Multivariate statistical approaches for bias adjustment of climate simulations and compound events analysis

Bastien Francois

université
PARIS-SACLAY

Multivariate statistical approaches for bias adjustment of climate simulations and compound events analysis

Approches statistiques multivariées pour l'ajustement des biais des simulations climatiques et l'analyse des événements composés

Thèse de doctorat de l'université Paris-Saclay

École doctorale n° 129 : sciences de l'environnement d'Île-de-France (SEIF)
Spécialité de doctorat : Sciences du climat, de l'atmosphère et des océans, terrestres et planétaire
Graduate School : Géosciences, climat, environnement et planètes,
Référent : Université de Versailles-Saint-Quentin-en-Yvelines

Thèse préparée dans l'unité de recherche LSCE (Université Paris-Saclay, CNRS, CEA, UVSQ), sous la direction de Mathieu VRAC, directeur de recherche.

Thèse soutenue à Paris-Saclay, le 23 septembre 2022, par

Bastien FRANÇOIS

Composition du jury

Membres du jury avec voix délibérative

Sylvie Joussaume Directrice de recherche, CNRS LSCE-IPSL, Gif-sur-Yvette	Présidente
Juliette Blanchet Chargée de recherche, CNRS Université Grenoble Alpes	Rapporteur & Examinatrice
José Manuel Gutiérrez Professeur, Université de Cantabria, Santander	Rapporteur & Examinateur
Sylvie Parey Ingénieure de recherche, EDF/R&D, Palaiseau	Examinatrice
Jakob Zscheischler Professeur assistant, UFZ, Leipzig	Examinateur

THESE DE DOCTORAT

NNT : 2022UPASJ018

université PARIS-SACLAY

ÉCOLE DOCTORALE
Sciences de l'Environnement d'Île-de-France (SEIF)

Titre : Approches statistiques multivariées pour l'ajustement des biais des simulations climatiques et l'analyse des événements composés

Mots clés : modèle climatique, statistiques, multivarié, biais, événements extrêmes

Résumé: Le climat est un système complexe qui est le résultat de multiples interactions entre ses différentes composantes et ses multiples variables. Cette thèse a pour but d'évaluer si et comment l'utilisation d'approches statistiques multivariées pour l'étude des simulations climatiques peut contribuer à une compréhension plus approfondie du changement climatique et des événements climatiques à forts impacts sur la société. Pour répondre à ces questions, je propose et applique de nouveaux outils statistiques multivariés pour, d'une part, la correction de biais des simulations climatiques, et d'autre part, l'étude des changements de probabilités d'événements conjoints à forts impacts. Le travail s'articule autour de trois objectifs : (i) comparer des méthodes de correction de biais multivariés (MBC) déjà existantes, (ii) développer une nouvelle méthode MBC pour l'ajustement des dépendances spatiales des simulations climatiques, (iii) évaluer la période d'émergence des probabilités d'événements conjoints et quantifier la contribution des propriétés univariées et multivariées aux changements de ces probabilités.

La comparaison de méthodes de correction de biais a permis d'une part d'informer les utilisateurs de leurs avantages et leurs inconvénients mais aussi d'identifier des pistes de développements pour de nouvelles méthodes. Une nouvelle méthode, basée sur une technique de Machine Learning appelée réseaux adverses génératifs (CycleGAN), a été développée. Elle donne des résultats satisfaisants, montrant ainsi le potentiel des techniques de Machine Learning pour la correction de biais multivariés. L'évaluation de la période d'émergence des probabilités d'événements conjoints, ainsi que la quantification de la contribution des propriétés univariées et multivariées aux changements de ces probabilités se révèlent être une procédure pertinente pour améliorer la compréhension de tels phénomènes climatiques. Il est trouvé que la non-stationnarité de la structure de dépendance inter-variable dans un contexte de changement climatique peut jouer un rôle important dans les probabilités futures d'événements conjoints.

Les travaux réalisés dans cette thèse ouvrent des perspectives pertinentes en termes méthodologiques mais participent aussi à une amélioration de la compréhension du climat et de ses évolutions en fournissant des outils statistiques adaptés à la nature intrinsèquement multivariée du système climatique.

université PARIS-SACLAY

ÉCOLE DOCTORALE
Sciences de l'Environnement d'Île-de-France (SEIF)

Title: Multivariate statistical approaches for bias adjustment of climate simulations and compound events analysis

Keywords: climate model, statistics, multivariate, bias, extreme events, climate change

Abstract: Climate is a complex system resulting from various interactions between its different components and its multiple variables. This thesis aims to assess whether and how the use of multivariate statistical approaches for the study of climate simulations can contribute to a deeper understanding of climate change and high-impact climate events. To answer these questions, I propose and apply new multivariate statistical tools for, on the one hand, bias correction of climate simulations, and on the other hand, the investigation of changes in the probabilities of compound climate events. The work conducted pursues three main objectives: (i) to intercompare existing multivariate bias correction (MBC) methods, (ii) to develop a new MBC method for adjusting spatial dependencies of climate simulations, (iii) to assess the time of emergence of compound events probabilities, as well as to quantify the contribution of marginal and dependence properties to these changes of probabilities.

The intercomparison of multivariate bias correction methods allowed, first, to better inform end-users of their advantages and disadvantages and, also, to identify avenues for the development of new methods. A new method, based on a Machine Learning technique named cycle-consistent generative adversarial networks (CycleGAN), has been developed. It gives satisfactory results, thus showing the potential of Machine Learning techniques for multivariate bias correction. Assessing the time of emergence of compound events probabilities and quantifying the contribution of univariate and multivariate properties to these changes has proved to be relevant to better investigate compound events. It is found that non-stationarity in inter-variable dependence structures under climate change can play a significant role in future probabilities of compound events.

The work carried out in this thesis opens up relevant perspectives in terms of methodology but also contributes to an improved understanding of the climate and its evolution. It provides new statistical tools that are adapted to the intrinsically multivariate nature of the climate system.

A mon père,
Arnaud François

Remerciements

Voilà, ça y est. Après ces trois années de travail, voici venu le moment d'écrire la partie la plus importante de mon manuscript de thèse: les remerciements! Quelle pression! Mais il est vrai que c'est un moment important: cette thèse n'aurait en réalité pas été possible sans l'aide de nombreuses personnes. Je profite donc de cette section pour remercier tous ceux qui ont contribué à faire de mes années de doctorat une aventure scientifique et humaine incroyable.

Je tiens tout d'abord à remercier sincèrement mon directeur de thèse Mathieu VRAC pour la confiance qu'il a toujours su m'accorder, et ce, dès le premier jour. Lorsque j'ai postulé en M1 pour une offre de stage au LSCE, en vérité, je ne savais pas vraiment où je mettais les pieds. En m'embarquant en tant que co-pilote à bord de sa navette (bas-carbone), Mathieu m'a alors fait découvrir la constellation de la recherche en science du climat et des statistiques appliquées. Un voyage de 3 ans à explorer en orbite les planètes fascinantes de la correction de biais et des évènements composés, et pendant lequel Mathieu a pu me transmettre son goût pour la recherche et la découverte. Je garderai comme un modèle à suivre sa rigueur scientifique et son écoute attentive, sachant toujours réajuster subtilement la trajectoire de mes idées parfois un peu farfelues.

Merci à Juliette y gracias a José Manuel, d'avoir accepté de réviser mon manuscrit de thèse résumant cette aventure de 3 ans. Merci également à Sylvie J., Sylvie P. et Jakob d'avoir été examinateurs, et d'avoir soulevé des éléments de discussions pertinents pendant ma soutenance. Merci à Aurélien et Julien, pour les discussions scientifiques pendant les comités de thèse et les nombreux conseils donnés sur la route à suivre. Merci à Jérôme, mon parrain de thèse, qui a toujours su garder un oeil bienveillant sur moi.

Merci à l'ensemble de l'équipe ESTIMR de bien avoir voulu m'accueillir pour cette mission. Merci au chef d'équipe Davide, pour son dévouement et sa bonne humeur à toute épreuve. A Pascal, pour son oreille (de musicien) attentive. Merci à Nathalie et Philippe, pour nos discussions pouvant passer d'un extrême à l'autre.

Bien sûr, comment ne pas remercier Meriem, co-pilote embarquée en même temps que moi pour explorer l'usage des analogues. Et avec qui j'ai pu avoir la chance de partager les meilleurs moments de la thèse, mais aussi les plus difficiles. Merci également à Soulivanh pour son soutien et nos discussions informelles sur les sciences du climat et les statistiques. Son aide m'a particulièrement été précieuse pendant ces 3 années, et je l'en remercie.

Merci à tous ceux que j'ai pu croiser sur mon chemin : les anciens thésards qui montrent la voie, comme Linh et Florentin, ainsi que les post-docs Andreia, Flavio, Pradeebane, Yann et Yoann, toujours plein de bons conseils, que ce soit pour préparer la soutenance, corriger des biais ou pour réaliser des spagghetti "cacio e pepe". Je tiens également à remercier les nouveaux (et

moins nouveaux) thésards pour l'entraide et les rigolades. Merci à Camille pour nos discussions profondes sur le vélo et les trottinettes à Paris; merci à Robin pour son analyse géopolitique la plus fine d'Ile-de-France; merci à Vaishnavi pour sa gentillesse et les échanges franco-indiens; merci à Clément, Gaëlle et Stella pour les promenades dans Vienne et leur bonne humeur autour d'un bon café; merci à Mireia et Lia de partager cet amour pour le reggaeton; merci à Paula pour me rappeler le bonheur de jouer de la musique en groupe (je vais m'y remettre, promis!); merci à Germain pour aimer le fromage. J'ai une pensée particulière pour Alizée[2], Nemo et Nicolas pour leur sympathie. Merci également à Mats, parti vers de nouveaux horizons après ses stages. Merci à Pierre, Will et Didier, les vélotaffeurs déchaînés, d'avoir bien voulu m'initier au trajet Paris–Gif-sur-Yvette. Merci à Florence et John pour leur gentillesse de tous les jours.

