J., & Ogura, Y. (1960). A numerical model for thermal convection in the atmosphere. *Journal of the Meteorological Society of Japan. Ser. II*, 38(6), 19a–19a.

Cited on page 68.

Charney, J. G., Arakawa, A., Baker, D. J., Bolin, B., Dickinson, R. E., Goody, R. M., ... Wunsch, C. I. (1979). *Carbon dioxide and climate: a scientific assessment*. National Academy of Sciences, Washington, DC.

Cited on page 168.

Chazette, P., Marnas, F., Totems, J., & Shang, X. (2014, sep). Comparison of IASI water vapor retrieval with H2O-Raman lidar in the framework of the Mediterranean HyMeX and ChArMEx programs. *Atmospheric Chemistry and Physics*, 14(18), 9583–9596. doi: 10.5194/acp-14-9583-2014

Cited on page 197.

Chen, L., & Frauenfeld, O. W. (2014). Surface air temperature changes over the twentieth and twenty-first centuries in china simulated by 20 cmip5 models. *Journal of Climate*, 27(11), 3920–3937.

Cited on page 25.

Cheruy, F., Ducharne, A., Hourdin, F., Musat, I., Vignon, É., Gastineau, G., ... others (2020). Improved near-surface continental climate in ipsl-cm6a-lr by combined evolutions of atmospheric and land surface physics. *Journal of Advances in Modeling Earth Systems*, 12(10), e2019MS002005.

Cited on page 143.

Chien, S.-Y., Ma, S.-Y., & Chen, L.-G. (2002). Efficient moving object segmentation algorithm using background registration technique. *IEEE Transactions on Circuits and Systems for Video Technology*, 12(7), 577–586.

Cited on page 32.

Chlond, A., & Wolkau, A. (2000). Large-eddy simulation of a nocturnal stratocumulus-topped marine atmospheric boundary layer: An uncertainty analysis. *Boundary-Layer Meteorology*, 95(1), 31–55.

Cited on page 127.

- Chou, C., Neelin, J. D., Chen, C.-A., & Tu, J.-Y. (2009). Evaluating the “rich-get-richer” mechanism in tropical precipitation change under global warming. *Journal of Climate*, 22(8), 1982–2005. Cited on page 15.
- Claussen, M., Mysak, L., Weaver, A., Crucifix, M., Fichefet, T., Loutre, M.-F., ... others (2002). Earth system models of intermediate complexity: closing the gap in the spectrum of climate system models. *Climate dynamics*, 18(7), 579–586. Cited on page 49.
- Comte, A. (1835). *Cours de philosophie positive: première leçon* (Vol. 1). Bachelier. Cited on pages 57 and 216.
- Couvreux, F., Hourdin, F., & Rio, C. (2010). Resolved versus parametrized boundary-layer plumes. part i: A parametrization-oriented conditional sampling in large-eddy simulations. *Boundary-layer meteorology*, 134(3), 441–458. Cited on pages 32, 97, and 98.
- Couvreux, F., Hourdin, F., Williamson, D., Roehrig, R., Volodina, V., Villefranche, N., ... others (2021). Process-based climate model development harnessing machine learning: I. a calibration tool for parameterization improvement. *Journal of Advances in Modeling Earth Systems*, 13(3), e2020MS002217. Cited on pages 23, 109, 129, 130, and 226.
- Craig, G. C., & Cohen, B. G. (2006). Fluctuations in an equilibrium convective ensemble. part i: Theoretical formulation. *Journal of the atmospheric sciences*, 63(8), 1996–2004. Cited on pages 65, 77, 78, 79, 87, and 219.
- Craig Goff, R. (1976). Vertical structure of thunderstorm outflows. *Monthly Weather Review*, 104(11), 1429–1440. Cited on pages 52, 176, and 215.
- Craver, C. F. (2007). *Explaining the brain: Mechanisms and the mosaic unity of neuroscience*. Oxford University Press. Cited on page 36.
- Craver, C. F. (2013). Functions and mechanisms: A perspectivalist view. In *Functions: Selection and mechanisms* (pp. 133–158). Springer. Cited on page 36.
- Cubasch, U., & et al. (2001). *Climate change 2001: The scientific basis* (J. Houghton et al., Eds.). Cambridge University Press. Cited on page 152.
- Cuijpers, J., & Duynkerke, P. (1993). Large eddy simulation of trade wind cumulus clouds. *Journal of the Atmospheric Sciences*, 50(23), 3894–3908. doi: [https://doi.org/10.1175/1520-0469\(1993\)050<3894:LESOTW>2.0.CO;2](https://doi.org/10.1175/1520-0469(1993)050<3894:LESOTW>2.0.CO;2) Cited on pages 59 and 178.
- Dalmedico, A. D. (2001). History and epistemology of models: Meteorology (1946–1963) as a case study. *Archive for history of exact sciences*, 55(5), 395–422. Cited on page 25.

- Danabasoglu, G., Yeager, S. G., Bailey, D., Behrens, E., Bentsen, M., Bi, D., ... others (2014). North atlantic simulations in coordinated ocean-ice reference experiments phase ii (core-ii). part i: mean states. *Ocean Modelling*, 73, 76–107.
Cited on page 100.
- Darden, L. (2006). *Reasoning in biological discoveries: Essays on mechanisms, interfield relations, and anomaly resolution*. Cambridge University Press.
Cited on page 36.
- Davidson, B. (1968). The barbados oceanographic and meteorological experiment. *Bulletin of the American Meteorological Society*, 49(9), 928–935.
Cited on page 40.
- Davis, C., Brown, B., & Bullock, R. (2006). Object-based verification of precipitation forecasts. part i: Methodology and application to mesoscale rain areas. *Monthly Weather Review*, 134(7), 1772–1784.
Cited on page 32.
- Deardorff, J. W. (1972). Parameterization of the planetary boundary layer for use in general circulation models. *Monthly Weather Review*, 100(2), 93–106.
Cited on page 46.
- Dee, D. P., Uppala, S. M., Simmons, A., Berrisford, P., Poli, P., Kobayashi, S., ... others (2011). The era-interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the royal meteorological society*, 137(656), 553–597.
Cited on pages 139 and 144.
- DelGenio, A. D., & Yao, M.-S. (1993). Efficient cumulus parameterization for long-term climate studies: The giss scheme. In *The representation of cumulus convection in numerical models* (pp. 181–184). Springer.
Cited on pages 18 and 213.
- de Szoeke, S. P., Skillingstad, E. D., Zuidema, P., & Chandra, A. S. (2017). Cold pools and their influence on the tropical marine boundary layer. *Journal of the Atmospheric Sciences*, 74(4), 1149–1168. doi: 10.1175/JAS-D-16-0264.1
Cited on pages 54, 176, 180, 185, and 192.
- Dipankar, A., Stevens, B., Heinze, R., Moseley, C., Zängl, G., Giorgetta, M., & Brdar, S. (2015). Large eddy simulation using the general circulation model icon. *Journal of Advances in Modeling Earth Systems*, 7(3), 963–986.
Cited on pages 58, 60, and 180.
- Dopplack, T. G. (1972). Radiative heating of the global atmosphere. *Journal of the Atmospheric Sciences*, 29(7), 1278–1294. doi: [https://doi.org/10.1175/1520-0469\(1972\)029\(1278:RHOTGA\)2.0.CO;2](https://doi.org/10.1175/1520-0469(1972)029(1278:RHOTGA)2.0.CO;2).
Cited on page 202.
- Drager, A. J., & van den Heever, S. C. (2017). Characterizing convective cold pools. *Journal of Advances in Modeling Earth Systems*, 9(2), 1091–1115.
Cited on pages 53, 55, 180, and 216.

- Droegemeier, K. K., & Wilhelmson, R. B. (1985). Three-dimensional numerical modeling of convection produced by interacting thunderstorm outflows. part ii: Variations in vertical wind shear. *Journal of the atmospheric sciences*, 42(22), 2404–2414.
Cited on pages 52, 176, and 215.
- Dueben, P. D., & Bauer, P. (2018). Challenges and design choices for global weather and climate models based on machine learning. *Geoscientific Model Development*, 11(10), 3999–4009.
Cited on page 24.
- Duhem, P. (1954). *The aim and structure of physical theory*. Princeton University Press.
Cited on page 158.
- Dunlap, R. E. (2013). Climate change skepticism and denial: An introduction. *American behavioral scientist*, 57(6), 691–698.
Cited on page 21.
- Dupont, S., Brunet, Y., & Finnigan, J. J. (2008). Large-eddy simulation of turbulent flow over a forested hill: Validation and coherent structure identification. *Quarterly Journal of the Royal Meteorological Society*, 134(636), 1911–1929.
Cited on page 45.
- Dyroff, C., Zahn, A., Christner, E., Forbes, R., Tompkins, A. M., & van Velthoven, P. F. (2015). Comparison of ecmwf analysis and forecast humidity data with caribic upper troposphere and lower stratosphere observations. *Quarterly Journal of the Royal Meteorological Society*, 141(688), 833–844. doi: <https://doi.org/10.1002/qj.2400>
Cited on page 210.
- Edwards, P. N. (2001). Representing the global atmosphere: Computer models, data, and knowledge about climate change. *Changing the atmosphere: Expert knowledge and environmental governance*, 31, 33.
Cited on pages 27, 28, 66, 166, and 217.
- Edwards, P. N. (2010). *A vast machine: Computer models, climate data, and the politics of global warming*. The MIT Press.
Cited on pages 43, 66, 87, 157, 161, 231, and 232.
- Emanuel, K., Wing, A. A., & Vincent, E. M. (2014). Radiative-convective instability. *Journal of Advances in Modeling Earth Systems*, 75–90. doi: 10.1002/2013MS000270.Received
Cited on page 211.
- Emanuel, K. A. (1983). Elementary aspects of the interaction between cumulus convection and the large-scale environment. In *Mesoscale meteorology—theories, observations and models* (pp. 551–575). Springer.
Cited on page 77.
- Emanuel, K. A. (1991). A scheme for representing cumulus convection in large-scale models. *Journal of Atmospheric Sciences*, 48(21), 2313–2329.
Cited on pages 71 and 85.
- Emanuel, K. A. (1993). A cumulus representation based on the episodic mixing model: the importance

- of mixing and microphysics in predicting humidity. In *The representation of cumulus convection in numerical models* (pp. 185–192). Springer.
Cited on pages 84, 87, and 120.
- Emanuel, K. A., et al. (1994). *Atmospheric convection*. Oxford University Press on Demand.
Cited on pages 38 and 53.
- Emanuel, K. A., & Živković-Rothman, M. (1999). Development and evaluation of a convection scheme for use in climate models. *Journal of the Atmospheric Sciences*, 56(11), 1766–1782.
Cited on pages 18, 82, and 213.
- Engerer, N. A., Stensrud, D. J., & Coniglio, M. C. (2008). Surface characteristics of observed cold pools. *Monthly Weather Review*, 136(12), 4839–4849.
Cited on page 54.
- Ent, R. J., & Tuinenburg, O. A. (2017). The residence time of water in the atmosphere revisited. *Hydrology and Earth System Sciences*, 21(2), 779–790.
Cited on page 15.
- EUREC⁴A official website. (2021). *Eurec4a platforms*. <http://eurec4a.eu/press-blog/press>. ([Online; accessed 3-May-2021])
Cited on page 17.
- European Centre for Medium-Range Weather Forecasts. (2017). *Era5 reanalysis*. Boulder CO: Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory. doi: 10.5065/D6X34W69
Cited on pages 198 and 199.
- Everitt, B., & Skrondal, A. (2002). *The cambridge dictionary of statistics* (Vol. 106). Cambridge University Press Cambridge.
Cited on page 107.
- Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the coupled model intercomparison project phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, 9(5), 1937–1958. doi: <https://doi.org/10.5194/gmd-9-1937-2016>
Cited on pages 148, 153, 154, 162, and 230.
- Feng, Z., Hagos, S., Rowe, A. K., Burleyson, C. D., Martini, M. N., & de Szoeke, S. P. (2015). Mechanisms of convective cloud organization by cold pools over tropical warm ocean during the amie/-dynamo field campaign. *Journal of Advances in Modeling Earth Systems*, 7(2), 357–381. Retrieved from <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2014MS000384>
doi: <https://doi.org/10.1002/2014MS000384>
Cited on page 176.
- Flato, G., & Marotzke, J. (2014). Evaluation of climate models. In *Climate change 2014: Synthesis report. contribution of working groups i, ii and iii to the fifth assessment report of the inter-governmental panel on climate change* (Mach, K. J. and Planton, S. and von Stechow, C. ed., p. 895). IPCC [Core Writing Team, R. K. Pachauri and L.A. Meyer (eds.)].
Cited on page 44.