Je tiens également à saluer tous mes amis. Les olympiens, les skieurs-skaters, les mangeurs d'aligots et de magrets, les escaladeurs, les runners, les fans de F1 et de karaoké, ils se reconnaîtront.

Je voudrais remercier ma famille qui m'a toujours accompagné et soutenu dans mes choix. Merci à ma mère Laurence et mon père Arnaud, à qui je dois tout. Merci à ma soeur Manon pour partager ces précieux souvenirs ensemble. Merci au Tonton Nico et au Zio Polo. Merci aux grands-parents, Bonp' et Bonne, Papito et Mamie.

Je tiens enfin à remercier Valérie, qui a joué un rôle déterminant dans la réalisation de cette thèse. A vivre mes périodes stress-stress et mes "sursauts" d'humeur. Merci pour ton soutien et tout ce que tu m'apportes. Merci d'être à mes côtés.

Summary

Climate is a complex system resulting from various interactions between its different components and its multiple variables. This thesis aims to assess whether and how the use of multivariate statistical approaches for the study of climate simulations can contribute to a deeper understanding of climate change and high-impact climate events. To answer these questions, I propose and apply new multivariate statistical tools for, on the one hand, bias correction of climate simulations, and on the other hand, the investigation of changes in the probabilities of compound climate events. The work conducted pursues three main objectives: (i) to intercompare existing multivariate bias correction (MBC) methods, (ii) to develop a new MBC method for adjusting spatial dependencies of climate simulations, (iii) to assess the time of emergence of compound events probabilities, as well as to quantify the contribution of marginal and dependence properties to these changes of probabilities.

The intercomparison of multivariate bias correction methods allowed, first, to better inform end-users of their advantages and disadvantages and, also, to identify avenues for the development of new methods. A new method, based on a Machine Learning technique named cycle-consistent generative adversarial networks (CycleGAN), has been developed. It gives satisfactory results, thus showing the potential of Machine Learning techniques for multivariate bias correction. Assessing the time of emergence of compound events probabilities and quantifying the contribution of univariate and multivariate properties to these changes has proved to be relevant to better investigate compound events. It is found that non-stationarity in inter-variable dependence structures under climate change can play a significant role in future probabilities of compound events.

The work carried out in this thesis opens up relevant perspectives in terms of methodology but also contributes to an improved understanding of the climate and its evolution. It provides new statistical tools that are adapted to the intrinsically multivariate nature of the climate system.

Résumé

De nombreuses catastrophes liées au climat, tels que les feux de forêts ou les pertes de récoltes, résultent souvent de la combinaison de plusieurs phénomènes climatiques, également appelés "événements conjoints". En interagissant les uns avec les autres, ces évènements climatiques peuvent avoir des impacts environnementaux et sociétaux considérables, à une échelle potentiellement bien supérieure que ce qu'auraient pu causer séparement ces événements climatiques. Une compréhension approfondie de ces événements conjoints nécessite donc l'étude de leurs interactions ou dépendances. Pour étudier les caractéristiques futures de ces événements, les scientifiques utilisent des projections climatiques. Ces projections sont obtenues à partir de modèles climatiques qui simulent les différents processus du climat de la Terre à l'aide d'équations mathématiques basées sur la physique. Les modèles climatiques ont prouvé leur capacité à simuler des aspects importants du climat actuel et à fournir des informations fiables sur les climats passés et les changements climatiques futurs. Cependant, les simulations issues des modèles climatiques peuvent tout de même présenter des erreurs - ou biais statistiques -, c'est-à-dire qu'elles ne fournissent souvent pas une représentation appropriée du système climatique, que ce soit dans les valeurs simulées ou les dépendances entre les variables climatiques. Ceci est particulièrement problématique car des simulations avec des propriétés statistiques réalistes sont nécessaires pour étudier de manière robuste non seulement les événements conjoints à forts impacts, mais aussi les événements climatiques plus habituels et ayant des impacts moins importants.

Cette thèse a pour but d'évaluer si et comment l'utilisation d'approches statistiques multivariées pour l'étude des simulations climatiques peut contribuer à une compréhension plus approfondie du changement climatique et des événements climatiques à forts impacts sur la société. Pour répondre à ces questions, je propose et applique de nouveaux outils statistiques multivariés pour, d'une part, la correction de biais des simulations climatiques, et d'autre part, l'étude des changements de probabilités d'événements conjoints à forts impacts. Le travail s'articule autour de trois objectifs : (i) comparer des méthodes de correction de biais multivariés (MBC) déjà existantes, (ii) développer une nouvelle méthode MBC pour l'ajustement des dépendances spatiales des simulations climatiques, (iii) évaluer la période d'émergence des probabilités d'événements conjoints et quantifier la contribution des propriétés univariées et multivariées aux changements de ces probabilités.

La comparaison de méthodes de correction de biais a permis d'une part d'informer les utilisateurs de leurs avantages et leurs inconvénients mais aussi d'identifier des pistes de développements pour de nouvelles méthodes. Une nouvelle méthode, basée sur une technique de Machine Learning appelée réseaux adverses génératifs (CycleGAN), a été développée. Elle donne des résultats satisfaisants pour l'ajustement des propriétés spatiales, montrant ainsi le potentiel des techniques de Machine Learning pour la correction de biais multivariés. L'évaluation de la période d'émergence des probabilités d'événements conjoints, ainsi que la quantification de

la contribution des propriétés univariées et multivariées aux changements de ces probabilités se révèlent être une procédure pertinente pour améliorer la compréhension de tels phénomènes climatiques. Il est trouvé notamment que la non-stationnarité de la structure de dépendance inter-variable dans un contexte de changement climatique peut jouer un rôle important dans les probabilités futures d'événements conjoints.

Les travaux réalisés dans cette thèse ouvrent des perspectives pertinentes en termes méthodologiques mais participent aussi à une amélioration de la compréhension du climat et de ses évolutions en fournissant des outils statistiques adaptés à la nature intrinsèquement multivariée du système climatique.

Contents

1 Introduction, context and objectives **1**
- 1.1 General context . . . 1
- 1.2 Multivariate statistics: a view to investigate climate . . . 3
- 1.3 Compound events . . . 8
- 1.4 Adjusting climate simulations using statistical tools . . . 12
- 1.5 Objectives of the thesis . . . 17
- 1.6 Outline of the manuscript . . . 18
- Résumé . . . 19

2 Intercomparison of MBC methods for climate simulations **21**
- 2.1 Preamble and methodological summary . . . 21
- 2.2 Article published in *Earth System Dynamics*: Multivariate bias corrections of climate simulations: which benefits for which losses? . . . 24
- 2.3 Summary and reminder of the conclusions . . . 51
- 2.4 Perspectives . . . 52
- Résumé . . . 54

3 Development of a new MBC method using Machine Learning **57**
- 3.1 A brief overview of neural networks . . . 57
- 3.2 CNNs in climate science . . . 59
- 3.3 CNNs and GANs for MBC: how did I get there? . . . 60
- 3.4 Article published in *Climate Dynamics:* Adjusting spatial dependence of climate model outputs with cycle-consistent adversarial networks . . . 61
- 3.5 Summary and conclusions . . . 94
- 3.6 A few comments on what did not work . . . 94
- 3.7 Perspectives . . . 95
- Résumé . . . 99

4 Time of Emergence of compound events: contribution of univariate and dependence properties **101**
- 4.1 Introduction . . . 101
- 4.2 Article submitted to *Natural Hazards and Earth System Sciences* . . . 104
- 4.3 Summary and conclusions . . . 150
- 4.4 Perspectives . . . 151
- Résumé . . . 154

5 Conclusion **155**
- 5.1 Conclusion . . . 155

5.2 Perspectives 157

Bibliography **163**

Appendix **189**

A Supplement of the article "Multivariate bias corrections of climate simulations: which benefits for which losses?" **191**

B Supplement of the article "Adjusting spatial dependence of climate model outputs with cycle-consistent adversarial networks" **205**

C Supplement of the submitted article "Time of Emergence of compound events: contribution of univariate and dependence properties" **217**

Chapter 1

Introduction, context and objectives

1.1 General context

In 2019, Australia has experienced high temperatures and prolonged dry conditions, which resulted in one of the worst bush fire seasons in its recorded history. It caused massive damages throughout the country, affecting both people and ecosystems[1]. In October 2020, Central Vietnam was struck by heavy rain induced by a sequence of typhoons, leading to extensive floodings and landslides. During summer 2021, a persistent high-pressure system (or "heat dome") occurred over Western North America, and particularly Western Canada, producing an unprecedented heat wave which caused important economic losses and human casualties[2]. In August 2021, 50.3°C were recorded in Kairouan (Tunisia)[3], the highest temperature ever observed for this country since the start of reliable measures. In April 2021[4] and 2022[5], consecutive days of severe frost occurred after bud burst over Central Europe. It caused severe damages to crops, including grapevine and fruit trees. These climate events, which occurred during my thesis all over the world, are some examples of events with dramatic consequences on both human societies and ecosystems. This is however not new: high-impact climate events have always occurred as they are an expression of climate variability, and improving the understanding of these phenomena has always been crucial. When these specific climate phenomena are studied in more detail, they often turn out to be the result of multiple climate hazards occurring simultaneously or successively. By interacting with each other, these hazards can lead to exacerbate impacts, at a scale potentially far greater than any of these climate events could have caused individually. A thorough understanding of these climate phenomena, also known as "compound events", therefore requires the study of these interactions or dependencies. Moreover, it is now unequivocal that greenhouse gases emitted by human activities have warmed the atmosphere, ocean and land inducing many changes across the climate system (IPCC, 2021). Assessing whether these dependencies change, how and with which consequences is a major challenge for mitigation and adaptation issues.

[1] https://reliefweb.int/sites/reliefweb.int/files/resources/IBAUbf050220.pdf

[2] https://www.bbc.com/news/world-us-canada-57678054

[3] https://watchers.news/2021/08/11/tunis-kairouan-record-temperature-tunisia-august-2021/

[4] https://www.worldweatherattribution.org/human-caused-climate-change-increased-the-likelihood-of-early-growing-period-frost-in-france/

[5] https://www.washingtonpost.com/weather/2022/04/04/europe-record-cold-france-agriculture/

To investigate the future characteristics of compounding events such as their frequencies or intensities, scientists use climate projections. These projections are obtained from climate models that simulate the various processes of the Earth's climate using mathematical equations based on physical, biological and chemical knowledge. Climate models have proven their ability not only to simulate important aspects of the current climate but also to provide reliable and useful information on past climates and future climate changes. Our understanding of compound events, and more generally the climate system, comes in part through the examination of the huge amount of simulated data produced by climate models. The study of climate using statistics, also called statistical climatology, is particularly relevant as it provides quantitative tools to summarise and model the climate system to investigate its properties. In particular, the use of multivariate statistics, for which the relationships - or dependence - between the different variables is an essential aspect, allows to further improve our understanding of the complex interactions involved in the Earth's climate. These multivariate tools can be applied to investigate evolution of dependencies in simulations from climate models. However, as the British statistician George E.P. Box puts it "Essentially, all models are wrong, but some are useful"[6]. While it originally referred to statistical models, it can be applied to scientific models in general, including climate models. Consequently, simulated outputs from climate models can present errors — or statistical biases —, i.e., they often fail to provide an appropriate representation of the climate system, either in the simulated values or dependencies between climate variables. This is particularly problematic as simulations with realistic statistical properties are required to robustly investigate not only compounding hazards leading to high impacts but also more usual and less impacting climate events.

The goal of this thesis is to (partly) assess whether and how the use of multivariate statistical approaches for the study of climate simulations can contribute to a deeper understanding of climate change and high-impact climate events. The work presented in this thesis was carried out in this context, by evaluating and designing new multivariate statistical tools that take into account the dependence between different climate variables. First, I was interested in applying and designing multivariate statistical methods able to increase the realism of simulations by adjusting their statistical biases on dependence properties. Then, a new methodology based on multivariate statistical tools has been developed to evaluate significant changes (and emergence, see later) of compound events probabilities and investigate the contribution of dependence to the changes of compound events probability. The rest of this chapter will introduce the main concepts of statistical climatology used in this thesis. It will also provide a context for the work done in this manuscript. The goal of this chapter is not to be exhaustive, as several books that give a broader overview of statistical climatology already exist (e.g., Storch and Zwiers, 1999; Wilks, 2006). It simply gives a few reasons why the use of multivariate statistical tools to adjust statistical biases of climate models and investigate compound events properties is of interest for the scientific community.

[6]This quote comes from Box's book *Empirical Model-Building and Response Surfaces*, which was co-authored with Norman Draper.

1.2 Multivariate statistics: a view to investigate climate

Climate as a physical object

Earth's climate is complex and constantly evolving. It results from various interactions between its different components, which include the atmosphere, oceans, cryosphere, lithosphere and land surface. Since Antiquity, human societies have tried to develop theories to explain the observed climatic phenomena such as rain or thunders (e.g., in *Meteorologica*, Aristotle). However, it was the development of physical theories in the 18th and 19th centuries that contributed greatly to an improved understanding of the structure and dynamics of the Earth's climate. For example, in 1735, George Hadley proposed a physical mechanism to explain the motion of air masses in the atmosphere with Earth's rotation (Hadley, 1735). Nowadays, this mechanism, known as the Hadley cell, remains a relevant explanation of the Earth's atmospheric circulation that occurs in both hemispheres near the equator. In 1824, Joseph Fourier was the first to propose that the Earth's atmosphere acts to increase the planet's temperature, a phenomenon now known as the greenhouse effect (Fourier, 1824). In 1856, Eunice Newton Foote experimented on the warming effect of sunlight on different gases (Foote, 1856). Her work led to improve the understanding of how chemical composition of the atmosphere could affect climate. In 1837, after observing erratic blocks of Alpine rocks, the geologist Louis Agassiz was the first to scientifically propose that the Earth had been subject to a past ice age, and thus a different climate (Agassiz, 1837). These past climate changes would then be explained by Arrhenius (1896) and Chamberlin (1899) as resulting from past changes in the concentration of carbon dioxide in the atmosphere. It was only later that past climate changes were also explained by the slow variations of the Earth's orbit (Milankovitch, 1930). This list of groundbreaking findings during the 19th century in climate physics is of course not exhaustive, and a more complete overview of climate science history can be found in Le Treut et al. (2007) or in the thesis of Jézéquel (2018).

Based on experiments and careful examination of facts, the development of physical theories continued during the 20th century. It attempted to better understanding the complex internal processes of the climate and the influence of external forcings (such as solar, volcanic and anthropogenic greenhouse gases emissions) on its variations. Bjerknes (1904) – and later Richardson (1922) – proposed to use equations derived from physical principles to describe interactions between atmospheric processes and to perform predictions. However, these differential equations were not solvable in their time due to limited computing power. It was not until the 1950s that these computational limits were overcome thanks to major advances in computer technology. Using a programmable computer, John von Neumann and his assistant Jule Gregory Charney performed the first numerical short-term weather forecasts for the United States based on simplified equations of the atmosphere (Charney et al., 1950). Following this success, scientists continued to develop more and more sophisticated models to reproduce the climate system at longer time scales. These dynamical models, also known as general circulation models (GCMs) aim at simulating the main features of the Earth's climate on time scales ranging from decades to centuries. For this purpose, scientists combined atmospheric processes with other components that play an important role on climate variations at longer time scales such as ocean, sea-ice or vegetation. In 1969, Manabe and Bryan (1969) published the first simulations of the climate by a coupled ocean-atmosphere model[7]. In the 1970s, the development of satellites and climate sta-

[7]During my thesis, Syukuro Manabe was named among the winners of the 2021 Nobel Prize in physics "for the physical modeling of Earth's climate, quantifying variability and reliably predicting

tions networks allowed to routinely collect observations. Collecting high-precision observations participated in the improvement of physical knowledge by providing a better representation of the state of the atmosphere. By producing simulations of current and past climates in agreement with observations, GCMs became an important tool to simulate and improve our understanding of the different physical processes involved in the Earth's climate. By solving numerically the equations that describe the dynamics of the climate system, these models provide simulations over a discretized three-dimensional space formed by grid cells. Based on well-founded physical principles, GCMs are used nowadays to explore future climate changes and their potential impacts by considering several scenarios of plausible futures for greenhouse gas emissions (IPCC, 2021).

As climate results from very complex interactions between processes governed by physical laws, it is thus only natural that scientists studied through the years its functioning using physical sciences. However, our understanding of the climate system comes in part through the manipulation and examination of the huge mass of data, such as those observed or generated by GCMs. It thus has led to the development of statistical tools to extract relevant information about natural phenomena and complement the physical sciences to better understand the climate system.

Climate as a statistical object

Statistics is the discipline that concerns the collection, analysis and interpretation of data. It is thus not surprising that statistics have many roles to play in climate science to analyse not only observational datasets, but also outputs from climate models. Although all statistical tools aim at improving the understanding of phenomenon under study, they can be divided into two categories depending on their approaches: tools for the extraction of useful information from dataset, and tools for the generation of new data. Both approaches are relevant and widely used in climate science.

When confronted with large batches of data, descriptive statistics are quite useful as they provide simple summaries about the characteristics of datasets, for example with measures of central tendency (means, medians, ...) or measures of variability (standard deviation, variance, ...). The use of descriptive statistics in climate science takes its roots from the first attempts by geographers and climatologists to classify the different Earth's climates over the globe (such as dry or tropical climates) by computing averages of meteorological variables (e.g., Köppen, 1931). Quoting from the glossary of the IPCC report (IPCC, 2014), *"Climate in a narrow sense is usually defined as the average weather, or more rigorously, as the statistical description in terms of the mean and variability of relevant quantities over a period of time ranging from months to thousands or millions of years."* This emphasises that, when dealing with climate, it is not only the values of the various atmospheric variables that are important in themselves, but rather their statistics, and the evolution of these statistics (e.g., trends, changes over time). In addition to providing simple summaries of data, statistics have proven their usefulness to describe the whole distribution of climate variables using probability density functions (PDFs). PDFs define the relationship between the possible values for the variables and their probability of occurrences. From my point of view, this is one of the places where statistics are particularly relevant to investigate climate: it allows to characterise the set of possible climate events and their associated probabilities. When used to describe distributions of climate variables, PDFs

global warming" https://www.nobelprize.org/prizes/physics/2021/manabe/facts/

can permit to investigate individual climate extremes (e.g. extreme surface temperature or precipitation), including their occurrences and intensities in both observations and climate model outputs (e.g., Kharin and Zwiers, 2000; Coles, 2001). In a climate change context, PDFs can be used to perform extreme event attribution (EEA) studies. EEA studies aim at quantifying to what extent an individual weather event is attributable to anthropogenic climate change (Stott et al., 2016). To address this question, a methodology based on probability distribution fitting have been developed (called the "risk-based" approach, Stott et al., 2004; Shepherd, 2016). This methodology consists in estimating the parameters of the PDFs for climate variables in a factual world (i.e., in a world with anthropogenic climate change) and a counterfactual world (i.e. in a world in which anthropogenic emissions had never occurred) to then evaluate probabilities of a particular event in these two worlds, respectively. Comparing the probabilities between the two worlds permits to evaluate the effects of climate change on the occurrence of climate events.