- Fleagle, R. G. (1972). Bomex: An appraisal of results. *Science*, 176(4039), 1079–1084.
Cited on page 41.
- Franklin, A. (1989). *The neglect of experiment*. Cambridge University Press.
Cited on page 101.
- Franklin, A. (1990). *Experiment, right or wrong*. Cambridge University Press.
Cited on page 101.
- Franklin, A. (1999). Calibration. In *Can that be right?* (pp. 237–272). Springer.
Cited on pages 101 and 223.
- Frigg, R., & Reiss, J. (2009). The philosophy of simulation: hot new issues or same old stew? *Synthese*, 169(3), 593–613.
Cited on page 48.
- Frigg, R., Thompson, E., & Werndl, C. (2015). Philosophy of climate science part ii: Modelling climate change. *Philosophy Compass*, 10(12), 965–977.
Cited on page 27.
- Frisch, M. (2019). Calibration, validation, and confirmation. In C. Beisbart & N. J. Saam (Eds.), *Computer simulation validation: Fundamental concepts, methodological frameworks, and philosophical perspectives* (pp. 981–1004). Springer International Publishing. doi: 10.1007/978-3-319-70766-2_41
Cited on pages 157 and 231.
- Gabaix, X., & Laibson, D. (2008). The seven properties of good models. In *The foundations of positive and normative economics*. Oxford University Press. doi: 10.1093/acprof:oso/9780195328318.003.0012
Cited on page 149.
- Gaetani, M., & Mohino, E. (2013). Decadal prediction of the Sahelian precipitation in CMIP5 simulations. *Journal of Climate*, 26(19), 7708–7719. doi: 10.1175/JCLI-D-12-00635.1
Cited on page 155.
- Gao, S., Cui, X., Zhou, Y., & Li, X. (2005). Surface rainfall processes as simulated in a cloud-resolving model. *Journal of Geophysical Research: Atmospheres*, 110(D10).
Cited on page 18.
- Garand, L., Turner, D. S., Larocque, M., Bates, J., Boukabara, S., Brunel, P., ... Woolf, H. (2001, October). Radiance and Jacobian intercomparison of radiative transfer models applied to HIRS and AMSU channels. *J. Geophys. Res.*, 106(D20), 24017–24031. doi: 10.1029/2000JD000184
Cited on page 199.
- Gates, W. L. (1992). An AMS continuing series: Global change–AMIP: The atmospheric model intercomparison project. *Bulletin of the American Meteorological Society*, 73(12), 1962–1970.
Cited on pages 150, 158, and 231.
- Gates, W. L. (2015). *Comments of the history of the working group on numerical experimentation*. (http://www.wmo.int/pages/prog/arep/wwrp/new/Presentations_wgne30_March_2015.html)

Cited on page 149.

Gates, W. L., Boyle, J. S., Covey, C., Dease, C. G., Doutriaux, C. M., Drach, R. S., ... Williams, D. N. (1999). An overview of the results of the atmospheric model intercomparison project (AMIP I). *Bulletin of the American Meteorological Society*, 80(1), 29-56.
Cited on pages 150, 151, and 158.

Gentine, P., Garelli, A., Park, S.-B., Nie, J., Torri, G., & Kuang, Z. (2016). Role of surface heat fluxes underneath cold pools. *Geophysical research letters*, 43(2), 874–883.
Cited on pages 55 and 216.

Gentine, P., Pritchard, M., Rasp, S., Reinaudi, G., & Yacalis, G. (2018). Could machine learning break the convection parameterization deadlock? *Geophysical Research Letters*, 45(11), 5742–5751.
Cited on pages 23, 66, and 166.

George, G., Stevens, B., Bony, S., Pincus, R., Fairall, C., Schulz, H., ... others (2021). Joanne: Joint dropsonde observations of the atmosphere in tropical north atlantic meso-scale environments. *Earth System Science Data Discussions*, 1–33.
Cited on pages 178, 194, 197, and 198.

Gettelman, A., Truesdale, J., Bacmeister, J., Caldwell, P., Neale, R., Bogenschutz, P., & Simpson, I. (2019). The single column atmosphere model version 6 (scam6): Not a scam but a tool for model evaluation and development. *Journal of Advances in Modeling Earth Systems*, 11(5), 1381–1401.
Cited on page 23.

Giorgi, F., & et al. (2001). *Climate change 2001: The scientific basis* (J. Houghton et al., Eds.). Cambridge University Press.
Cited on pages 48 and 152.

Golaz, J.-C., Horowitz, L. W., & Levy, H. (2013). Cloud tuning in a coupled climate model: Impact on 20th century warming. *Geophysical Research Letters*, 40(10), 2246–2251.
Cited on pages 104 and 105.

Gough, D. (1969). The anelastic approximation for thermal convection. *Journal of Atmospheric Sciences*, 26(3), 448–456.
Cited on page 79.

Grabowski, W. W. (2001). Coupling cloud processes with the large-scale dynamics using the cloud-resolving convection parameterization (crp). *Journal of the Atmospheric Sciences*, 58(9), 978–997.
Cited on page 22.

Grabowski, W. W. (2014). Extracting microphysical impacts in large-eddy simulations of shallow convection. *Journal of the Atmospheric Sciences*, 71(12), 4493–4499.
Cited on page 127.

Grabowski, W. W., & Smolarkiewicz, P. K. (1999). Crp: A cloud resolving convection parameterization for modeling the tropical convecting atmosphere. *Physica D: Nonlinear Phenomena*, 133(1-4), 171–178.
Cited on page 22.

- Grandpeix, J.-Y., & Lafore, J.-P. (2010). A density current parameterization coupled with emanuel's convection scheme. part i: The models. *Journal of the Atmospheric Sciences*, 67(4), 881–897. Cited on pages 54, 55, 59, 82, 85, 86, 87, 90, 94, 95, 117, 176, 215, 216, and 220.
- Grandpeix, J.-Y., Phillips, V., & Tailleux, R. (2004). Improved mixing representation in emanuel's convection scheme. *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography*, 130(604), 3207–3222. Cited on pages 82, 84, and 120.
- Grant, L. D., Moncrieff, M. W., Lane, T. P., & van den Heever, S. C. (2020). Shear-parallel tropical convective systems: Importance of cold pools and wind shear. *Geophysical Research Letters*, 47(12), e2020GL087720. doi: <https://doi.org/10.1029/2020GL087720> Cited on pages 61 and 192.
- Gray, W., & Jacobson, R. (1977). Diurnal variation of deep cumulus convection. *Monthly Weather Review*, 105, 1171–1187. doi: 10.1175/1520-0493(1977)105(1171:DVODCC)2.0.CO;2 Cited on page 211.
- Gray, W. M., & Jacobson Jr, R. W. (1977). Diurnal variation of deep cumulus convection. *Monthly Weather Review*, 105(9), 1171–1188. doi: 10.1175/1520-0493(1977)105(1171:DVODCC)2.0.CO;2 Cited on pages 202 and 211.
- Guemas, V., García-Serrano, J., Mariotti, A., Doblas-Reyes, F., & Caron, L.-P. (2015). Prospects for decadal climate prediction in the Mediterranean region. *Quarterly Journal of the Royal Meteorological Society*, 141(687), 580–597. doi: 10.1002/qj.2379 Cited on page 155.
- Guichard, F., Parsons, D., & Miller, E. (2000). Thermodynamic and radiative impact of the correction of sounding humidity bias in the tropics. *Journal of Climate*, 13(20), 3611–3624. doi: 10.1175/1520-0442(2000)013(3611:TARIOT)2.0.CO;2 Cited on pages 197 and 199.
- Guichard, F., Petch, J., Redelsperger, J.-L., Bechtold, P., Chaboureaud, J.-P., Cheinet, S., ... others (2004). Modelling the diurnal cycle of deep precipitating convection over land with cloud-resolving models and single-column models. *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography*, 130(604), 3139–3172. Cited on page 20.
- Guillemot, H. (2009). Comment évaluer un modèle numérique de climat? *Revue d'anthropologie des connaissances*, 3(2), 273–293. Cited on page 16.
- Guillemot, H. (2017). How to develop climate models? the “gamble” of improving climate model parameterizations. *Cultures of prediction in atmospheric and climate science: Epistemic and cultural shifts in computer-based modelling and simulation*, 120–136. Cited on pages 26 and 66.
- Gutleben, M., Groß, S., Wirth, M., & Mayer, B. (2020). Radiative effects of long-range-transported

- saharan air layers as determined from airborne lidar measurements. *Atmospheric Chemistry and Physics*, 20(20), 12313–12327.
Cited on page 199.
- Gutleben, M., Groß, S. M., & Wirth, M. (2019). Cloud macro-physical properties in saharan-dust-laden and dust-free north atlantic trade wind regimes: a lidar case study. *Atmospheric Chemistry and Physics (ACP)*, 19(16), 10659–10673. doi: 10.5194/acp-19-10659-2019
Cited on pages 197, 199, 202, and 207.
- Haerter, J. O. (2019). Convective self-aggregation as a cold pool-driven critical phenomenon. *Geophysical Research Letters*, 46(7), 4017–4028.
Cited on pages 55 and 216.
- Haerter, J. O., Böing, S. J., Henneberg, O., & Nissen, S. B. (2019). Circling in on convective organization. *Geophysical Research Letters*, 46(12), 7024–7034.
Cited on pages 55 and 216.
- Hane, C. E. (1973). The squall line thunderstorm: Numerical experimentation. *Journal of Atmospheric Sciences*, 30(8), 1672–1690.
Cited on page 53.
- Hardwig, J. (1985). Epistemic dependence. *Journal of Philosophy*, 82(7), 335–349. doi: jphil198582747
Cited on page 159.
- Harvey, B., Cook, P., Shaffrey, L., & Schiemann, R. (2020). The response of the northern hemisphere storm tracks and jet streams to climate change in the cmip3, cmip5, and cmip6 climate models. *Journal of Geophysical Research: Atmospheres*, 125(23), e2020JD032701.
Cited on page 168.
- Hawkins, D. M. (2004). The problem of overfitting. *Journal of chemical information and computer sciences*, 44(1), 1–12.
Cited on page 107.
- Heggem, T., Lende, R., & Løvseth, J. (1998). Analysis of long time series of coastal wind. *Journal of the atmospheric sciences*, 55(18), 2907–2917.
Cited on pages 22, 76, and 89.
- Heinze, R., Dipankar, A., Henken, C. C., Moseley, C., Sourdeval, O., Trömel, S., ... others (2017). Large-eddy simulations over germany using icon: A comprehensive evaluation. *Quarterly Journal of the Royal Meteorological Society*, 143(702), 69–100.
Cited on page 58.
- Heisenberg, W. (1970). Physics and beyond-encounters and conversations. *SCIENCE AND PUBLIC AFFAIRS-BULLETIN OF THE ATOMIC SCIENTISTS*, 26(9), 33–36.
Cited on page 63.
- Held, I. (2005). The gap between simulation and understanding in climate modeling. *Bulletin of the American Meteorological Society*(86), 1609–1614.
Cited on pages 21 and 51.

- Held, I. M., & Soden, B. J. (2006). Robust responses of the hydrological cycle to global warming. *Journal of climate*, 19(21), 5686–5699.
Cited on pages 15 and 168.
- Heymann, M., & Hundebøl, N. R. (2017). From heuristic to predictive: Making climate models into political instruments. In M. Heymann, G. Gramelsberger, & M. Mahony (Eds.), *Cultures of prediction in atmospheric and climate science*. London: Routledge.
Cited on pages 24 and 25.
- Hirt, M., & Craig, G. C. (2021). A cold pool perturbation scheme to improve convective initiation in convection-permitting models. *Quarterly Journal of the Royal Meteorological Society*.
Cited on pages 55 and 216.
- Hirt, M., Craig, G. C., Schäfer, S. A., Savre, J., & Heinze, R. (2020). Cold pool driven convective initiation: using causal graph analysis to determine what convection permitting models are missing. *Quarterly Journal of the Royal Meteorological Society*.
Cited on pages 45 and 56.
- Hoffman, A. J. (2011). The culture and discourse of climate skepticism. *Strategic Organization*, 9(1), 77–84.
Cited on page 21.
- Hohenegger, C., Brockhaus, P., Bretherton, C. S., & Schär, C. (2009). The soil moisture–precipitation feedback in simulations with explicit and parameterized convection. *Journal of Climate*, 22(19), 5003–5020.
Cited on page 15.
- Holland, J. Z. (1970). Preliminary report on the bomex sea-air interaction program. *Bulletin of the American Meteorological Society*, 51(9), 809–821.
Cited on page 40.
- Holton, J., & Hakim, G. (2012). An introduction to dynamic meteorology, 5th edn., 552 pp. *Academic, Cambridge*.
Cited on page 75.
- Horner, J. K., & Symons, J. F. (2020). Software engineering standards for epidemiological models. *History and Philosophy of the Life Sciences*, 42(4), 1–24.
Cited on page 170.
- Houghton, D. D., & Irvine, W. S. (1976). A case study comparison of the performance of operational prediction models used in the United States. *Monthly Weather Review*, 104(7), 817–827.
Cited on page 149.
- Houghton, J., Callander, B., & Varney, S. (Eds.). (1992). *Climate change 1992: the supplementary report to the IPCC scientific assessment*. Cambridge University Press.
Cited on page 151.
- Houghton, J., Jenkins, G., & Ephraums, J. (Eds.). (1990). *IPCC first assessment report 1990*. Cambridge University Press.
Cited on pages 151 and 152.