Not only do statistics can be used for descriptive purposes, they can also be part of procedures that permit to generate — i.e., simulate — new synthetic sets of data. Generations of new data using statistical modelling are performed for various applications in climate science. For example, weather variables (e.g., precipitation) can be parameterised using probability distributions to produce realistic simulations with statistical properties similar to those of observations (Stochastic Weather Generators, Wilks and Wilby, 1999; Allard et al., 2015; Vaittinada Ayar et al., 2016). These simulations are of particular interest to explore the different possibilities of weather conditions: for instance, simulated rainfall events can serve as inputs to rainfall-runoff models in order to investigate hydrological responses and their associated uncertainties (e.g., Russo et al., 2006; Jonsdottir et al., 2006; Gabellani et al., 2007). Statistical modelling has also been employed to generate predictions of atmospheric variables (statistical forecasting, e.g., Vislocky and Fritsch, 1997; Wilks, 2006) or to predict spatial distributions of endangered species in ecological studies (e.g., Phillips et al., 2006; Elith et al., 2006). In addition, a relevant application area using statistical approaches to produce new climate data is stochastic parameterization. This research field consists in investigating how processes that are too small-scale to be physically represented in climate models (e.g., convection) can be replaced by stochastic processes. Including stochastic parameterization schemes in climate modelling showed to be effective to reduce climate model biases and to improve weather forecasts (e.g., Teixeira and Reynolds, 2008; Batté and Doblas-Reyes, 2015; Berner et al., 2017), although this is not necessarily systematic (e.g., Maher et al., 2018).

An important hypothesis in statistical modelling is that of stationarity, i.e. the assumption that the statistics to be reproduced are constant over the space–time modelling domain. In their most basic unconditional form when calibrated only based on a reference dataset, statistical models are able to infer the statistics from this reference dataset and to reproduce them in simulations. Such statistical approaches are thus stationary. However, statistics of climate variables can be nonstationary, e.g., due to interannual climate variations or in a context of global climate change. It implies for statistical modelling that parameters of probability functions to estimate can change over time. One common way to deal with nonstationarity is to use covariates that permit to model statistics *conditionally* on some relevant information. Such covariates can include information from large-scale climate indices, local climate or seasonal information. For example, many versions of stochastic weather generators have been developed to simulate weather sequences conditioned on some aspect of large-scale atmospheric circulation (e.g., Bardossy and Plate, 1992; Wallis and Griffiths, 1997; Wilks and Wilby, 1999). In this manuscript, nonstationarity of climate variables will be investigated. It will be achieved by studying the change of their statistical properties between different time periods.

From univariate to multivariate statistics in climate science

Statistical treatment can be applied to one-dimensional variable: its univariate characteristics can be summarised using descriptive statistics, and its probability density function can be investigated to generate individual random realisations. Univariate statistical methods are quite popular as they are often simple to implement and results are easy to interpret. However, climate datasets are often composed of several variables, and are therefore *multivariate*. For example, multivariate climate data could include observations of several variables at one location, observations for one variable at different grid points (spatial fields), or several variables at several grid points. In addition, all these variables can be selected at different altitudes or pressure levels. When multiple variables are studied, univariate methods can be applied to individual variables. However, a fundamental aspect when analysing multiple variables is the statistical relationship — or dependence structure — between the different variables, i.e, analysing their joint variability (Anderson, 1958). Methods that consider individual variables separately are not designed to describe such relationships between variables, and using them to investigate dependent climate variables is therefore limited.

To study the climate system where the relations among variables are important, climate scientists have long adapted and applied multivariate statistical tools from statistical theory. For example, analysis of correlations (Galton, 1889; Pearson and Henrici, 1896) that describe the strength of associations between atmospheric variables allowed the discovering of teleconnections, i.e., that climate variables are related to each other at large distances throughout the troposphere (Walker, 1925). Studying multivariate statistical distributions (e.g., Laplace, 1811; Gauss, 1823; Johnson and Kotz, 1972; Johnson, 1987) allowed to investigate multivariate density functions of climate variables. In particular, the use of the multivariate cumulative distribution functions named copulas (Sklar, 1959) has shown to be useful to gain a better understanding of the dependence between various climate variables. Copulas have been widely applied in various climate applications, e.g., to cluster atmospheric variables (Vrac et al., 2005), to assess return periods of multivariate hydroclimatic extremes (Salvadori and De Michele, 2004; Favre et al., 2004), to examine river and coastal flooding (Brunner et al., 2017; Serinaldi and Kilsby, 2017; Didier et al., 2019), study crop yields (Alidoost et al., 2019) and sediment transport (Shojaeezadeh et al., 2020). Given the complexity of highly dimensional datasets, researchers have also applied approaches to summarise multiple dependencies through data reduction (e.g., Principal Component Analysis, Pearson, 1901; Wilks, 2006) which allows, for instance, to generate synthetic multivariate rainfall and streamflow time series (Westra et al., 2007) or to better investigate atmospheric circulation patterns (Corti et al., 1999; Scherrer et al., 2006) and large-scale modes of variability (Dai and Wigley, 2000; Kessler, 2001). Another important research field that make uses of multivariate statistics is data assimilation (e.g., Kalnay et al., 1996), for which one of the main objectives is to provide a multivariate, spatially complete, and coherent record of the global atmospheric circulation. Such datasets are called reanalysis datasets (e.g., ERA5, Hersbach et al., 2020), and are largely used in climate science to understand climate change and current weather extremes.

Machine Learning: a promising multivariate tool?

Recent advances in computing power and data storage capacity have revolutionised multivariate statistics with the emergence of machine learning (ML) techniques. ML tools gather specific algorithms that are able to learn complex relationships between statistical variables.

A profusion of ML algorithms has been developed to solve a wide variety of problems (e.g., classification, regression or clustering) for different applications, such as in medicine, security, social media or finance. Among the most employed ML methods are random forest (Breiman, 2001), artificial neural networks (McCulloch and Pitts, 1943; Rosenblatt, 1958; Rumelhart et al., 1986; Lecun and Bengio, 1995), support-vector machines (Cortes and Vapnik, 1995), k-nearest neighbors (Fix and Hodges, 1989) and k-means algorithms (Steinhaus, 1956; MacQueen, 1967). One particularity of ML tools compared to classical statistics is that they learn statistical relationships in an automated way directly from sample data (also called "training data") and by making minimal assumptions about these relationships. Although their modelling procedures differ, classical statistics and ML are in fact closely related: classical statistics can be used within ML tools to learn statistical relationships. One basic example of this is linear regression. Developed in the field of statistics to model linear relationship between an output and one or more explanatory variables, linear regression can be used within ML algorithms such as neural networks to compute gradient descent which drives the learning procedure. ML tools demonstrated to be particularly effective to deal with large datasets where there is a large volume of information available, and in cases where it is difficult to use classical statistics, e.g., when the number of statistical variables exceeds the number of observations, in the presence of complicated nonlinear interactions or when it is difficult to make any assumptions about the learning to perform. ML techniques thus offer exciting new opportunities for expanding the knowledge about the Earth climate from the huge amount of data that has become available. By offering the possibility to circumvent some of the difficulties encountered using classical statistical approaches, ML tools have proven to be successful for various climate science applications such as:

- **clouds classification** using artificial neural networks (Lee et al., 1990; Tian et al., 1999) and convolutional neural networks (Zhang et al., 2018),
- **spatial estimation of soil composition** using random forest (Grimm et al., 2008) and gradient boosting (Hengl et al., 2017),
- **detection of extreme weather** using convolutional neural networks (Liu et al., 2016; Racah et al., 2017),
- **short and near term weather forecasting** using recurrent neural networks (Salman et al., 2015) and long short-term memory networks (Shi et al., 2015; Zaytar and Amrani, 2016),
- **seasonal forecasts** using hierarchical clustering (Cohen et al., 2019) and gradient boosting (Qian et al., 2021),
- **modelling of vegetation-climate relationships** using artificial neural networks (Hilbert and Ostendorf, 2001; Jahan and Gan, 2011),
- **land-use and change detection** using convolutional neural networks (Zhao and Du, 2016; Carranza-García et al., 2019),
- **transport modelling** using convolutional neural networks (Tompson et al., 2016; de Bezenac et al., 2017) and artificial neural networks (Vlasenko et al., 2021),
- **modelling of rainfall-runoff processes** using artificial neural networks (Smith and Eli, 1995) and long short-term memory networks (Mao et al., 2021),

- **statistical downscaling** using multitask learning (Vandal et al., 2017), generative adversarial networks (Leinonen et al., 2020), convolutional neural networks (Baño-Medina et al., 2020) and random forest (Legasa et al., 2022).

Other examples of applications of ML tools in climate science can be found in Reichstein et al. (2019).

New tools to consider dependence are needed

All these multivariate tools applied in climate science, either based on Machine Learning or more classical statistical techniques, have proven their value by being able to take into account the relationships between variables. However, the use of multivariate tools and the analysis of their results is not always straightforward. Indeed, multivariate methods can sometimes be more complicated to understand and implement than univariate methods, as they require more data, computing capacity and statistical knowledge. This is particularly true when using Machine Learning tools: despite showing successful results, the lack of an explicit model can make solutions from ML tools difficult to directly link with physical knowledge. Moreover, in a context of climate change, another difficulty can be added when investigating multivariate data, as assessing dependence changes between multiple atmospheric variables can present a key challenge for climate scientists.

Research is therefore needed to improve the understanding of dependence between climate variables and their potential changes. In addition to improving physical knowledge, statistical tools are needed to complement physical science research and contribute to a better consideration of dependencies, which can be done through the development of multivariate statistical tools. In this thesis, I focused on how to exploit multivariate statistical tools to investigate the dependencies of climate variables leading to extreme impact events (compound events). Improving the overall understanding of these climate events and how dependencies evolve and contribute to change the characteristics of compound events is of obvious relevance to society because of their extreme impacts. Then, since our understanding of future climate change depends in part on the realism of dependencies of simulated climate variables obtained from climate models, I have also been interested in applying and designing multivariate bias correction methods able to adjust multivariate properties of climate simulations. These adjustments are potentially crucial to robustly investigate not only compound events but also less impacting climate events. In the rest of this chapter, I will introduce the topics of compound events and bias correction.

1.3 Compound events

What are compound events?

Climate-related disasters can result from an individual extreme weather event. For example, a sudden and localised temperature extreme can affect agricultural yields by damaging crops. However, the most damaging disasters often arise when several climate variables interact between each other, also referred to as "compound events" (CEs). The individual climate variables forming the CEs may not be extreme themselves, but their simultaneous or successive occurrences across multiple spatial and temporal scales can cause extreme impacts (Leonard et al., 2014; Zscheischler et al., 2014, 2018). For example, consecutive dry days with high temperatures

(i.e., not necessarily extremes) can exacerbate the risk of low crop yields (e.g., Ciais et al., 2005).

Firstly introduced by Seneviratne et al. (2012), compound-event research has developed very rapidly in an interdisciplinary framework involving climate science, impact research and statistics. By taking advantage of this interdisciplinarity, research on compound events aims to enhance the understanding of physical processes leading to compound events, to better assess their statistical characteristics and predictability, as well as their impacts on society and ecosystems. To frame compound event research, Zscheischler et al. (2020) proposes a general definition for the different characteristics that form a compound event: modulators, drivers, hazards and impacts. Fig. 1.1 illustrates how these different elements can interact to produce an impact. Hazards are climate phenomena, such as heatwaves, droughts, frosts or wildfires, that cause an impact by interacting between each other. These hazards themselves result from climate drivers such as atmospheric blockings, tropical cyclones or storms. Finally, drivers are affected by modulators such as sea-surface temperature patterns or variability of large-scale climate modes (e.g., El Nino- southern Oscillation). It is through the interactions of these different elements (modulators, drivers and hazards) that extreme impacts are obtained. In a climate change context, all physical elements can be potentially affected: for example, heatwaves can be simply hotter, location of atmospheric blockings causing heatwaves and droughts can be modified; and a change of sea-surface temperature patterns due to climate change can affect the occurrence of atmospheric blocking events.

Different types of interactions

Hazards and drivers constituting compound events can interact in different ways. Although subjective, Zscheischler et al. (2020) proposes a typology of compound events to categorise in four classes the different types of events depending on the underlying interactions:

- "Preconditioned" events for which a weather- or climate-driven background can exacerbate the impact;
- "Multivariate" events for which the simultaneous co-occurrences of individual drivers and/or hazards at the same location result in impacts,
- "Temporally compounding" events for which impacts are due to the successive occurrences of hazards, and
- "Spatially compounding" events for which individual hazards in several locations lead to an impact.

This categorisation of CEs is an illustration that the climate processes causing extreme impacts can be diverse and complex to understand. Of course, compound events do not always fit perfectly the presented typology by being assigned to a unique class. For example, distinguishing preconditioning CEs from temporally CEs is not always straightforward. As illustration, consecutive days of severe frost occurring after a warm winter and bud burst can be interpreted as either preconditioning or temporally compounding events. In addition, compound events can be the results of climate events from two or more categories. For example, impact from wind and precipitation extremes (multivariate event) can be exacerbated if it is preconditioned by saturated soils (preconditioned event) (e.g., van den Hurk et al., 2015). Still, the topology provides great guidance to analyse the causal mechanisms underlying compound events, as it

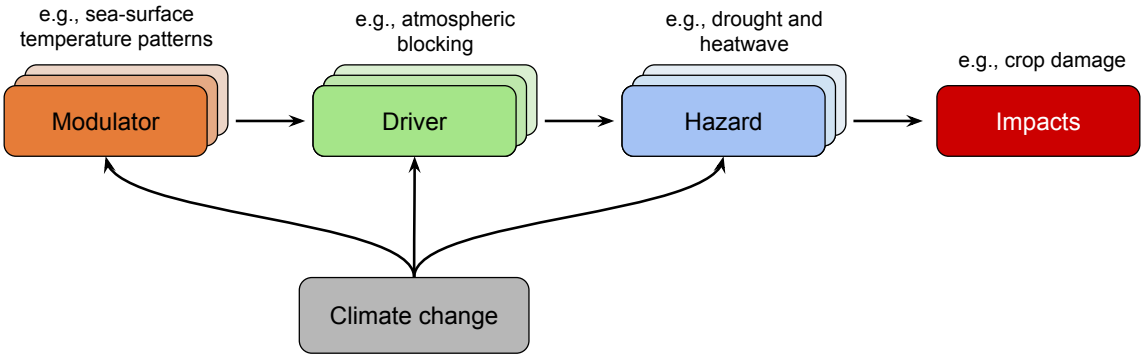

Figure 1.1: Physical elements of a compound climate event. Inspired and modified from Zscheischler et al. (2020).

helps to identify and separate the different elements of compound events. It also shows that dependence between drivers and hazards is key and can be involved in several ways to produce high-impact events.

From univariate to multivariate approaches to better investigate CEs

Before compound-event research was developed, much of the analysis of high-impact events was focused on univariate climate extreme events, even when multiple climate variables were involved (Seneviratne et al., 2012; Collins et al., 2013). Univariate statistical tools were then used to analyse the characteristics and the projections of such climate extreme events (e.g., extreme rainfall, extreme temperatures, Kharin et al., 2013; Lenderink et al., 2007). As by definition, CEs are the results of multiple meteorological processes at play, they require different analysis methods than those used for univariate events. Indeed, considering univariate tools implies that climate drivers and hazards of compound events are considered independent of each other and that their dependence is neglected. It can result in an underestimation of their probability of occurrence and consequently their risks which limits the ability to develop effective adaptation strategies (Kew et al., 2013; van den Hurk et al., 2015; Zscheischler and Seneviratne, 2017; Hillier et al., 2020). For example, considering only high-temperature for heat-stress-adaptation strategies is limited as human health impacts depend on both temperature and humidity (e.g., Coffel et al., 2017; Raymond et al., 2020). Similarly, the design of flood protection measures must take into account *together* the multiple hazards that cause flooding (e.g., sea-level and river discharge, Wahl et al., 2015; Moftakhari et al., 2017; Ward et al., 2018).

Thus, multivariate statistical tools naturally emerge as a promising tool to study not only univariate properties but also dependence relationships between the multiple variables forming the CEs. By providing quantitative tools such as correlation analysis, multivariate regression or multivariate modelling, multivariate statistics permit to describe and investigate the statistical relationships between the different elements leading to extreme impacts. Recently, research has focused on studying a wide variety of compound events in observations or climate models outputs using various multivariate techniques:

- **Compound dry and hot events** using count-based approaches (Wu et al., 2021a; Manning et al., 2022), copula modelling (Zscheischler and Seneviratne, 2017; Sarhadi et al.,

2018; Zhou and Liu, 2018; Manning et al., 2019; Ribeiro et al., 2020) and logistic regression model (Hao et al., 2019; Wu et al., 2021b),

- **Concurrent wind and precipitation extremes** using conditional exceedance model (Zheng et al., 2014), count-based approaches (Ridder et al., 2018; Owen et al., 2021), logistic regression (Martius et al., 2016), copula modelling (Bevacqua et al., 2019), dynamical systems metrics (De Luca et al., 2020a), extremal dependence metrics (Zscheischler et al., 2021) and Bayesian models (Couasnon et al., 2018),
- **Spatially co-occurring climate extremes** using extremal dependence metrics (Brunner et al., 2020), max-stable processes (Blanchet and Davison, 2011; Nicolet et al., 2016; Wang et al., 2014; Oesting and Stein, 2018), multivariate skew-t distribution (Ghizzoni et al., 2010, 2012), multiple linear regression (Singh et al., 2021) and generative adversarial networks (Boulaguiem et al., 2022),
- **Temporal sequences of heavy-precipitation events** using clustering algorithms (Barton et al., 2016), Poisson regression (Villarini et al., 2011), Cox regression (Mallakpour et al., 2017; Yang and Villarini, 2019, 2021) and count-based approach (Kopp et al., 2021),
- **Concurrent heat and air pollution** using clustering algorithm (Schnell and Prather, 2017), Poisson regression (Breitner et al., 2014) and generalized additive model (Cheng and Kan, 2012; Wang et al., 2020),
- **Frost events in spring following a warm winter** using conditional modelling (Sgubin et al., 2018; Vautard et al., 2021) and count-based approaches (Pfleiderer et al., 2019).

By taking dependence into account, all these studies are able to provide valuable information on statistical properties of compound events. Moreover, changes of dependence under climate change have been found to significantly affecting the occurrence of CEs in reanalyses data for the recent decades (e.g., Abatzoglou et al., 2020) or in climate models in future periods (e.g., temperature-precipitation correlations, Vrac et al., 2021). In particular, the latter study shows that the different climate models do not agree on the simulation of future dependencies. This inter-model variability can be partly explained by the different capacities of the climate models to represent dependencies appropriately. This further motivates a better understanding of dependence changes, their influences on CEs properties and how they are represented by climate models.

The importance of an appropriate dependence in simulations

As climate models are one of the main tools for understanding CEs, it is crucial that their representations in climate simulations are correct not only for historical periods but also for future conditions. Appropriate simulations of CEs can then permit to get robust estimates of their occurrence probabilities and their evolutions in a changing climate. In particular, detecting *when* the probabilities of CEs are significantly different from those associated with natural variability is quite relevant for society and policy-makers, not only to better anticipate their emerging risks but also to raise awareness among the general public about climate change. Such change detection corresponds in the climate science literature to the notion of "Time of Emergence" (ToE), which consists in determining the time period at which a climate signal emerges from (i.e., goes out of) the natural variability (e.g., Christensen et al., 2007). For this purpose, climate simulations with appropriate representations of CEs are needed to investigate

robustly their emergence and thus to provide relevant information for adaptation planning. Appropriate simulations of CEs would also be particularly crucial for impact studies, in which simulated data are given as inputs to impact models to explore how climate hazards would impact society and ecosystems in current and future periods.