- Hourdin, F., Couvreux, F., & Menut, L. (2002). Parameterization of the dry convective boundary layer based on a mass flux representation of thermals. *Journal of the atmospheric sciences*, 59(6), 1105–1123.
Cited on pages 82, 83, 84, and 87.
- Hourdin, F., Dufresne, J. L., & Fabart, V. (2010). La lettre des utilisateurs du modèle de circulation générale de l’institut pierre simon laplace (ipsl). *Laboratoire de Meteorologie Dynamique, LMDZinfo*(7).
Cited on page 47.
- Hourdin, F., Grandpeix, J.-Y., Rio, C., Bony, S., Jam, A., Cheruy, F., ... others (2013). Lmdz5b: the atmospheric component of the ipsl climate model with revisited parameterizations for clouds and convection. *Climate Dynamics*, 40(9-10), 2193–2222.
Cited on pages 92, 104, 159, and 232.
- Hourdin, F., Mauritsen, T., Gettelman, A., Golaz, J.-C., Balaji, V., Duan, Q., ... others (2017a). The art and science of climate model tuning. *Bulletin of the American Meteorological Society*, 98(3), 589–602.
Cited on pages 22, 92, 97, 99, 100, 103, 104, 106, 107, 108, 157, 221, 222, and 226.
- Hourdin, F., Mauritsen, T., Gettelman, A., Golaz, J.-C., Balaji, V., Duan, Q., ... others (2017b). Supplement: The art and science of climate model tuning. *Bulletin of the American Meteorological Society*, 98(3), ES61–ES66.
Cited on pages 99, 103, 109, and 222.
- Hourdin, F., Rio, C., Grandpeix, J.-Y., Madeleine, J.-B., Cheruy, F., Rochetin, N., ... others (2020). Lmdz6a: The atmospheric component of the ipsl climate model with improved and better tuned physics. *Journal of Advances in Modeling Earth Systems*, 12(7), e2019MS001892.
Cited on page 143.
- Hourdin, F., Williamson, D., Rio, C., Couvreux, F., Roehrig, R., Villefranque, N., ... Volodina, V. (2020). Process-based climate model development harnessing machine learning: Ii. model calibration from single column to global. *Journal of Advances in Modeling Earth Systems*, e2020MS002225.
Cited on pages 23, 107, 109, 130, 131, 132, 133, 136, and 226.
- Houze Jr, R. A., & Betts, A. K. (1981). Convection in gate. *Reviews of Geophysics*, 19(4), 541–576.
Cited on pages 41, 53, 56, and 193.
- Hu, Y., Tao, L., & Liu, J. (2013). Poleward expansion of the Hadley circulation in CMIP5 simulations. *Advances in Atmospheric Sciences*, 30(3), 790–795. doi: 10.1007/s00376-012-2187-4
Cited on page 155.
- Hurrell, J., Meehl, G. A., Bader, D., Delworth, T. L., Kirtman, B., & Wielicki, B. (2009). A unified modeling approach to climate system prediction. *Bulletin of the American Meteorological Society*, 90(12), 1819–1832. doi: 10.1175/2009BAMS2752.1
Cited on page 161.
- Hwong, Y.-L., Song, S., Sherwood, S., Stirling, A., Rio, C., Roehrig, R., ... others (2021). Characterizing convection schemes using their responses to imposed tendency perturbations. *Journal of Advances in Modeling Earth Systems*, 13(5), e2021MS002461.

Cited on pages 133 and 165.

IPCC. (2014). Detection and attribution of climate change: from global to regional. In *Climate change 2014: Synthesis report. contribution of working groups i, ii and iii to the fifth assessment report of the intergovernmental panel on climate change* (Mach, K. J. and Planton, S. and von Stechow, C. ed., p. 895). IPCC [Core Writing Team, R. K. Pachauri and L.A. Meyer (eds.)].

Cited on page 48.

Ishida, H. (1989). Spectra of surface wind speed and air temperature over the ocean in the mesoscale frequency range in jasin-1978. *Boundary-layer meteorology*, 47(1), 71–84.

Cited on pages 22, 76, and 89.

Jakob, C. (2010). Accelerating progress in global atmospheric model development through improved parameterizations: Challenges, opportunities, and strategies. *Bulletin of the American Meteorological Society*, 91(7), 869–876.

Cited on pages 110, 124, and 168.

Jakob, C., & Klein, S. A. (1999). The role of vertically varying cloud fraction in the parametrization of microphysical processes in the ecmwf model. *Quarterly Journal of the Royal Meteorological Society*, 125(555), 941–965.

Cited on page 124.

Jakob, C., & Klein, S. A. (2000). A parametrization of the effects of cloud and precipitation overlap for use in general-circulation models. *Quarterly Journal of the Royal Meteorological Society*, 126(568), 2525–2544.

Cited on pages 120, 121, 122, 123, 124, 125, 144, and 227.

Jam, A., Hourdin, F., Rio, C., & Couvreux, F. (2013). Resolved versus parametrized boundary-layer plumes. part iii: Derivation of a statistical scheme for cumulus clouds. *Boundary-layer meteorology*, 147(3), 421–441.

Cited on pages 83, 96, and 119.

Jansson, F., Edeling, W., Attema, J., & Crommelin, D. (2021). Assessing uncertainties from physical parameters and modelling choices in an atmospheric large eddy simulation model. *Philosophical Transactions of the Royal Society A*, 379(2197), 20200073.

Cited on page 49.

Jayasankar, C. B., Surendran, S., & Rajendran, K. (2015). Robust signals of future projections of indian summer monsoon rainfall by IPCC AR5 climate models: Role of seasonal cycle and inter-annual variability. *Geophysical Research Letters*, 42(9), 3513–3520. doi: 10.1002/2015GL063659

Cited on pages 155, 156, and 230.

Jebeile, J. (2013). *Expliquer et comprendre dans les sciences empiriques : les modèles scientifiques et le tournant computationnel* (Thèse de doctorat dirigée par Dubucs, Jacques et Barberousse, Anouk). Université Paris 1. (<http://www.theses.fr/2013PA010695>)

Cited on page 21.

Jebeile, J., & Crucifix, M. (2020). Multi-model ensembles in climate science: Mathematical structures and expert judgements. *Studies in History and Philosophy of Science Part A*, 83, 44–52.

Cited on page 115.

- Jebeile, J., Lam, V., & R  z, T. (2020). Understanding climate change with statistical downscaling and machine learning. *Synthese*, 1–21.
Cited on page 24.
- Jeevanjee, N., & Fueglistaler, S. (2020). Simple spectral models for atmospheric radiative cooling. *Journal of the Atmospheric Sciences*, 77(2), 479–497. doi: <https://doi.org/10.1175/JAS-D-18-0347.1>.
Cited on page 202.
- Jeevanjee, N., Hassanzadeh, P., Hill, S., & Sheshadri, A. (2017). A perspective on climate model hierarchies. *Journal of Advances in Modeling Earth Systems*, 9(4), 1760–1771.
Cited on pages 21 and 51.
- Jeevanjee, N., & Romps, D. M. (2015). Effective buoyancy, inertial pressure, and the mechanical generation of boundary layer mass flux by cold pools. *Journal of the Atmospheric Sciences*, 72(8), 3199–3213. doi: <https://doi.org/10.1175/JAS-D-14-0349.1>
Cited on page 176.
- Jiang, X., Waliser, D. E., Xavier, P. K., Petch, J., Klingaman, N. P., Woolnough, S. J., ... others (2015). Vertical structure and physical processes of the madden-julian oscillation: Exploring key model physics in climate simulations. *Journal of Geophysical Research: Atmospheres*, 120(10), 4718–4748.
Cited on pages 104 and 223.
- Johnson, G. C., Lyman, J. M., & Loeb, N. G. (2016). Improving estimates of earth’s energy imbalance. *Nature Climate Change*, 6(7), 639–640.
Cited on page 138.
- Jones, N. (2017). How machine learning could help to improve climate forecasts. *Nature News*, 548(7668), 379.
Cited on page 24.
- Jones, T. R., & Randall, D. A. (2011). Quantifying the limits of convective parameterizations. *Journal of Geophysical Research: Atmospheres*, 116(D8).
Cited on page 76.
- Joshi, M. M., Gregory, J. M., Webb, M. J., Sexton, D. M., & Johns, T. C. (2008). Mechanisms for the land/sea warming contrast exhibited by simulations of climate change. *Climate Dynamics*, 30(5), 455–465.
Cited on page 15.
- J  rgensen, S. E. (2008). Overview of the model types available for development of ecological models. *Ecological Modelling*, 215(1), 3 - 9. doi: <https://doi.org/10.1016/j.ecolmodel.2008.02.041>
Cited on page 149.
- Kasahara, A. (1961). A numerical experiment on the development of a tropical cyclone. *Journal of Atmospheric Sciences*, 18(3), 259–282.
Cited on pages 69 and 218.
- Kato, S., Ackerman, T. P., Clothiaux, E. E., Mather, J. H., Mace, G. G., Wesely, M. L., ... Michalsky, J. (1997). Uncertainties in modeled and measured clear-sky surface shortwave irradiances. *Jour-*

- nal of Geophysical Research: Atmospheres*, 102(D22), 25881–25898. doi: 10.1029/97JD01841
Cited on page 197.
- Katzav, J. (2013). Hybrid models, climate models, and inference to the best explanation. *The British journal for the philosophy of science*, 64(1), 107–129.
Cited on page 66.
- Katzav, J. (2014). The epistemology of climate models and some of its implications for climate science and the philosophy of science. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 46, 228–238.
Cited on page 26.
- Katzav, J., & Parker, W. S. (2015). The future of climate modeling. *Climatic Change*, 132(4), 475–487. doi: 10.1007/s10584-015-1435-x
Cited on page 161.
- Katzav, J., & Parker, W. S. (2018). Issues in the theoretical foundations of climate science. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 63, 141 - 149.
Cited on page 24.
- Kent, C., Chadwick, R., & Rowell, D. P. (2015). Understanding uncertainties in future projections of seasonal tropical precipitation. *Journal of Climate*, 28(11), 4390–4413. doi: 10.1175/JCLI-D-14-00613.1
Cited on page 158.
- Kerr, R. A. (1994). Climate modeling’s fudge factor comes under fire. *Science*, 265(5178), 1528–1529.
Cited on page 99.
- Kerr, R. A. (1997). Climate change: Model gets it right—without fudge factors. *Science*, 276(5315).
Cited on page 99.
- Khairoutdinov, M., Randall, D., & DeMott, C. (2005). Simulations of the atmospheric general circulation using a cloud-resolving model as a superparameterization of physical processes. *Journal of the Atmospheric Sciences*, 62(7), 2136–2154.
Cited on page 22.
- Kirchner, J. W. (2003). The gaia hypothesis: conjectures and refutations. *Climatic Change*, 58(1), 21–45.
Cited on page 37.
- Klein, E. E., & Herskovitz, P. J. (2005). Philosophical foundations of computer simulation validation. *Simulation & Gaming*, 36(3), 303–329.
Cited on page 48.
- Klemp, J. B., & Wilhelmson, R. B. (1978). The simulation of three-dimensional convective storm dynamics. *Journal of the Atmospheric Sciences*, 35(6), 1070–1096.
Cited on page 53.
- Knupp, K. R., & Cotton, W. R. (1982). An intense, quasi-steady thunderstorm over mountainous terrain. Part II: Doppler radar observations of the storm morphological structure. *Journal of the*

Atmospheric Sciences, 39(2), 343–358. doi: [https://doi.org/10.1175/1520-0469\(1982\)039<0343:AIQSTO>2.0.CO;2](https://doi.org/10.1175/1520-0469(1982)039<0343:AIQSTO>2.0.CO;2)
Cited on page 183.

Knutti, R. (2010). The end of model democracy? *Climatic Change*, 102(3), 395–404.
Cited on pages 107 and 157.

Knutti, R. (2018). Climate model confirmation: From philosophy to predicting climate in the real world. In *Climate modelling* (pp. 325–359). Springer.
Cited on page 26.

Knutti, R., Furrer, R., Tebaldi, C., Cermak, J., & Meehl, G. A. (2010). Challenges in combining projections from multiple climate models. *Journal of Climate*, 23(10), 2739–2758.
Cited on pages 149, 156, and 157.

Knutti, R., Masson, D., & Gettelman, A. (2013). Climate model genealogy: Generation CMIP5 and how we got there. *Geophysical Research Letters*, 40(6), 1194–1199.
Cited on page 157.

Knutti, R., Sedláček, J., Sanderson, B. M., Lorenz, R., Fischer, E. M., & Eyring, V. (2017). A climate model projection weighting scheme accounting for performance and interdependence. *Geophysical Research Letters*, 44(4), 1909–1918. doi: 10.1002/2016GL072012
Cited on pages 140 and 157.

Koutroulis, A., Grillakis, M., Tsanis, I., & Papadimitriou, L. (2016). Evaluation of precipitation and temperature simulation performance of the cmip3 and cmip5 historical experiments. *Climate Dynamics*, 47(5), 1881–1898.
Cited on page 25.

Kuettner, J. P. (1971). Cloud bands in the earth's atmosphere: Observations and theory. *Tellus*, 23(4-5), 404–426.
Cited on page 40.

Kuhn, T. S. (1962). The structure of scientific revolutions: University of chicago press. *Original edition*.
Cited on pages 114, 117, and 227.

Kurowski, M. J., Suselj, K., Grabowski, W. W., & Teixeira, J. (2018). Shallow-to-deep transition of continental moist convection: Cold pools, surface fluxes, and mesoscale organization. *Journal of the Atmospheric Sciences*, 75(12), 4071–4090.
Cited on pages 52, 176, and 215.

Lac, C., Chaboureaud, J.-P., Masson, V., Pinty, J.-P., Tulet, P., Escobar, J., . . . others (2018). Overview of the meso-nh model version 5.4 and its applications. *Geoscientific Model Development*, 11(5), 1929–1969.
Cited on pages 118 and 126.