Hence, simulations with not only appropriate univariate properties but also appropriate dependence structures between climate hazards leading to extreme events are required to derive meaningful results. However, it is now known that simulations from climate models can have various biases in their univariate and dependence properties: statistical properties of simulated outputs are not always in line with those obtained for observations or reanalyses. These statistical biases necessarily affect the robustness of the simulated CEs and their associated analyses. Many univariate bias correction methods have been developed: they are able to adjust univariate properties of the simulations but disregard multivariate aspects (see subsection 1.4). In fact, climate simulations with appropriate multivariate distributions is not only needed to study high-impact climate events, but is required more generally for any climate or impact studies using simulated data and for which dependence properties are important. Statistical methods to adjust the simulated relationships between climate variables, also called multivariate bias correction (MBC) methods, have been recently developed in the literature. Multivariate bias correction will be one of the major topics of this manuscript, namely how to adjust not only univariate properties but also dependence structure between simulated variables. The topic of bias correction of climate simulations is introduced in the following subsection.

1.4 Adjusting climate simulations using statistical tools

Univariate and multivariate biases

Climate models such as GCMs are the main tools to study the climate system and its evolutions at different time scales. However, despite considerable progress in recent years, simulations obtained from climate models can present different statistical properties than those observed, which are called *statistical biases* (e.g., Christensen et al., 2008). For example, a simulated temperature time series at one grid cell can present univariate properties, such as mean, variance or extremes, different from those obtained based on reanalyses over a historical period. Multivariate properties of simulated climate variables can also differ from those observed, such as the dependence structures between several physical variables and/or sites. For example, the correlation between simulated temperature and precipitation at a specific grid cell can be under- or overestimated compared to reanalyses or observations.

The reasons for these statistical biases are multiple. I provide here some of the most obvious reasons for biased model simulations but this list is not exhaustive. One main reason is that, for computational purposes, GCMs discretize the Earth system in a three-dimensional space formed by grid cells. As a result, although reproducing adequately key large-scale physical processes, they often misrepresent important processes that occur at spatial scales smaller than the model resolution (e.g. convective processes). For instance, the Coupled Model Intercomparison Project Phase 5 (CMIP5, Taylor et al., 2011) gathers GCMs with spatial resolutions between 1 and 3°, which is often considered too coarse for a realistic representation of precipitation (e.g., Kundzewicz et al., 2007). A second reason is that, despite continued scientific progress, some biases can arise from incomplete physical knowledge of climate system processes. For example, the complexity of cloud processes is still identified as a major source of biases for climate models

despite recent improvements in its understanding (IPCC, 2021). Another main reason refers to errors in the parameterization chain: when physical processes are too complex to be physically represented, they can be replaced by a simplified process with parameters estimated using observations. The observed data used for parameterization are time series of finite length which may not cover the full range of climate dynamics. Therefore, parameterization is necessarily subject to sampling errors that can then alter the ability of climate models to represent the physical processes (e.g., Ehret et al., 2012).

Therefore, for these reasons, climate models can present statistical biases. This is particularly problematic when climate model simulations are used as inputs for impact models, for instance in hydrological (e.g., Bates et al., 2008; Chen et al., 2013), agronomical (e.g., Wheeler and von Braun, 2013) or epidemiological (Caminade et al., 2014; Chemison et al., 2021) studies. These impact models are calibrated to appropriately represent the statistics of some desired observed variables (such as runoff for hydrology or crop yields for agronomy), given observed climate input. Replacing observed input by simulations can be desirable, for example by selecting simulated projections data for future periods to investigate impacts of climate change. However, biases in climate simulations can significantly affect the realism of the impact model simulation and some bias corrections of climate simulations might be required to obtain more meaningful results from impact studies (e.g., Wilby et al., 2000). To increase the realism of simulations from climate models, two post-processing statistical approaches have been developed in the literature: statistical downscaling and bias correction. Statistical downscaling and bias correction post-processing methods in climate modelling are often treated as identical, but are not. Although the thesis is mainly about bias correction, some clarifications on their differences are here needed to better understand the rest of the manuscript.

Statistical downscaling and bias correction methods

To perform downscaling, two different approaches have been developed in the literature: dynamical and statistical downscaling. On the one hand, dynamical downscaling consists in using Regional Climate Models (RCMs) that numerically solve physical equations describing the dynamics of the climate system to produce local-scale climate information over a limited area. To simulate regional features of the climate at a higher resolution, RCMs use GCMs outputs as inputs, on the boundary of the domain (Rummukainen, 2010). On the other hand, statistical downscaling (SD) methods are conditional statistical models that establish empirical relationships between large-scale information (predictors, i.e., input data) and local-scale observations (predictands, i.e., output data) over the region of interest (Maraun and Widmann, 2018). Fig. 1.2a presents a simple illustration of the process of statistical downscaling methods applied to coarse model outputs. In contrast with dynamical models such as RCMs that resolve physical equations, SD methods have the advantage to generate very quickly random realisations with desired distributional properties. It has several implications such as facilitating the exploration of the uncertainty inherent in the climate system. However, the price to pay for this advantage is that, contrary to dynamical models, statistical models are not derived from physical theory, and thus potentially result in generating outputs with inappropriate physical laws. As this thesis focuses on statistical tools, we provide more detail below on the particularity of statistical downscaling methods.

The statistical relationships determined by SD methods could be based on either perfect prognosis (PP) or model output statistics (MOS) approaches (Maraun et al., 2010). The PP

approach consists in learning the synchronous relationships between a variable of interest from references (predictand) and one or several observed variables (predictors). The consequence is that PP approaches must be calibrated on synchronised predictors and predictands (e.g., reanalyses and observations). Once learned, these relationships can be applied to predictors from climate simulations (Fig. 1.2a). Thus, PP makes the assumption that these simulated predictors are realistically simulated (i.e., are unbiased with respect to the predictors over the calibration). One particularity of PP approaches is that, by being calibrated on large-scale and local-scale observations, they can then be applied to any climate models to derive local-scale data. For instance to downscale two (or more) different climate models, the PP approaches only need to be calibrated once and thus are not "model-dependent". Several PP methods have been developed in the literature, including regression methods (e.g., Huth, 2004; Vrac et al., 2007; Quesada-Chacón et al., 2021), analogs (e.g., Zorita and von Storch; Walton et al., 2020) or more recently machine learning (e.g., Leinonen et al., 2020; Baño-Medina et al., 2020). For the PP approaches to yield reasonable results, the predictors have to be perfectly simulated by the climate model, which justifies the term "perfect prognosis" (PP). However, in practice, this assumption is often not met (Maraun, 2016).

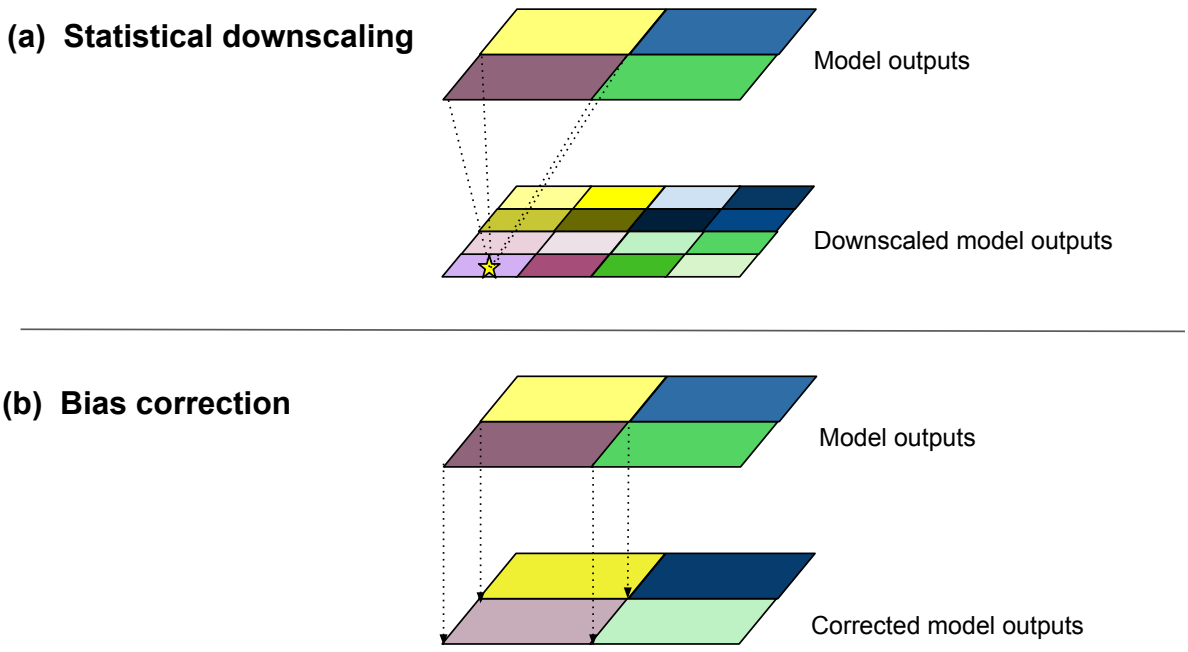

Figure 1.2: Illustrations for (a) statistical downscaling and (b) bias correction of climate model outputs.

Therefore, the MOS approach has been developed. MOS-based techniques aim to find a statistical transformation that directly links simulated variables from climate models (predictor) and the corresponding variables in the reference dataset (predictand). Contrary to the PP approach, predictors (i.e., simulated variables) and predictands (i.e., observations) are not considered to be synchronised in time, and biases relate to differences in statistical distributions or

	Statistical Downscaling	Bias Correction
Data	Predictors $\neq$ Predictands	Predictors = Predictands
Outputs	Local scale	Not necessarily local scale
Approach	"Perfect prognosis"	"Model Output Statistics"
Model-dependent	No	Yes

Table 1.1: Summary of statistical downscaling and bias correction features.

properties. MOS-based approaches consist in directly adjusting the simulated variables so that they have similar statistical properties as the corresponding variables in the reference datasets. Thus, MOS approaches in climate studies are often referred to as *statistical bias correction* methods. Contrary to PP approach, the statistical transformation designed using the MOS approach maps identical physical variables onto each other. For example, bias-corrected temperature is obtained from simulated temperature only without other predictors (such as precipitation) being used. Once calibrated over a reference period, the statistical transformation can be applied to adjust the modelled outputs for other periods (e.g. future) by assuming that the correction is still valid under climate change. However, as the MOS techniques are calibrated directly to link climate models outputs and observations, the derived statistical transformations are "model-dependent": they basically learn to *bias-correct* a single model, and this transformation is not necessarily valid to be applied to other climate models. In practice, MOS techniques are designed to adjust the statistical properties of climate models with respect to a reference dataset having similar spatial resolution (Fig. 1.2b).

Table 1.1 provides a concise summary of the main different attributes that make the statistical downscaling and bias correction methods distinct and that are discussed in the previous paragraphs. As this thesis mainly focuses on bias correction, the rest of the chapter is devoted to a deeper explanation of this particular topic.

Assumptions and limits of bias correction

Before going into the details of the different bias correction (BC) methods based on the MOS approach, here I provide some of the different assumptions that have to be fulfilled to successfully apply bias correction in a climate modelling context.

- First, a key assumption of bias correction is that the relevant processes of the climate system and their changes are reasonably well simulated by the climate model to be adjusted (Maraun et al., 2017). Indeed, this assumption is fundamental as BC methods cannot overcome major model errors, and basic implementation of BC methods tend to preserve some of the biases from the climate models to be adjusted, such as those in the atmospheric fields (Christensen et al., 2008). Before applying any BC methods, end users have to make sure that climate simulations provide reasonable representation of climate.
- Second, another common assumption of BC is that biases are assumed to be time-independent (i.e., do not vary in time). Indeed, bias correction methods often learn to adjust biases over a reference period (e.g. the recent past) and assume this correction to remain valid for other time periods (e.g. projections for the end of the century). This assumption is however questionable given that characteristics of climate can differ between the reference and projection periods, resulting in time-dependent biases. This may call into question the application of BC methods in such contexts.

- Basic implementation of BC methods often rely on pure statistical techniques and are not physically constrained. The underlying assumption of such BC methods is that they are skilled enough to produce realistic corrected outputs. This assumption is again questionable as the lack of physical constraints can result in obtaining corrected outputs with inappropriate physical laws, and, thereby, might be inappropriate.

When these hypotheses are not valid, the bias-corrected output should be handled with great care by end users. The reader is referred to Maraun (2016) for more details.

Bias correction: from univariate to multivariate

Several bias correction methods based on the MOS approach have been developed in order to adjust the statistical discrepancies of the simulations. Most of these BC methods are designed to adjust univariate properties of climate variables, such as the mean (e.g., Xu, 1999), the variance (e.g., Berg et al., 2012) and also all moments of higher order and quantiles (e.g., "quantile-mapping", Haddad and Rosenfeld, 1997). The last technique is certainly the most employed one, since it has been used to adjust many univariate aspects of simulated variables (mean, variance and any quantiles). Its flexible theoretical framework has also allowed the development of multiple versions of quantile-based methods (Panofsky and Brier, 1968; Déqué, 2007; Michelangeli et al., 2009; Gudmundsson et al., 2012; Kallache et al., 2011; Vrac et al., 2012; Tramblay et al., 2013; Cannon et al., 2015), contributing to its large success in various climate applications (e.g., Vigaud et al., 2013; Defrance et al., 2017; Famien et al., 2018; Bartok et al., 2019). However, one major issue of such quantile-mapping methods is that they are univariate: in practice, they are applied to correct separately one physical variable at a time at one grid cell at a time. Although univariate properties are adjusted according to references dataset, the potential biases in simulated multivariate dependence structures are not particularly adjusted (e.g., Maraun, 2013a; Wilcke et al., 2013; Vrac, 2018). By disregarding these potential multivariate biases, univariate BC methods can generate corrections where the dependence structure between variables and sites is misrepresented. These inappropriate multivariate situations can then affect subsequent analyses that depend on multivariate characteristics of climate variables such as in hydrology (e.g., Van de Velde et al., 2022) or, as explained previously for compound events analyses (e.g., Zscheischler et al., 2019). It is therefore of paramount importance to consider multivariate statistical methods that are able to generate corrections with not only appropriate univariate properties but also appropriate dependence structures between climate variables.

Recently, a few multivariate bias correction (MBC) methods have been developed in the literature. Not only do these methods adjust univariate properties, they are also able to correct the multivariate properties of climate simulations (e.g., Bárdossy and Pegram, 2012; Dekens et al., 2017; Nahar et al., 2018; Vrac, 2018; Cannon, 2018; Robin et al., 2019; Nguyen et al., 2019, among few others). As these methods are multivariate by definition, they can be implemented in different dimensional configurations depending on the need of corrections. For example, MBC methods can be applied to adjust several physical variables jointly, but at each individual grid cell. This would result in adjusting inter-variable properties of climate simulations. Another example of applications for MBC could be to adjust for a specific physical variable all grid cells jointly to adjust spatial properties. However, in most studies applying MBC methods, they are applied at each grid cell individually to only adjust inter-variable properties (Cannon, 2018; Meyer et al., 2019; Guo et al., 2019), disregarding spatial adjustments.

These MBC methods are often based on more sophisticated statistical techniques than their univariate counterparts, including rank resampling (e.g., Vrac, 2018; Mehrotra and Sharma, 2019), matrix recorrelation (e.g., Bárdossy and Pegram, 2012; Cannon, 2018) or optimal transport (e.g., Robin et al., 2019). They can also make different assumptions to produce corrected outputs. For example, methods can differ depending on how they consider the simulated non-stationary properties between calibration and projection periods in the correction procedure. All these differences between methods are sometimes not yet fully understood by end users, although being crucial for a thorough understanding of the quality of adjustments provided by MBC methods.

1.5 Objectives of the thesis

The first objective of the thesis is to better clarify the performances of existing multivariate statistical methods for the adjustment of multivariate properties of climate simulations. This will be done through an intercomparison study that will aim to:

- Compare and evaluate existing multivariate BC methods to adjust multivariate properties of climate simulations.
- Evaluate the sensitivity of multivariate BC methods in a highly dimensional context.
- Assess the performance of multivariate BC methods in a non-stationary context.
- Provide a guide for end users to help them choose methods depending on their needs.

The second objective of the thesis is to explore the capabilities of Machine Learning tools to adjust multivariate properties of climate simulations. The development of a new multivariate BC method based on Machine Learning will be carried out for the adjustment of simulated spatial properties only. The questions underlying this objective that will be addressed are:

- Can Machine Learning be exploited for multivariate BC?
- Can Machine Learning do better than existing multivariate methods?
- How this new developed method performs in a non-stationary context?

The third objective is to study how to determine the time period at which the probabilities of compound events emerge from natural variability. This will be done by investigating the changes in CEs probability in simulations, evaluating the significance of their changes and quantifying the importance of univariate and dependence properties in these changes. The questions underlying this objective that will be addressed are:

- How can we determine the time of emergence of compound events probabilities?
- How univariate and dependence properties of climate variables leading to compound events contribute to the change of their occurrences?
- Are there differences in time of emergence between climate models?

1.6 Outline of the manuscript

The rest of the thesis is organised as follows: Chapters 2 and 3 cover the subject of bias correction of climate simulations. Chapter 2 examines the benefits and losses of multivariate bias correction methods through an intercomparison study of four existing methods. Chapter 3 explores the potential of Machine Learning techniques for multivariate bias correction and provides a new method for the adjustments of spatial properties based on these innovative tools. Chapters 2 to 3 both contain a published article and additional perspectives.

Then, Chapter 4 focuses on investigating compound events probabilities and their evolutions. It presents a new statistical tool to assess the time of emergence of simulated compound events probability and to quantify the importance of univariate and dependence properties in these changes.

Finally, Chapter 5 will summarise the work carried out and the results obtained. It will also specify some perspectives for future research.

Résumé

Contexte

Ce chapitre introductif a pour but de poser le contexte et la problématique de cette thèse. Le but de cette thèse est de développer de nouveaux outils statistiques multivariés afin de permettre une compréhension plus approfondie du système climatique. L'aspect multivarié de ces outils qui seront appliqués dans cette thèse à des simulations issues de modèles climatiques permettra de prendre en compte de manière appropriée la dépendance entre les variables climatiques, et donc de potentiellement améliorer la compréhension de leurs interactions et leurs évolutions potentielles dans un contexte de changement climatique.

Objectifs : corrections de biais et études des événements composés.

J'applique et développe de nouveaux outils statistiques multivariés pour, d'une part, la correction de biais des simulations climatiques, et d'autre part, l'étude des changements de probabilités d'événements conjoints à forts impacts.
Cette thèse s'organise en 4 chapitres et a principalement pour but de répondre aux questions:

- Comment fonctionne la correction de biais multivariés de simulations climatiques et quels sont les avantages et les inconvénients de l'utilisation des méthodes existantes? (Chapitre 2)
- Les récents outils de Machine Learning peuvent-ils être exploités pour la correction des biais multivariés? (Chapitre 3)
- Comment évaluer la période d'émergence des probabilités des événements composés? Et comment les propriétés univariées et multivariées contribuent-elles aux changements de ces probabilités? (Chapitre 4)

Chapter 2

Intercomparison of MBC methods for climate simulations: which benefits for which losses?

The core of this chapter is an article published in the scientific journal *Earth System Dynamics*. The article is preceded by a preamble giving some methodological information on the work carried out. It will then be completed by a discussion recalling the main conclusions and highlighting some perspectives.