Langhans, W., & Roms, D. M. (2015). The origin of water vapor rings in tropical oceanic cold pools. *Geophysical Research Letters*, 42(18), 7825–7834.
Cited on pages 52, 176, and 215.

- Leahy, L., Wood, R., Charlson, R., Hostetler, C., Rogers, R., Vaughan, M., & Winker, D. (2012). On the nature and extent of optically thin marine low clouds. *Journal of Geophysical Research: Atmospheres*, 117(D22). doi: 10.1029/2012JD017929
Cited on page 207.
- Leary, C. A., & Houze Jr, R. A. (1979). The structure and evolution of convection in a tropical cloud cluster. *Journal of the Atmospheric Sciences*, 36(3), 437–457.
Cited on pages 53 and 193.
- L’Ecuyer, T. S., Wood, N. B., Haladay, T., Stephens, G. L., & Stackhouse Jr., P. W. (2008). Impact of clouds on atmospheric heating based on the r04 cloudsat fluxes and heating rates data set. *Journal of Geophysical Research: Atmospheres*, 113(D8). Retrieved from <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2008JD009951> doi: 10.1029/2008JD009951
Cited on page 197.
- Le Moigne, J.-L. (1990). La science informatique va-t-elle construire sa propre épistémologie? *Culture technique*.
Cited on page 20.
- Lenderink, G., Siebesma, A. P., Cheinet, S., Irons, S., Jones, C. G., Marquet, P., ... others (2004). The diurnal cycle of shallow cumulus clouds over land: A single-column model intercomparison study. *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography*, 130(604), 3339–3364.
Cited on page 20.
- Lenhard, J., & Winsberg, E. (2010). Holism, entrenchment, and the future of climate model pluralism. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 41(3), 253–262.
Cited on pages 24, 27, 47, 104, 107, 115, 158, 159, 224, and 231.
- Lenschow, D. H., & Sun, J. (2007). The spectral composition of fluxes and variances over land and sea out to the mesoscale. *Boundary-Layer Meteorology*, 125(1), 63–84. doi: <https://doi.org/10.1007/s10546-007-9191-8>
Cited on page 186.
- Levins, R. (1966). The strategy of model building in population biology. *American Scientist*, 54(4), 421–431. Retrieved from <http://www.jstor.org/stable/27836590>
Cited on page 51.
- Lévy-Leblond, J.-M. (2000). Mots & maux de la physique quantique: Critique épistémologique et problèmes terminologiques. *Revue internationale de philosophie*, 243–265.
Cited on page 32.
- Lilly, D. K. (1962). On the numerical simulation of buoyant convection. *Tellus*, 14(2), 148–172.
Cited on pages 69 and 218.
- Lilly, D. K. (1967). The representation of small-scale turbulence in numerical simulation experiments. *IBM Form*, 195–210.
Cited on page 44.
- Lilly, D. K. (1968). Models of cloud-topped mixed layers under a strong inversion. *Quarterly Journal*

of the *Royal Meteorological Society*, 94(401), 292–309.

Cited on page 89.

Lin, J.-L. (2007). The double-itz problem in ipcc ar4 coupled gcms: Ocean–atmosphere feedback analysis. *Journal of Climate*, 20(18), 4497–4525.

Cited on pages 104 and 223.

Lin, J.-L., Kiladis, G. N., Mapes, B. E., Weickmann, K. M., Sperber, K. R., Lin, W., ... others (2006). Tropical intraseasonal variability in 14 ipcc ar4 climate models. part i: Convective signals. *Journal of climate*, 19(12), 2665–2690.

Cited on pages 104 and 223.

Lloyd, E. A. (2009). I—elisabeth a. lloyd: Varieties of support and confirmation of climate models. In *Aristotelian society supplementary volume* (Vol. 83, pp. 213–232).

Cited on pages 25, 28, 92, 217, and 221.

Lloyd, E. A. (2010). Confirmation and robustness of climate models. *Philosophy of Science*, 77(5), 971–984.

Cited on pages 26, 92, and 221.

Loeb, N. G., Wielicki, B. A., Doelling, D. R., Smith, G. L., Keyes, D. F., Kato, S., ... Wong, T. (2009). Toward optimal closure of the earth’s top-of-atmosphere radiation budget. *Journal of Climate*, 22(3), 748–766.

Cited on pages 37, 136, and 138.

Long, D. J., & Collins, M. (2013). Quantifying global climate feedbacks, responses and forcing under abrupt and gradual CO2 forcing. *Climate Dynamics*, 41(9), 2471–2479. doi: 10.1007/s00382-013-1677-0

Cited on page 158.

Lonitz, K., Stevens, B., Nuijens, L., Sant, V., Hirsch, L., & Seifert, A. (2015). The signature of aerosols and meteorology in long-term cloud radar observations of trade wind cumuli. *Journal of the Atmospheric Sciences*, 72(12), 4643–4659.

Cited on page 199.

Lord, S. J. (1982). Interaction of a cumulus cloud ensemble with the large-scale environment. part iii: Semi-prognostic test of the arakawa-schubert cumulus parameterization. *Journal of Atmospheric Sciences*, 39(1), 88–103.

Cited on pages 38 and 70.

Lord, S. J., & Arakawa, A. (1980). Interaction of a cumulus cloud ensemble with the large-scale environment. part ii. *Journal of Atmospheric Sciences*, 37(12), 2677–2692.

Cited on pages 38 and 70.

Luo, B., & Minnett, P. (2020). Evaluation of the era5 sea surface skin temperature with remotely-sensed shipborne marine-atmospheric emitted radiance interferometer data. *Remote Sensing*, 12(11), 1873. doi: <https://doi.org/10.3390/rs12111873>

Cited on page 209.

Luomaranta, A., Ruosteenoja, K., Jylhä, K., Gregow, H., Haapala, J., & Laaksonen, A. (2014). Multimodel estimates of the changes in the baltic sea ice cover during the present century. *Tellus*

- A: Dynamic Meteorology and Oceanography*, 66(1), 22617. doi: 10.3402/tellusa.v66.22617
Cited on page 155.
- Luther, F. M., Ellingson, R. G., Fouquart, Y., Fels, S., Scott, N. A., & Wiscombe, W. J. (1988). Intercomparison of radiation codes in climate models (ICRCCM): Longwave clear-sky results—a workshop summary. *Bulletin of the American Meteorological Society*, 69(1), 40–48.
Cited on page 150.
- Mach, E. (1882). *Die ökonomische natur der physikalischen forschung*. KK Hof-und Staatsdruckerei.
Cited on page 63.
- Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of science*, 67(1), 1–25.
Cited on page 36.
- Maddy, E. S., & Barnet, C. D. (2008, August). Vertical Resolution Estimates in Version 5 of AIRS Operational Retrievals. *IEEE Trans. Geosci. Remote Sensing*, 46(8), 2375–2384. doi: 10.1109/TGRS.2008.917498
Cited on page 197.
- Madeleine, J.-b., Hourdin, F., Grandpeix, J.-y., Rio, C., Dufresne, J.-L., Vignon, E., . . . others (2020). Improved representation of clouds in the atmospheric component lmdz6a of the ipsl-cm6a earth system model. *Journal of Advances in Modeling Earth Systems*, 12(10), e2020MS002046.
Cited on pages 119, 123, and 165.
- Maher, P., Gerber, E. P., Medeiros, B., Merlis, T. M., Sherwood, S., Sheshadri, A., . . . Zurita-Gotor, P. (2019). Model hierarchies for understanding atmospheric circulation. *Reviews of Geophysics*, 57(2), 250–280. doi: 10.1029/2018RG000607
Cited on page 160.
- Mahoney, K. M., Lackmann, G. M., & Parker, M. D. (2009). The role of momentum transport in the motion of a quasi-idealized mesoscale convective system. *Monthly weather review*, 137(10), 3316–3338. doi: <https://doi.org/10.1175/2009MWR2895.1>
Cited on pages 61 and 192.
- Malkus, J. S. (1958). *On the structure of the trade wind moist layer*. Massachusetts Institute of Technology.
Cited on pages 196 and 202.
- Manabe, S., & Bryan, K. (1969). Climate calculations with a combined ocean-atmosphere model. *Journal of the Atmospheric Sciences*, 26, 786–789.
Cited on page 151.
- Manabe, S., Smagorinsky, J., & Strickler, R. F. (1965). Simulated climatology of a general circulation model with a hydrologic cycle. *Monthly Weather Review*, 93(12), 769–798.
Cited on pages 69, 70, and 218.
- Manabe, S., & Strickler, R. F. (1964). Thermal equilibrium of the atmosphere with a convective adjustment. *Journal of Atmospheric Sciences*, 21(4), 361–385.
Cited on pages 67, 69, and 218.

- Manabe, S., & Wetherald, R. T. (1967). Thermal equilibrium of the atmosphere with a given distribution of relative humidity. *Journal of the Atmospheric Sciences*, *24*(3), 241–258.
Cited on pages 67, 168, and 172.
- Manabe, S., & Wetherald, R. T. (1975). The effects of doubling the CO₂ concentration on the climate of a general circulation model. *Journal of the Atmospheric Sciences*, *32*(1), 3–15. doi: 10.1175/1520-0469(1975)032<0003:TEODTC>2.0.CO;2
Cited on pages 161 and 232.
- Manrique, J. A., Lopez-Reyes, G., Cousin, A., Rull, F., Maurice, S., Wiens, R. C., ... others (2020). Supercam calibration targets: design and development. *Space Science Reviews*, *216*(8), 1–27.
Cited on pages 92 and 221.
- Mapes, B. E. (2001). Water’s two height scales: The moist adiabat and the radiative troposphere. *Quart. J. Roy. Meteor. Soc.*, *127*(577), 2353–2366. doi: 10.1175/1520-0469(1998)055<1354:ETGMSO>2.0.CO;2
Cited on page 211.
- Marotzke, J., & Stone, P. H. (1995). Atmospheric transports, the thermohaline circulation, and flux adjustments in a simple coupled model. *Journal of Physical Oceanography*, *25*(6), 1350–1364.
Cited on pages 98 and 222.
- Martinich, A., & Stroll, A. (2021). Epistemology. In *Encyclopedia britannica*. Retrieved 2021-06-25, from <https://www.britannica.com/topic/epistemology>
Cited on page 20.
- Masmoudi, N. (2007). Rigorous derivation of the anelastic approximation. *Journal de mathématiques pures et appliquées*, *88*(3), 230–240.
Cited on page 79.
- Masson, D., & Knutti, R. (2011). Climate model genealogy. *Geophysical Research Letters*, *38*(8). doi: 10.1029/2011GL046864
Cited on page 157.
- Mauritsen, T., & Roeckner, E. (2020). Tuning the mpi-esm1.2 global climate model to improve the match with instrumental record warming by lowering its climate sensitivity. *Journal of advances in modeling earth systems*, *12*(5), e2019MS002037.
Cited on pages 92, 111, and 221.
- Mauritsen, T., Stevens, B., Roeckner, E., Crueger, T., Esch, M., Giorgetta, M., ... others (2012). Tuning the climate of a global model. *Journal of advances in modeling Earth systems*, *4*(3).
Cited on pages 92 and 221.
- McAvaney, & et al. (2001). *Climate change 2001: The scientific basis* (J. Houghton et al., Eds.). Cambridge University Press.
Cited on page 152.
- McCaul Jr, E. W., Goodman, S. J., LaCasse, K. M., & Cecil, D. J. (2009). Forecasting lightning threat using cloud-resolving model simulations. *Weather and Forecasting*, *24*(3), 709–729.
Cited on page 18.

- McFarlane, N. (2011). Parameterizations: representing key processes in climate models without resolving them. *Wiley Interdisciplinary Reviews: Climate Change*, 2(4), 482–497.
Cited on page 49.
- McGovern, A., Lagerquist, R., Gagne, D. J., Jergensen, G. E., Elmore, K. L., Homeyer, C. R., & Smith, T. (2019). Making the black box more transparent: Understanding the physical implications of machine learning. *Bulletin of the American Meteorological Society*, 100(11), 2175–2199.
Cited on page 24.
- Medeiros, B., & Nuijens, L. (2016). Clouds at barbados are representative of clouds across the trade wind regions in observations and climate models. *Proceedings of the National Academy of Sciences*, 113(22), E3062–E3070. doi: 10.1073/pnas.1521494113
Cited on pages 176 and 199.
- Meehl, G., Covey, C., McAvaney, B., Latif, M., & Ronald, S. (2005). Overview of the coupled model intercomparison project. *Bulletin of the American Meteorological Society*(86), 89–93.
Cited on page 152.
- Meehl, G. A. (1995). Global coupled general circulation models. *Bulletin of the American Meteorological Society*. (Meeting Summary)
Cited on page 151.
- Meehl, G. A., Boer, G. J., Covey, C., Latif, M., & Stouffer, R. J. (1997). Intercomparison makes for a better climate model. *Eos, Transactions American Geophysical Union*, 78(41), 445–451.
Cited on page 151.
- Meehl, G. A., Covey, C., Delworth, T., Latif, M., McAvaney, B., Mitchell, J. F. B., ... Taylor, K. E. (2007). The WCRP CMIP3 multimodel dataset: A new era in climate change research. *Bulletin of the American Meteorological Society*, 88(9), 1383–1394.
Cited on pages 152 and 153.
- Meehl, G. A., Goddard, L., Boer, G., Burgman, R., Branstator, G., Cassou, C., ... Yeager, S. (2014). Decadal climate prediction: An update from the trenches. *Bulletin of the American Meteorological Society*, 95(2), 243–267. doi: 10.1175/BAMS-D-12-00241.1
Cited on page 155.
- Menzel, W. P., Schmit, T. J., Zhang, P., & Li, J. (2018, March). Satellite-Based Atmospheric Infrared Sounder Development and Applications. *Bulletin of the American Meteorological Society*, 99(3), 583–603. doi: 10.1175/BAMS-D-16-0293.1
Cited on page 197.
- Merriam-Webster. (2021). epistemology. In *Merriam-webster.com dictionary*. Retrieved 2021-05-15, from <https://www.merriam-webster.com/dictionary/adhoc>
Cited on page 69.
- Meyer, B., & Haerter, J. O. (2020). Mechanical forcing of convection by cold pools: Collisions and energy scaling. *Journal of Advances in Modeling Earth Systems*, 12(11), e2020MS002281. Retrieved from <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020MS002281> (e2020MS002281 10.1029/2020MS002281) doi: <https://doi.org/10.1029/2020MS002281>
Cited on page 176.

- Mignot, J., Hourdin, F., Deshayes, J., Boucher, O., Gastineau, G., Musat, I., ... others (2021). The tuning strategy of ipsl-cm6a-lr. *Journal of Advances in Modeling Earth Systems*, e2020MS002340.
Cited on pages 92, 99, 100, 111, 221, and 223.
- Mintz, Y. (1968). Very long-term global integration of the primitive equations of atmospheric motion: An experiment in climate simulation. In *Causes of climatic change* (pp. 20–36). Springer.
Cited on pages 69 and 70.
- Mitchell, J., Wilson, C., & Cunningham, W. (1987). On co2 climate sensitivity and model dependence of results. *Quarterly Journal of the Royal Meteorological Society*, 113(475), 293–322.
Cited on page 15.
- Mlawer, E., Clough, S., & Kato, S. (1998). Shortwave clear-sky model measurement intercomparison using rrtm. In *Proceedings of the eighth arm science team meeting* (pp. 23–27).
Cited on page 197.
- Mlawer, E. J., Payne, V. H., Moncet, J. L., Delamere, J. S., Alvarado, M. J., & Tobin, D. C. (2012, jun). Development and recent evaluation of the MT-CKD model of continuum absorption. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 370(1668), 2520–2556. doi: 10.1098/rsta.2011.0295
Cited on page 198.
- Molinari, J., & Dudek, M. (1992). Parameterization of convective precipitation in mesoscale numerical models: A critical review. *Monthly Weather Review*, 120(2), 326–344.
Cited on pages 76 and 89.
- Moncrieff, M. W. (1992). Organized convective systems: Archetypal dynamical models, mass and momentum flux theory, and parametrization. *Quarterly Journal of the Royal Meteorological Society*, 118(507), 819–850. doi: <https://doi.org/10.1002/qj.49711850703>
Cited on pages 61 and 192.
- Moody, D. E. (2012). Seven misconceptions regarding the gaia hypothesis. *Climatic Change*, 113(2), 277–284.
Cited on page 37.
- Mucchielli, A. (2000). *La nouvelle communication: épistémologie des sciences de l'information-communication*. Armand Colin.
Cited on page 20.
- Muller, C. J., & Held, I. M. (2012). Detailed investigation of the self-aggregation of convection in cloud-resolving simulations. *Journal of the Atmospheric Sciences*, 69(8), 2551–2565. doi: 10.1175/JAS-D-11-0257.1
Cited on pages 17 and 205.
- Muller, C. J., O’Gorman, P. A., & Back, L. E. (2011). Intensification of precipitation extremes with warming in a cloud-resolving model. *Journal of Climate*, 24(11), 2784–2800.
Cited on page 15.
- Murphy, J. (1995). Transient response of the hadley centre coupled ocean-atmosphere model to increasing carbon dioxide. part 1: control climate and flux adjustment. *Journal of Climate*,

8(1), 36–56.

Cited on pages 98 and 222.

Nadeau, R. (1999). *Vocabulaire technique et analytique de l'épistémologie* (PUF, Ed.).

Cited on page 20.

Nagarajan, B., & Aiyer, A. R. (2004). Performance of the ecmwf operational analyses over the tropical indian ocean. *Monthly weather review*, 132(9), 2275–2282. doi: [https://doi.org/10.1175/1520-0493\(2004\)132<2275:POTEOA>2.0.CO;2](https://doi.org/10.1175/1520-0493(2004)132<2275:POTEOA>2.0.CO;2)

Cited on page 210.

Nakicenovic, N., & Swart, R. (2000). *Emissions scenarios - special report of the intergovernmental panel on climate change*. Cambridge University Press.

Cited on page 152.

Nam, C., Bony, S., Dufresne, J.-L., & Chepfer, H. (2012). The ‘too few, too bright’ tropical low-cloud problem in cmip5 models. *Geophysical Research Letters*, 39(21).

Cited on page 106.

National Academy of Sciences. (1975). Survey of the climate simulation capability of global circulation models. *Understanding Climatic Change*, 196–239.

Cited on page 150.

Naumann, A. K., Stevens, B., & Hohenegger, C. (2019, 05). A Moist Conceptual Model for the Boundary Layer Structure and Radiatively Driven Shallow Circulations in the Trades. *Journal of the Atmospheric Sciences*, 76(5), 1289-1306.

Cited on pages 205 and 211.

Naumann, A. K., Stevens, B., Hohenegger, C., & Mellado, J. P. (2017). A conceptual model of a shallow circulation induced by prescribed low-level radiative cooling. *Journal of the Atmospheric Sciences*, 74(10), 3129-3144. doi: 10.1175/JAS-D-17-0030.1

Cited on pages 45, 46, 47, and 51.

Newton, C. W. (1950). Structure and mechanism of the prefrontal squall line. *Journal of meteorology*, 7(3), 210–222.

Cited on page 52.

Newton, C. W. (1963). Dynamics of severe convective storms. In *Severe local storms* (pp. 33–58). Springer.

Cited on pages 52 and 215.

Norton, S., & Suppe, F. (2001). Why atmospheric modeling is good science. *Changing the atmosphere: Expert knowledge and environmental governance*, 67–105.

Cited on page 21.

Notz, D., Haumann, F. A., Haak, H., Jungclaus, J. H., & Marotzke, J. (2013, 6). Arctic sea-ice evolution as modeled by Max Planck Institute for Meteorology's earth system model. *Journal of Advances in Modeling Earth Systems*, 5(2), 173–194. doi: 10.1002/jame.20016

Cited on pages 159 and 232.

Nuijens, L., Medeiros, B., Sandu, I., & Ahlgrimm, M. (2015). The behavior of trade-wind cloudiness

- in observations and models: The major cloud components and their variability. *Journal of Advances in Modeling Earth Systems*, 7(2), 600–616. doi: 10.1002/2014MS000390
Cited on page 199.
- O, K.-T., Wood, R., & Tseng, H.-H. (2018). Deeper, precipitating pbls associated with optically thin veil clouds in the sc-cu transition. *Geophysical Research Letters*, 45(10), 5177–5184. Retrieved from <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018GL077084> doi: 10.1029/2018GL077084
Cited on pages 202 and 207.
- O’Gorman, P. A., & Dwyer, J. G. (2018). Using machine learning to parameterize moist convection: Potential for modeling of climate, climate change, and extreme events. *Journal of Advances in Modeling Earth Systems*, 10(10), 2548–2563.
Cited on page 23.
- O’Gorman, P. A., & Schneider, T. (2009). The physical basis for increases in precipitation extremes in simulations of 21st-century climate change. *Proceedings of the National Academy of Sciences*, 106(35), 14773–14777.
Cited on page 15.
- Ooyama, K. (1964). A dynamical model for the study of tropical cyclone development. *Geofisica Internacional (Mexico)*, 4, 187–198.
Cited on pages 72 and 218.
- Oreopoulos, L., Mlawer, E., Delamere, J., Shippert, T., Cole, J., Fomin, B., . . . Rossow, W. B. (2012). The continual intercomparison of radiation codes: Results from phase I. *Journal of Geophysical Research: Atmospheres*, 117(D6). doi: 10.1029/2011JD016821
Cited on page 159.
- Oreskes, N., Shrader-Frechette, K., & Belitz, K. (1994). Verification, validation, and confirmation of numerical models in the earth sciences. *Science*, 263(5147), 641–646.
Cited on pages 25 and 48.
- Oueslati, B., & Bellon, G. (2015). The double itcz bias in cmip5 models: interaction between sst, large-scale circulation and precipitation. *Climate dynamics*, 44(3-4), 585–607.
Cited on pages 104 and 223.
- Palmer, T., & Stevens, B. (2019). The scientific challenge of understanding and estimating climate change. *Proceedings of the National Academy of Sciences*, 116(49), 24390–24395. doi: 10.1073/pnas.1906691116
Cited on pages 66, 161, and 217.
- Palmer, T. N. (2012). Towards the probabilistic Earth-system simulator: a vision for the future of climate and weather prediction. *Quarterly Journal of the Royal Meteorological Society*, 138(665), 841–861. doi: 10.1002/qj.1923
Cited on page 161.
- Park, S. (2014). A unified convection scheme (unicon). part i: Formulation. *Journal of the Atmospheric Sciences*, 71(11), 3902–3930.
Cited on pages 54 and 215.

- Parker, D. J. (1996). Cold pools in shear. *Quarterly Journal of the Royal Meteorological Society*, 122(535), 1655–1674.
Cited on page 53.
- Parker, W. S. (2006). Understanding pluralism in climate modeling. *Foundations of Science*, 11(4), 349–368.
Cited on page 66.
- Parker, W. S. (2008). Computer simulation through an error-statistical lens. *Synthese*, 163(3), 371–384.
Cited on page 48.
- Parker, W. S. (2009). Ii—confirmation and adequacy-for-purpose in climate modelling. In *Aristotelian society supplementary volume* (Vol. 83, pp. 233–249).
Cited on page 25.
- Parker, W. S. (2011). When climate models agree: The significance of robust model predictions. *Philosophy of Science*, 78(4), 579–600.
Cited on pages 92, 105, 149, 157, 221, and 224.
- Parker, W. S. (2013). Ensemble modeling, uncertainty and robust predictions. *WIREs Clim Change*, 4, 213–223.
Cited on page 26.
- Parker, W. S. (2014). Simulation and understanding in the study of weather and climate. *Perspectives on Science*, 22(3), 336–356.
Cited on page 24.
- Parker, W. S. (2018). The significance of robust climate projections. In *Climate modelling* (pp. 273–296). Springer.
Cited on page 26.
- Parker, W. S. (2020). Model evaluation: An adequacy-for-purpose view. *Philosophy of Science*, 87(3), 457–477.
Cited on page 25.
- Pawlowska, H., & Brenguier, J.-L. (2003). An observational study of drizzle formation in stratocumulus clouds for general circulation model (gcm) parameterizations. *Journal of Geophysical Research: Atmospheres*, 108(D15).
Cited on page 105.
- Penalba, O., & Rivera, J. (2013, 09). Future changes in drought characteristics over Southern South America projected by a CMIP5 multi-model ensemble. *American Journal of Climate Change*, 2, 173–182. doi: 10.4236/ajcc.2013.23017
Cited on page 155.
- Pennell, W., & LeMone, M. (1974). An experimental study of turbulence structure in the fair-weather trade wind boundary layer. *Journal of the Atmospheric Sciences*, 31(5), 1308–1323. doi: [https://doi.org/10.1175/1520-0469\(1974\)031<1308:AESOTS>2.0.CO;2](https://doi.org/10.1175/1520-0469(1974)031<1308:AESOTS>2.0.CO;2)
Cited on pages 59 and 178.

- Pergaud, J., Masson, V., Malardel, S., & Couvreur, F. (2009). A parameterization of dry thermals and shallow cumuli for mesoscale numerical weather prediction. *Boundary-layer meteorology*, 132(1), 83–106.
Cited on page 44.
- Petersen, A. C. (2012). *Simulating nature: A philosophical study of computer-simulation uncertainties and thier role in climate science and policy advice* (Chapman & C. Hall, Eds.).
Cited on pages 24, 27, 166, and 217.
- Phillips, N. A. (1954). Energy transformations and meridional circulations associated with simple baroclinic waves in a two-level, quasi-geostrophic model. *Tellus*, 6(3), 274–286.
Cited on page 68.
- Phillips, T. J., Potter, G. L., Williamson, D. L., Cederwall, R. T., Boyle, J. S., Fiorino, M., . . . Yio, J. J. (2004). Evaluating parameterizations in general circulation models: Climate simulation meets weather prediction. *Bulletin of the American Meteorological Society*, 85(12), 1903–1916.
Cited on page 49.
- Phillips, V. T., Donner, L. J., & Garner, S. T. (2007). Nucleation processes in deep convection simulated by a cloud-system-resolving model with double-moment bulk microphysics. *Journal of the atmospheric sciences*, 64(3), 738–761.
Cited on page 18.
- Pincus, R., Beljaars, A., Buehler, S. A., Kirchengast, G., Ladstaedter, F., & Whitaker, J. S. (2017, October). The Representation of Tropospheric Water Vapor Over Low-Latitude Oceans in (Re-)analysis: Errors, Impacts, and the Ability to Exploit Current and Prospective Observations. *Surveys in Geophysics*, 38(6), 1399–1423. doi: 10.1007/s10712-017-9437-z
Cited on page 197.
- Pincus, R., Mlawer, E. J., & Delamere, J. S. (2019). Balancing accuracy, efficiency, and flexibility in radiation calculations for dynamical models. *Journal of Advances in Modeling Earth Systems*, 11(10), 3074-3089. doi: 10.1029/2019MS001621
Cited on pages 159 and 198.
- Pincus, R., Mlawer, E. J., Oreopoulos, L., Ackerman, A. S., Baek, S., Brath, M., . . . Schwarzkopf, D. M. (2015). Radiative flux and forcing parameterization error in aerosol-free clear skies. *Geophysical Research Letters*, 42(13), 5485–5492. doi: 10.1002/2015GL064291
Cited on page 159.
- Pincus, R., & Stevens, B. (2013). Paths to accuracy for radiation parameterizations in atmospheric models. *Journal of Advances in Modeling Earth Systems*, 5(2), 225–233. doi: 10.1002/jame.20027
Cited on page 159.
- Pirtle, Z., Meyer, R., & Hemilton, A. (2010). What does it mean when climate models agree? A case for assessing independence among general circulation models. *Environmental Science and Policy*.
Cited on pages 26 and 156.
- Pitt, J. (2000). *Thinking about technology: Foundations of the philosophy of technology* (N. Y. S. B. Press, Ed.).

Cited on pages 149, 160, and 232.

Plant, R., & Craig, G. C. (2008). A stochastic parameterization for deep convection based on equilibrium statistics. *Journal of the Atmospheric Sciences*, 65(1), 87–105.
Cited on page 78.

Plant, R. S., & Yano, J.-I. (2016). Quasi-equilibrium. In R. S. Plant & J.-I. Yano (Eds.), *Parameterization of atmospheric convection* (p. 101-146). World Scientific.
Cited on pages 76 and 78.

Popper, K. R. (1982). Science: Conjecture and refutations. *Philosophy of science and the occult*, 104–111.
Cited on page 70.

Potter, G. L., & Gates, W. L. (1984). A preliminary intercomparison of the seasonal response of two atmospheric climate models. *Monthly Weather Review*, 112(5), 909-917.
Cited on page 150.

Prandtl, L. (1952). *Essentials of fluid dynamics, with application to hydraulics, aeronautics, meteorology and other subjects*. Blackie.
Cited on page 53.

Prein, A. F., Langhans, W., Fosser, G., Ferrone, A., Ban, N., Goergen, K., ... others (2015). A review on regional convection-permitting climate modeling: Demonstrations, prospects, and challenges. *Reviews of geophysics*, 53(2), 323–361.
Cited on page 18.

Pritchard, M. S., & Somerville, R. C. (2009). Empirical orthogonal function analysis of the diurnal cycle of precipitation in a multi-scale climate model. *Geophysical Research Letters*, 36(5).
Cited on page 139.

Provod, M., Marsham, J., Parker, D., & Birch, C. (2016). A characterization of cold pools in the west african sahel. *Monthly Weather Review*, 144(5), 1923–1934.
Cited on page 54.

Purdom, J. F. (1976). Some uses of high-resolution GOES imagery in the mesoscale forecasting of convection and its behavior. *Monthly Weather Review*, 104(12), 1474–1483. doi: [https://doi.org/10.1175/1520-0493\(1976\)104<1474:SUOHRG>2.0.CO;2](https://doi.org/10.1175/1520-0493(1976)104<1474:SUOHRG>2.0.CO;2)
Cited on page 183.

Qian, L., Young, G. S., & Frank, W. M. (1998). A convective wake parameterization scheme for use in general circulation models. *Monthly weather review*, 126(2), 456–469.
Cited on pages 54 and 215.

Qu, X., & Hall, A. (2014, Jan 01). On the persistent spread in snow-albedo feedback. *Climate Dynamics*, 42(1), 69–81. doi: 10.1007/s00382-013-1774-0
Cited on page 158.

Raïsaänen, J. (2007). How reliable are climate models? *Tellus A: Dynamic Meteorology and Oceanography*, 59(1), 2–29.
Cited on page 21.

- Ramaswamy, V., Schwarzkopf, M., Randel, W., Santer, B., Soden, B. J., & Stenchikov, G. (2006). Anthropogenic and natural influences in the evolution of lower stratospheric cooling. *Science*, *311*(5764), 1138–1141.
Cited on page 67.
- Randall, D., Khairoutdinov, M., Arakawa, A., & Grabowski, W. (2003). Breaking the cloud parameterization deadlock. *Bulletin of the American Meteorological Society*, *84*(11), 1547–1564.
Cited on pages 19, 22, 66, 110, 213, and 217.
- Randall, D. A. (2013). Beyond deadlock. *Geophysical Research Letters*, *40*(22), 5970–5976.
Cited on pages 66 and 213.
- Randall, D. A., Dazlich, D. A., & Corsetti, T. G. (1989). Interactions among radiation, convection, and large-scale dynamics in a general circulation model. *Journal of Atmospheric Sciences*, *46*(13), 1943–1970.
Cited on page 48.
- Randall, D. A., & Tjemkes, S. (1991). Clouds, the earth’s radiation budget, and the hydrologic cycle. *Global and Planetary Change*, *4*(1-3), 3–9. doi: 10.1016/0921-8181(91)90063-3
Cited on page 202.
- Randall, D. A., & Wielicki, B. A. (1997). Measurements, models, and hypotheses in the atmospheric sciences. *Bulletin of the American Meteorological Society*, *78*(3), 399–406.
Cited on pages 21, 27, 92, 101, 104, 221, and 224.
- Randall, D. A., Xu, K.-M., Somerville, R. J., & Iacobellis, S. (1996). Single-column models and cloud ensemble models as links between observations and climate models. *Journal of Climate*, *9*(8), 1683–1697.
Cited on page 100.
- Randel, W. J., Shine, K. P., Austin, J., Barnett, J., Claud, C., Gillett, N. P., ... others (2009). An update of observed stratospheric temperature trends. *Journal of Geophysical Research: Atmospheres*, *114*(D2).
Cited on page 67.
- Rasch, P., & Kristjánsson, J. (1998). A comparison of the ccm3 model climate using diagnosed and predicted condensate parameterizations. *Journal of Climate*, *11*(7), 1587–1614.
Cited on page 141.
- Rauber, R. M., Stevens, B., Ochs III, H. T., Knight, C., Albrecht, B. A., Blyth, A., ... others (2007). Rain in shallow cumulus over the ocean: The rico campaign. *Bulletin of the American Meteorological Society*, *88*(12), 1912–1928.
Cited on pages 20, 117, 176, and 228.
- Redelsperger, J.-L., & Lafore, J.-P. (1988). A three-dimensional simulation of a tropical squall line: Convective organization and thermodynamic vertical transport. *Journal of the atmospheric sciences*, *45*(8), 1334–1356.
Cited on page 53.
- Richter, I., & Xie, S.-P. (2008). Muted precipitation increase in global warming simulations: A surface

- evaporation perspective. *Journal of Geophysical Research: Atmospheres*, 113(D24).
Cited on page 15.
- Riehl, H., & Malkus, J. (1958). On the heat balance of the equatorial trough zone. *Geophysica*, 6, 503–538.
Cited on pages 71, 72, and 218.
- Riehl, H., Yeh, T., Malkus, J. S., & La Seur, N. E. (1951). The north-east trade of the pacific ocean. *Quarterly Journal of the Royal Meteorological Society*, 77(334), 598–626.
Cited on page 196.
- Rio, C., Del Genio, A. D., & Hourdin, F. (2019). Ongoing breakthroughs in convective parameterization. *Curr. Clim. Change Rep.*, 5(2), 95–111. doi: 10.1007/s40641-019-00127-w
Cited on pages 18, 22, 23, 35, 66, 168, and 217.
- Rio, C., & Hourdin, F. (2008). A thermal plume model for the convective boundary layer: Representation of cumulus clouds. *Journal of the atmospheric sciences*, 65(2), 407–425.
Cited on pages 82, 83, 94, and 95.
- Rio, C., Hourdin, F., Couvreux, F., & Jam, A. (2010). Resolved versus parametrized boundary-layer plumes. part II: Continuous formulations of mixing rates for mass-flux schemes. *Boundary-Layer Meteorology*, 135(3), 469–483. doi: 10.1007/s10546-010-9478-z
Cited on page 161.
- Rio, C., Hourdin, F., Grandpeix, J.-Y., & Lafore, J.-P. (2009). Shifting the diurnal cycle of parameterized deep convection over land. *Geophysical Research Letters*, 36(7).
Cited on pages 117 and 216.
- Rochetin, N., Couvreux, F., Grandpeix, J.-Y., & Rio, C. (2014). Deep convection triggering by boundary layer thermals. part I: Les analysis and stochastic triggering formulation. *Journal of the Atmospheric Sciences*, 71(2), 496–514. doi: 10.1175/JAS-D-12-0336.1
Cited on pages 83, 84, and 161.
- Rochetin, N., Hohenegger, C., Touzé-Peiffer, L., & Villefranque, N. (2021). A physically-based robust definition of convectively generated density currents: detection and characterization in convection-permitting simulations. *Journal of Advances in Modeling Earth Systems*, 88(XXX), 117–135.
Cited on pages 29, 56, 57, 58, 59, 60, 62, 165, 178, 214, and 216.
- Romps, D. M. (2010). A direct measure of entrainment. *Journal of the Atmospheric Sciences*, 67(6), 1908–1927.
Cited on pages 97, 98, and 106.
- Romps, D. M., & Jeevanjee, N. (2016). On the sizes and lifetimes of cold pools. *Quarterly Journal of the Royal Meteorological Society*, 142(696), 1517–1527.
Cited on page 56.
- Rotunno, R., Klemp, J. B., & Weisman, M. L. (1988). A theory for strong, long-lived squall lines. *Journal of the Atmospheric Sciences*, 45(3), 463–485.
Cited on pages 52, 53, and 176.

- Ruppert, J. H., & Johnson, R. H. (2016). On the cumulus diurnal cycle over the tropical warm pool. *Journal of Advances in Modeling Earth Systems*, 8(2), 669–690. doi: 10.1002/2015MS000610
Cited on page 202.
- Ruppert Jr, J. H., & O’Neill, M. E. (2019). Diurnal cloud and circulation changes in simulated tropical cyclones. *Geophysical Research Letters*, 46(1), 502–511. doi: 10.1029/2018GL081302
Cited on page 211.
- Sandu, I., & Stevens, B. (2011). On the factors modulating the stratocumulus to cumulus transitions. *Journal of the Atmospheric Sciences*, 68(9), 1865–1881.
Cited on pages 126 and 228.
- Sarewitz, D. R., & Byerly, R. (2000). *Prediction: science, decision making, and the future of nature*. Island Press.
Cited on page 171.
- Satoh, M., Stevens, B., Judt, F., Khairoutdinov, M., Lin, S.-J., Putman, W. M., & Düben, P. (2019). Global cloud-resolving models. *Current Climate Change Reports*, 5(3), 172–184.
Cited on pages 18, 49, and 66.
- Schlemmer, L., & Hohenegger, C. (2014). The formation of wider and deeper clouds as a result of cold-pool dynamics. *Journal of the Atmospheric Sciences*, 71(8), 2842–2858.
Cited on pages 52, 176, and 215.
- Schlesinger, R. E. (1973). A numerical model of deep moist convection: Part i. comparative experiments for variable ambient moisture and wind shear. *Journal of the Atmospheric Sciences*, 30(5), 835–856.
Cited on page 53.
- Schmidt, G. A., Bader, D., Donner, L. J., Elsaesser, G. S., Golaz, J.-C., Hannay, C., ... Saha, S. (2017). Practice and philosophy of climate model tuning across six us modeling centers. *Geoscientific Model Development*, 10(9), 3207–3223.
Cited on pages 22, 99, 100, and 103.
- Schmidt, G. A., Kelley, M., Nazarenko, L., Ruedy, R., Russell, G. L., Aleinov, I., ... others (2014). Configuration and assessment of the giss modele2 contributions to the cmip5 archive. *Journal of Advances in Modeling Earth Systems*, 6(1), 141–184.
Cited on page 104.
- Schmidt, O. T., Mengaldo, G., Balsamo, G., & Wedi, N. P. (2019). Spectral empirical orthogonal function analysis of weather and climate data. *Monthly Weather Review*, 147(8), 2979–2995.
Cited on page 139.
- Schmit, T. J., Li, J., Ackerman, S. A., & Gurka, J. J. (2009, November). High-Spectral- and High-Temporal-Resolution Infrared Measurements from Geostationary Orbit. *Journal of Atmospheric and Oceanic Technology*, 26(11), 2273–2292. doi: 10.1175/2009JTECHA1248.1
Cited on page 197.
- Schneider, T., Lan, S., Stuart, A., & Teixeira, J. (2017). Earth system modeling 2.0: A blueprint for models that learn from observations and targeted high-resolution simulations. *Geophysical Research Letters*, 44(24), 12–396.

Cited on page 23.

Schneider, T., Teixeira, J., Bretherton, C. S., Brient, F., Pressel, K. G., Schär, C., & Siebesma, A. P. (2017). Climate goals and computing the future of clouds. *Nature Climate Change*, 7(1), 3–5.
Cited on page 166.

Schubert, W. H. (2000). A retrospective view of arakawa’s ideas on cumulus parameterization. In *International geophysics* (Vol. 70, pp. 181–198). Elsevier.
Cited on pages 73, 74, 75, 76, 87, and 219.

Seifert, A., & Heus, T. (2013a). Large-eddy simulation of organized precipitating trade wind cumulus clouds. *Atmospheric Chemistry and Physics*, 13, 5631–5645. doi: <https://doi.org/10.5194/acp-13-5631-2013>
Cited on page 183.

Seifert, A., & Heus, T. (2013b). Large-eddy simulation of organized precipitating trade wind cumulus clouds. *Atmospheric Chemistry and Physics*, 13, 5631–5645. doi: 10.5194/acpd-13-1855-2013
Cited on page 211.

Seifert, A., Heus, T., Pincus, R., & Stevens, B. (2015). Large-eddy simulation of the transient and near-equilibrium behavior of precipitating shallow convection. *Journal of Advances in Modeling Earth Systems*, 7(4), 1918–1937. doi: 10.1002/2015MS000489
Cited on page 197.

Shackley, S. (2001). Epistemic lifestyles in climate change modeling. *Changing the atmosphere: Expert knowledge and environmental governance*, 107–33.
Cited on page 26.

Shackley, S., Risbey, J., Stone, P., & Wynne, B. (1999). Adjusting to policy expectations in climate change modeling. *Climatic Change*, 43(2), 413–454.
Cited on page 98.

Shackley, S., Young, P., Parkinson, S., & Wynne, B. (1998). Uncertainty, complexity and concepts of good science in climate change modelling: Are GCMs the best tools? *Climatic Change*, 38, 159–205.
Cited on pages 149, 160, and 163.

Sherwood, S., Roca, R., Weckwerth, T., & Andronova, N. (2010). Tropospheric water vapor, convection, and climate. *Reviews of Geophysics*, 48(2).
Cited on page 48.

Sherwood, S., Webb, M. J., Annan, J. D., Armour, K., Forster, P. M., Hargreaves, J. C., ... others (2020). An assessment of earth’s climate sensitivity using multiple lines of evidence. *Reviews of Geophysics*, 58(4), e2019RG000678.
Cited on page 15.

Sherwood, S. C., Bony, S., & Dufresne, J.-L. (2014). Spread in model climate sensitivity traced to atmospheric convective mixing. *Nature*, 505(7481), 37–42. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/24380952> doi: 10.1038/nature12829
Cited on page 196.

- Shukla, J., Hagedorn, R., Miller, M., Palmer, T. N., Hoskins, B., Kinter, J., ... Slingo, J. (2009). Strategies: Revolution in climate prediction is both necessary and possible: A declaration at the world modelling summit for climate prediction. *Bulletin of the American Meteorological Society*, 90(2), 175–178. doi: 10.1175/2008BAMS2759.1
Cited on page 161.
- Skamarock, W. C., Weisman, M. L., & Klemp, J. B. (1994). Three-dimensional evolution of simulated long-lived squall lines. *Journal of Atmospheric Sciences*, 51(17), 2563–2584.
Cited on page 53.
- Smagorinsky, J., Manabe, S., & Holloway Jr, J. L. (1965). Numerical results from a nine-level general circulation model of the atmosphere. *Monthly weather review*, 93(12), 727–768.
Cited on pages 69, 70, and 218.
- Soler, L., Wieber, F., Allamel-Raffin, C., Gangloff, J.-L., Dufour, C., & Trizio, E. (2013). Calibration: A conceptual framework applied to scientific practices which investigate natural phenomena by means of standardized instruments. *Journal for general philosophy of science*, 44(2), 263–317.
Cited on page 102.
- Srivastava, A. K., & DelSole, T. (2014). Robust forced response in South Asian summer monsoon in a future climate. *Journal of Climate*, 27(20), 7849–7860. doi: 10.1175/JCLI-D-13-00599.1
Cited on page 155.
- Stainforth, D. A., Downing, T. E., Washington, R., Lopez, A., & New, M. (2007). Issues in the interpretation of climate model ensembles to inform decisions. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 365(1857), 2163–2177. doi: 10.1098/rsta.2007.2073
Cited on pages 92, 156, and 221.
- Stephan, C. C., Schnitt, S., Schulz, H., Bellenger, H., Szoeké, S. P. d., Acquistapace, C., ... others (2020). Ship-and island-based atmospheric soundings from the 2020 eurec4a field campaign. *Earth System Science Data Discussions*, 1–35. doi: 10.5194/essd-2020-174
Cited on pages 178, 194, 197, and 198.
- Stephens, G. L., Li, J., Wild, M., Clayson, C. A., Loeb, N., Kato, S., ... Andrews, T. (2012). An update on earth’s energy balance in light of the latest global observations. *Nature Geoscience*, 5(10), 691–696. doi: 10.1038/ngeo1580
Cited on pages 103, 197, and 223.
- Stevens, B. (2002). Entrainment in stratocumulus-topped mixed layers. *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography*, 128(586), 2663–2690.
Cited on page 89.
- Stevens, B. (2006). Bulk boundary-layer concepts for simplified models of tropical dynamics. *Theoretical and Computational Fluid Dynamics*, 20(5), 279–304.
Cited on page 45.
- Stevens, B., Acquistapace, C., Hansen, A., Heinze, R., Klinger, C., Klocke, D., ... others (2020). The added value of large-eddy and storm-resolving models for simulating clouds and precipitation. *Journal of the Meteorological Society of Japan. Ser. II*.

Cited on pages 18, 22, 23, 49, and 66.

Stevens, B., & Bony, S. (2013). What are climate models missing? *Science*, *340*(6136), 1053–1054.
Cited on pages 66 and 217.

Stevens, B., Bony, S., Brogniez, H., Hentgen, L., Hohenegger, C., Kiemle, C., ... others (2019). Sugar, gravel, fish and flowers: Mesoscale cloud patterns in the trade winds. *Quarterly Journal of the Royal Meteorological Society*, 1–12.
Cited on pages 38, 39, 64, and 185.

Stevens, B., Bony, S., Brogniez, H., Hentgen, L., Hohenegger, C., Kiemle, C., ... Zuidema, P. (2020). Sugar, gravel, fish and flowers: Mesoscale cloud patterns in the trade winds. *Quarterly Journal of the Royal Meteorological Society*, *146*(726), 141–152. doi: 10.1002/qj.3662
Cited on pages 197 and 205.

Stevens, B., Bony, S., Farrell, D., Ament, F., Blyth, A., Fairall, C., ... Zöger, M. (2021). Eurec⁴a. *Earth System Science Data Discussions*, *2021*, 1–78. Retrieved from <https://essd.copernicus.org/preprints/essd-2021-18/> doi: 10.5194/essd-2021-18
Cited on pages 29, 39, 59, 165, 177, 180, 198, 203, and 214.

Stevens, B., Brogniez, H., Kiemle, C., Lacour, J.-L., Crevoisier, C., & Kiliani, J. (2017). Structure and dynamical influence of water vapor in the lower tropical troposphere. *Surveys in Geophysics*, *38*(6), 1371–1397. doi: 10.1007/s10712-017-9420-8
Cited on pages 196, 202, and 207.

Stevens, B., Farrell, D., Hirsch, L., Jansen, F., Nuijens, L., Serikov, I., ... Prospero, J. M. (2016). The barbados cloud observatory: Anchoring investigations of clouds and circulation on the edge of the itcz. *Bulletin of the American Meteorological Society*, *97*(5), 787–801. doi: 10.1175/BAMS-D-14-00247.1
Cited on pages 176, 177, 182, 189, and 194.

Stevens, B., & Lenschow, D. H. (2001). Observations, experiments, and large eddy simulation. *Bulletin of the American Meteorological Society*, *82*(2), 283–294.
Cited on pages 21 and 103.

Stevens, B., Satoh, M., Auger, L., Biercamp, J., Bretherton, C. S., Chen, X., ... others (2019). Dyamond: the dynamics of the atmospheric general circulation modeled on non-hydrostatic domains. *Progress in Earth and Planetary Science*, *6*(1), 1–17.
Cited on pages 49 and 169.

Stevens, B., & Schwartz, S. E. (2012). Observing and modeling earth’s energy flows. *Surveys in geophysics*, *33*(3), 779–816.
Cited on pages 14 and 37.

Stevenson, S. L. (2012). Significant changes to ENSO strength and impacts in the twenty-first century: Results from CMIP5. *Geophysical Research Letters*, *39*(17). doi: 10.1029/2012GL052759
Cited on page 155.

Stocker, T. F., Qin, D., Plattner, G.-K., Alexander, L. V., Allen, S. K., Bindoff, N. L., ... others (2013). Technical summary. In *Climate change 2013: the physical science basis. contribution of working group I to the fifth assessment report of the intergovernmental panel on climate change*

- (pp. 33–115). Cambridge University Press.
Cited on pages 15 and 67.
- Stone, P. H., & Risbey, J. S. (1990). On the limitations of general circulation climate models. *Geophysical Research Letters*, 17(12), 2173–2176.
Cited on pages 28, 66, and 217.
- Stouffer, R. J., Eyring, V., Meehl, G. A., Bony, S., Senior, C., Stevens, B., & Taylor, K. (2017). Cmp5 scientific gaps and recommendations for cmp6. *Bulletin of the American Meteorological Society*, 98(1), 95–105.
Cited on page 167.
- Sui, C., Lau, K., Tao, W., & Simpson, J. (1994). The tropical water and energy cycles in a cumulus ensemble model. part i: Equilibrium climate. *Journal of Atmospheric Sciences*, 51(5), 711–728.
Cited on page 18.
- Sundberg, M. (2007). Parameterizations as boundary objects on the climate arena. *Social Studies of Science*, 37(3), 473–488.
Cited on page 26.
- Sundqvist, H. (1988). Parameterization of condensation and associated clouds in models for weather prediction and general circulation simulation. In *Physically-based modelling and simulation of climate and climatic change* (pp. 433–461). Springer.
Cited on pages 94, 118, and 119.
- Suzuki, K., Golaz, J.-C., & Stephens, G. L. (2013). Evaluating cloud tuning in a climate model with satellite observations. *Geophysical Research Letters*, 40(16), 4464–4468.
Cited on page 103.
- Suzuki, K., Stephens, G. L., & Lebsock, M. D. (2013). Aerosol effect on the warm rain formation process: Satellite observations and modeling. *Journal of Geophysical Research: Atmospheres*, 118(1), 170–184.
Cited on page 105.
- Takeda, T. (1971). Numerical simulation of a precipitating convective cloud: the formation of a “long-lasting” cloud. *Journal of the Atmospheric Sciences*, 28(3), 350–376.
Cited on page 53.
- Taschetto, A. S., Gupta, A. S., Jourdain, N. C., Santoso, A., Ummenhofer, C. C., & England, M. H. (2014). Cold tongue and warm pool ENSO events in CMIP5: Mean state and future projections. *Journal of Climate*, 27(8), 2861–2885. doi: 10.1175/JCLI-D-13-00437.1
Cited on page 155.
- Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, 93(4), 485–498.
Cited on page 153.
- Tebaldi, C., & Knutti, R. (2007). The use of the multi-model ensemble in probabilistic climate projections. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 365(1857), 2053–2075. doi: 10.1098/rsta.2007.2076
Cited on pages 156, 157, 158, and 231.

- Thompson, D. W. J., Bony, S., & Li, Y. (2017). Thermodynamic constraint on the depth of the global tropospheric circulation. *Proceedings of the National Academy of Sciences*, 114(31), 8181–8186. doi: 10.1073/pnas.1620493114
Cited on page 211.
- Tian, B., & Dong, X. (2020). The double-itez bias in cmip3, cmip5, and cmip6 models based on annual mean precipitation. *Geophysical Research Letters*, 47(8), e2020GL087232.
Cited on page 168.
- Tompkins, A. M. (2001). Organization of tropical convection in low vertical wind shears: The role of cold pools. *Journal of the atmospheric sciences*, 58(13), 1650–1672.
Cited on pages 45, 52, 53, 176, and 215.
- Torri, G., Kuang, Z., & Tian, Y. (2015). Mechanisms for convection triggering by cold pools. *Geophysical Research Letters*, 42(6), 1943–1950.
Cited on pages 45, 55, 56, 81, and 216.
- Touzé-Peiffer, L., Barberousse, A., & Le Treut, H. (2020). The coupled model intercomparison project: History, uses, and structural effects on climate research. *Wiley Interdisciplinary Reviews: Climate Change*, 11(4), e648.
Cited on page 155.
- Touzé-Peiffer, L., Vogel, R., & Rochetin, N. (2021). Detecting cold pools from soundings during eurec⁴a. *Journal of applied meteorology and climatology*, 11(10), 3148–3166.
Cited on pages 29, 59, 62, 165, 214, and 217.
- Trenberth, K. E. (2011). Changes in precipitation with climate change. *Climate Research*, 47(1-2), 123–138.
Cited on page 15.
- Vaisala. (2020). *Rd41 technical data* (Tech. Rep.). <https://www.vaisala.com/sites/default/files/documents/{RD41}-Datasheet-{B211706EN}.pdf>: University Corporation for Atmospheric Research.
Cited on page 210.
- Van der Dussen, J., De Roode, S., Ackerman, A. S., Blossey, P. N., Bretherton, C. S., Kurowski, M. J., ... Siebesma, A. (2013). The gass/euclipse model intercomparison of the stratocumulus transition as observed during astex: Les results. *Journal of Advances in Modeling Earth Systems*, 5(3), 483–499.
Cited on page 127.
- VanZanten, M. C., Stevens, B., Nuijens, L., Siebesma, A. P., Ackerman, A., Burnet, F., ... others (2011). Controls on precipitation and cloudiness in simulations of trade-wind cumulus as observed during rico. *Journal of Advances in Modeling Earth Systems*, 3(2).
Cited on pages 118 and 127.
- Vial, J., Bony, S., Stevens, B., & Vogel, R. (2017). Mechanisms and model diversity of trade-wind shallow cumulus cloud feedbacks: a review. In *Shallow clouds, water vapor, circulation, and climate sensitivity* (pp. 159–181). Springer. doi: 10.1007/978-3-319-77273-8_8
Cited on page 103.

- Vial, J., Dufresne, J.-L., & Bony, S. (2013). On the interpretation of inter-model spread in cmip5 climate sensitivity estimates. *Climate Dynamics*, 41(11), 3339–3362. doi: 10.1007/s00382-013-1725-9
Cited on page 196.
- Vial, J., Vogel, R., Bony, S., Stevens, B., Winker, D. M., Cai, X., ... Brogniez, H. (2019a). A new look at the daily cycle of trade wind cumuli. *Journal of advances in modeling earth systems*, 11(10), 3148–3166.
Cited on pages 58 and 180.
- Vial, J., Vogel, R., Bony, S., Stevens, B., Winker, D. M., Cai, X., ... Brogniez, H. (2019b). A new look at the daily cycle of trade wind cumuli. *Journal of advances in modeling earth systems*, 11(10), 3148–3166. doi: 10.1029/2019MS001746
Cited on pages 202 and 211.
- Vial, J., Vogel, R., & Schulz, H. (2021). On the daily cycle of mesoscale cloud organization in the winter trades. *Quarterly Journal of the Royal Meteorological Society*.
Cited on page 39.
- Vinnichenko, N. (1970). The kinetic energy spectrum in the free atmosphere—1 second to 5 years. *Tellus*, 22(2), 158–166.
Cited on pages 22, 76, and 89.
- Vogel, R. (2017). *The influence of precipitation and convective organization on the structure of the trades* (Doctoral dissertation, Universität Hamburg Hamburg). doi: <https://doi.org/10.17617/2.2503092>
Cited on pages 54 and 176.
- Vogel, R., Bony, S., & Stevens, B. (2020). Estimating the shallow convective mass flux from the subcloud-layer mass budget. *Journal of Atmospheric Sciences*, 77(5), 1559–1574.
Cited on page 43.
- Voigt, A., Alber, N., Ceppi, P., Grise, K., Li, Y., & Medeiros, B. (2021). Clouds, radiation, and atmospheric circulation in the present-day climate and under climate change. *Wiley Interdisciplinary Reviews: Climate Change*, 12(2), e694.
Cited on page 67.
- Von Karman, T., et al. (1940). The engineer grapples with nonlinear problems. *Bulletin of the American Mathematical Society*, 46(8), 615–683.
Cited on page 85.
- Warner, C., Simpson, J., Martin, D., Suchman, D., Mosher, F., & Reinking, R. (1979). Shallow convection on day 261 of gate/mesoscale arcs. *Monthly Weather Review*, 107(12), 1617–1635.
Cited on pages 39, 41, 52, 176, and 215.
- Washington, W. M., Semtner, A. J., Meehl, G. A., Knight, D. J., & Mayer, T. A. (1980). A general circulation experiment with a coupled atmosphere, ocean and sea ice model. *Journal of Physical Oceanography*, 10(12), 1887–1908. doi: 10.1175/1520-0485(1980)010<1887:AGCEWA>2.0.CO;2
Cited on pages 161 and 232.

- Webb, M., Senior, C., Bony, S., & Morcrette, J.-J. (2001). Combining erbe and isccp data to assess clouds in the hadley centre, ecmwf and lmd atmospheric climate models. *Climate Dynamics*, 17(12), 905–922.
Cited on page 106.
- Weber, S. L. (2010). The utility of earth system models of intermediate complexity (emics). *Wiley Interdisciplinary Reviews: Climate Change*, 1(2), 243–252.
Cited on page 49.
- Webster, P. J., & Lukas, R. (1992). Toga coare: The coupled ocean–atmosphere response experiment. *Bulletin of the American Meteorological Society*, 73(9), 1377–1416.
Cited on pages 54 and 176.
- Weisman, M. L., & Rotunno, R. (2004). “a theory for strong long-lived squall lines” revisited. *Journal of the Atmospheric Sciences*, 61(4), 361–382. doi: [https://doi.org/10.1175/1520-0469\(2004\)061<0361:ATFSLS>2.0.CO;2](https://doi.org/10.1175/1520-0469(2004)061<0361:ATFSLS>2.0.CO;2)
Cited on pages 176 and 183.
- Weisman, M. L., Skamarock, W. C., & Klemp, J. B. (1997). The resolution dependence of explicitly modeled convective systems. *Monthly Weather Review*, 125(4), 527–548.
Cited on page 18.
- WGCM. (2006). Report of the tenth session of the JSC/CLIVAR working group on coupled modelling [WCRP Informal Report No.5/2007].
Cited on page 153.
- WGCM. (2012). Report of the sixteenth session of the working group on coupled modelling [WCRP Report No.2/2013].
Cited on page 153.
- WGCM. (2014). Report of the eighteenth session of the working group on coupled modelling [WCRP Report No.2/2015].
Cited on page 153.
- Williamson, D., Blaker, A. T., Hampton, C., & Salter, J. (2015). Identifying and removing structural biases in climate models with history matching. *Climate dynamics*, 45(5-6), 1299–1324.
Cited on pages 107 and 130.
- Williamson, D., Goldstein, M., Allison, L., Blaker, A., Challenor, P., Jackson, L., & Yamazaki, K. (2013). History matching for exploring and reducing climate model parameter space using observations and a large perturbed physics ensemble. *Climate dynamics*, 41(7-8), 1703–1729.
Cited on pages 130 and 131.
- Williamson, D. B., Blaker, A. T., & Sinha, B. (2017). Tuning without over-tuning: parametric uncertainty quantification for the nemo ocean model. *Geoscientific Model Development*, 10(4), 1789–1816.
Cited on pages 130 and 131.
- Winsberg, E. (2009). Computer simulation and the philosophy of science. *Philosophy Compass*, 4(5), 835–845.
Cited on page 48.

- Winsberg, E. (2018a). *Philosophy and climate science*. Cambridge University Press. doi: 10.1017/9781108164290
Cited on pages 106 and 157.
- Winsberg, E. (2018b, 11). What does robustness teach us in climate science: a re-appraisal. *Synthese*. doi: 10.1007/s11229-018-01997-7
Cited on pages 24, 106, and 225.
- Wood, R. (2012). Stratocumulus Clouds. *Monthly Weather Review*, 140(8), 2373–2423. doi: 10.1175/MWR-D-11-00121.1
Cited on page 196.
- Wood, R., O, K.-T., Bretherton, C. S., Mohrmann, J., Albrecht, B. A., Zuidema, P., ... others (2018). Ultraclean layers and optically thin clouds in the stratocumulus-to-cumulus transition. part i: Observations. *Journal of the Atmospheric Sciences*, 75(5), 1631–1652. doi: 10.1175/JAS-D-17-0213.1
Cited on pages 202 and 207.
- Wood, R., Tseng, H.-H., et al. (2018). Deeper, precipitating pbls associated with optically thin veil clouds in the sc-cu transition. *GeoRL*, 45(10), 5177–5184. doi: 10.1029/2018GL077084
Cited on pages 202 and 207.
- Xie, S.-P., Deser, C., Vecchi, G. A., Ma, J., Teng, H., & Wittenberg, A. T. (2010). Global warming pattern formation: Sea surface temperature and rainfall. *Journal of Climate*, 23(4), 966–986.
Cited on page 15.
- Xu, K.-M., Arakawa, A., & Krueger, S. K. (1992). The macroscopic behavior of cumulus ensembles simulated by a cumulus ensemble model. *Journal of Atmospheric Sciences*, 49(24), 2402–2420.
Cited on page 18.
- Xu, Q. (1992). Density currents in shear flows-a two-fluid model. *Journal of the atmospheric sciences*, 49(6), 511–524.
Cited on page 53.
- Yano, J.-I. (2014). Formulation structure of the mass-flux convection parameterization. *Dynamics of Atmospheres and Oceans*, 67, 1–28.
Cited on pages 79, 80, 81, 87, and 219.
- Yano, J.-I. (2016). Scale separation. In R. S. Plant & J.-I. Yano (Eds.), *Parameterization of atmospheric convection* (p. 101-146). World Scientific.
Cited on page 89.
- Yano, J.-I., Bengtsson, L., Geleyn, J.-F., & Brozkova, R. (2016). Towards a unified and self-consistent parameterization framework. In R. S. Plant & J.-I. Yano (Eds.), *Parameterization of atmospheric convection* (p. 423-436). World Scientific.
Cited on page 81.
- Yoneyama, K., Zhang, C., & Long, C. N. (2013). Tracking pulses of the madden–julian oscillation. *Bulletin of the American Meteorological Society*, 94(12), 1871–1891.
Cited on page 176.

- Yoo, C., Park, S., Kim, D., Yoon, J.-H., & Kim, H.-M. (2015). Boreal winter mjo teleconnection in the community atmosphere model version 5 with the unified convection parameterization. *Journal of Climate*, 28(20), 8135–8150.
Cited on page 141.
- Young, G. S., Perugini, S. M., & Fairall, C. W. (1995). Convective wakes in the equatorial western pacific during toga. *Monthly Weather Review*, 123(1), 110-123. Retrieved from [https://doi.org/10.1175/1520-0493\(1995\)123<0110:CWITEW>2.0.CO;2](https://doi.org/10.1175/1520-0493(1995)123<0110:CWITEW>2.0.CO;2) doi: 10.1175/1520-0493(1995)123<0110:CWITEW>2.0.CO;2
Cited on pages 54, 176, 192, and 215.
- Zängl, G., Reinert, D., Rípodas, P., & Baldauf, M. (2015). The icon (icosahedral non-hydrostatic) modelling framework of dwd and mpi-m: Description of the non-hydrostatic dynamical core. *Quarterly Journal of the Royal Meteorological Society*, 141(687), 563–579.
Cited on pages 58, 60, and 180.
- Zelinka, M. D., Myers, T. A., McCoy, D. T., Po-Chedley, S., Caldwell, P. M., Ceppi, P., ... Taylor, K. E. (2020). Causes of higher climate sensitivity in cmip6 models. *Geophysical Research Letters*, 47(1), e2019GL085782.
Cited on page 196.
- Zhang, X., Zwiers, F. W., Hegerl, G. C., Lambert, F. H., Gillett, N. P., Solomon, S., ... Nozawa, T. (2007). Detection of human influence on twentieth-century precipitation trends. *Nature*, 448(7152), 461–465.
Cited on page 15.
- Zhu, P., Bretherton, C. S., Köhler, M., Cheng, A., Chlond, A., Geng, Q., ... others (2005). Intercomparison and interpretation of single-column model simulations of a nocturnal stratocumulus-topped marine boundary layer. *Monthly weather review*, 133(9), 2741–2758.
Cited on page 20.
- Zimmer, M., Craig, G., Keil, C., & Wernli, H. (2011). Classification of precipitation events with a convective response timescale and their forecasting characteristics. *Geophysical Research Letters*, 38(5).
Cited on page 71.
- Zipser, E. (1977). Mesoscale and convective-scale downdrafts as distinct components of squall-line structure. *Monthly Weather Review*, 105(12), 1568–1589.
Cited on pages 41, 42, and 53.
- Zipser, E. J. (1969). The role of organized unsaturated convective downdrafts in the structure and rapid decay of an equatorial disturbance. *Journal of Applied Meteorology*, 8(5), 799–814.
Cited on pages 52, 53, 193, and 215.
- Zomer, R. J., Trabucco, A., Metzger, M. J., Wang, M., Oli, K. P., & Xu, J. (2014). Projected climate change impacts on spatial distribution of bioclimatic zones and ecoregions within the Kailash sacred landscape of China, India, Nepal. *Climatic Change*, 125(3), 445–460. doi: 10.1007/s10584-014-1176-2
Cited on page 155.

- Zou, L., & Zhou, T. (2015, 06). Asian summer monsoon onset in simulations and CMIP5 projections using four Chinese climate models. *Advances in Atmospheric Sciences*, 32. doi: 10.1007/s00376-014-4053-z
Cited on page 155.
- Zuidema, P., Li, Z., Hill, R. J., Bariteau, L., Rilling, B., Fairall, C., ... Hare, J. (2012). On trade wind cumulus cold pools. *Journal of the Atmospheric Sciences*, 69(1), 258-280. Retrieved from <https://doi.org/10.1175/JAS-D-11-0143.1> doi: 10.1175/JAS-D-11-0143.1
Cited on pages 176, 179, 183, 185, and 192.
- Zuidema, P., Torri, G., Muller, C., & Chandra, A. (2017). A survey of precipitation-induced atmospheric cold pools over oceans and their interactions with the larger-scale environment. *Surveys in Geophysics*, 38(6), 1283–1305.
Cited on pages 52 and 176.