2.1 Preamble and methodological summary

Recently, some multivariate bias correction (MBC) methods have been developed in the literature. However, these methods present differences in their applicability and assumptions that are not yet fully understood by end users. For instance, these methods are based on different statistical techniques to adjust simulated multivariate properties, or make different assumptions about how simulated non-stationary properties are considered. Consequently, the quality of the corrections can vary from one method to another, and it is hence crucial for end users to clarify the advantages and inconveniences associated with using these different multivariate statistical tools.

This chapter aims at comparing the performances of different multivariate bias correction approaches to adjust multivariate properties of climate simulations. In this comparison, a representative sample of MBC methods will be carefully selected based on their differences in terms of statistical techniques used, assumptions and methodologies in order to provide as much information as possible on these different features. In particular, MBC methods can be categorised in three different families depending on how they adjust multivariate properties of climate simulations. The first category is named "marginal/dependence", and gathers MBC methods that adjust separately univariate distributions and dependence structures. The second category of approaches, named "all-in-one" is made up of MBC methods that adjust both univariate and dependence properties altogether at the same time. In the last category called "successive conditional", MBC methods adjust the variables successively and conditionally on the previous adjusted ones. Fig. 2.1 summarises the three different methodologies employed by MBC meth-

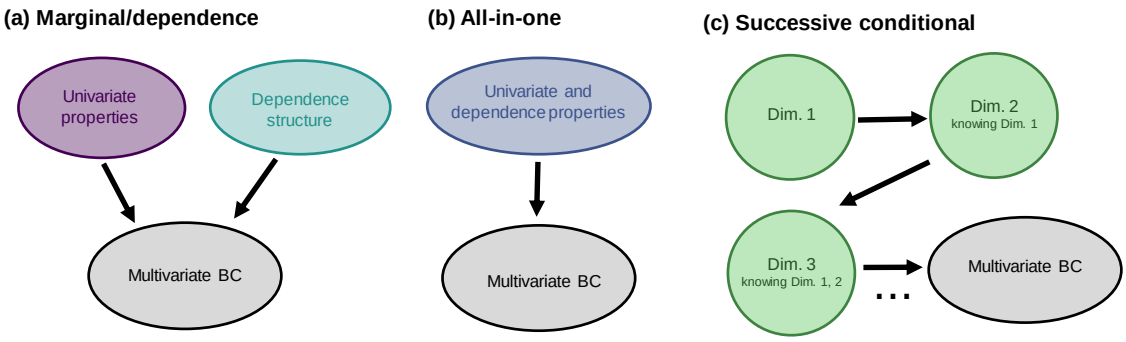

Figure 2.1: The three different families of multivariate bias correction methods.

ods to adjust multivariate properties of climate simulations. More details on these categories will be provided later in Section 1 of the article.

In the following study, four MBC methods — R^2D^2 (Vrac, 2018), dOTC (Robin et al., 2019), MBC-n (Cannon, 2018) and MRec (Bárdossy and Pegram, 2012) — will be applied to adjust simulated temperature and precipitation time series over France and three subregions for the 1979-2016 period. In addition, the univariate BC method named CDF-t (Vrac et al., 2012) will be implemented and used as a benchmark to measure the benefits of considering multivariate aspects in the correction process. For each MBC method, three dimensional configurations — with different objectives of corrections for multivariate properties — will be implemented (Fig. 2.2):

- a 2-dimensional ("2d-") version, which consists in applying the MBC method independently at each grid cell but to jointly adjust both temperature and precipitation time series. Thus, the 2d-version aims to adjust inter-variable properties at each grid cell (Fig. 2.2a).
- a spatial-dimensional ("Spatial-") version, for which the MBC method is applied to jointly adjust all time series for a specific physical variable but independently from the other physical variable. By doing so, Spatial-versions aim to adjust simulated spatial properties for each physical variable separately (Fig. 2.2b).
- a full-dimensional ("Full-") version, where all temperature and precipitation times series are adjusted jointly over the entire domain. The Full-version hence aims to adjust together inter-variable and spatial properties of climate simulations (Fig. 2.2c).

While most of the studies apply MBC methods grid cell by grid cell to adjust inter-variable properties of climate simulations (e.g., Meyer et al., 2019; Guo et al., 2019; Van de Velde et al., 2022), including the Spatial- and Full-versions in the study will permit to assess the potential sensitivity of multivariate BC methods to produce realistic corrections in a high-dimensional context.

As a reminder, BC methods will be applied to adjust time series over the 1979-2016 period. This period will be divided into two intervals of 19 years: 1979-1997 and 1998-2016, and two protocols will be defined to assess the performances of MBC methods. The first protocol will consist in using the 1979-1997 portion as the calibration period for MBC methods to derive corrections for the 1998-2016 projection period. Then, swapping of the two periods will be

performed so that each period is used once for calibration and projection. Using this protocol will permit to evaluate the global performances of MBC methods in adjusting multivariate properties (e.g., Cannon, 2018). The second protocol will consist in applying MBC methods without cross-validation: both the 1979-1997 and 1998-2016 periods will be corrected by using the 1979-1997 portion as calibration period. By doing so, the potential simulated climate change signal will not be distorted, which will allow to evaluate how simulated non-stationary properties are considered by BC methods.

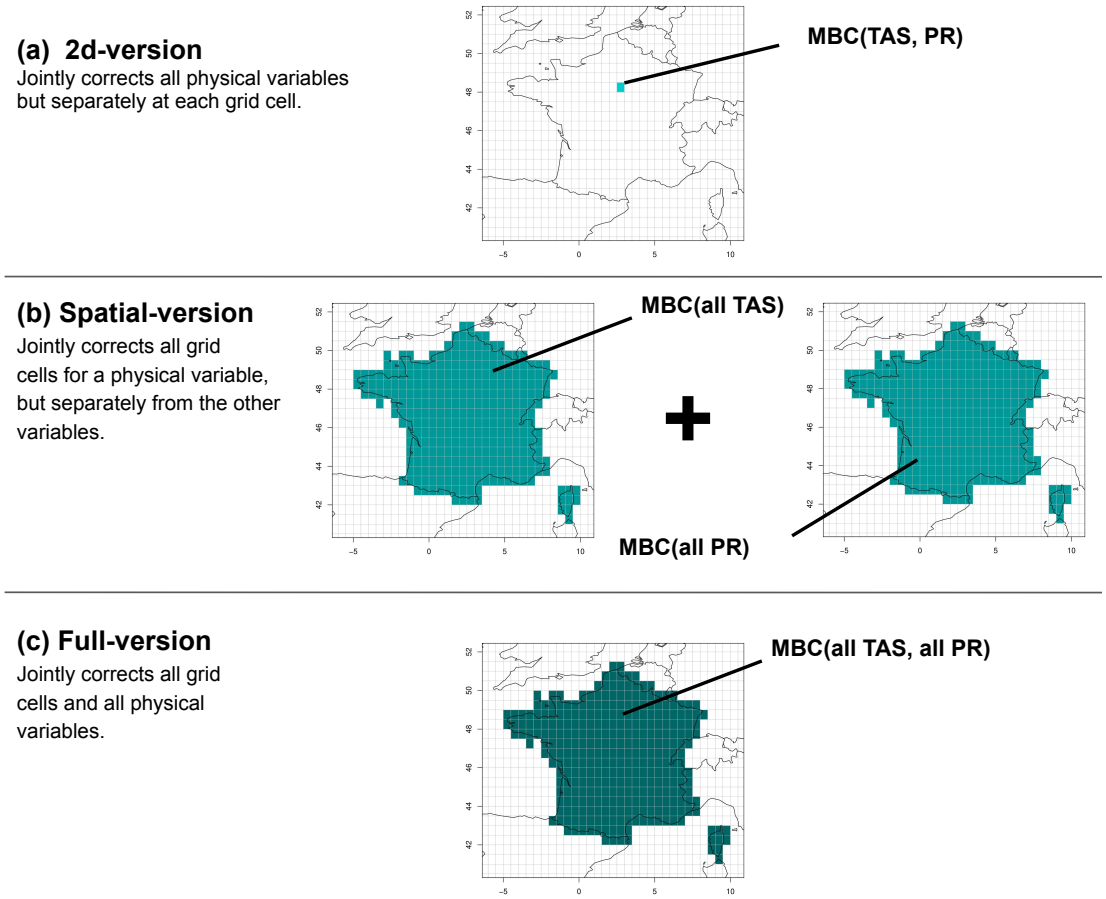

Figure 2.2: The three different dimensional configurations implemented in the article to perform multivariate bias correction: (a) 2d-, (b) Spatial-, and (c) Full-versions.

The choice to include precipitation in this study was motivated by the fact that precipitation shows strong spatial and temporal variability and is often poorly represented by climate models while being the main driver of hydrological processes (e.g., Ehret et al., 2012). Evaluating the performances of multivariate BC in providing realistic precipitation time series is therefore a crucial issue for scientists working on impacts, e.g., in hydrology. Adjusting precipitation from climate simulations in this intercomparison study would permit to evaluate multivariate BC methods in a challenging context, as important corrections are likely to be required.

I emphasise that through this study, I am not trying to find the "best" MBC method. Indeed, this effort would be in vain as there is probably no single method that can "perfectly" correct all statistical properties of climate simulations at the same time (Ehret et al., 2012). The study rather aims at drawing general recommendations to help end users in the choice of bias correction methods that best meet their needs for their applications. The evaluation of the methods will

be conducted from a climatological point of view using relevant statistical metrics to analyse the adjustments of univariate, inter-variable, spatial, temporal and non-stationary properties. It will provide an intercomparison framework to which other MBC methods, e.g. relying on different assumptions, could be easily added. This evaluation work will also permit to identify potential avenues for the development of new MBC methods.

2.2 Article published in *Earth System Dynamics*: Multivariate bias corrections of climate simulations: which benefits for which losses?

Earth Syst. Dynam., 11, 537–562, 2020
https://doi.org/10.5194/esd-11-537-2020

Earth System Dynamics
Open Access
EGU

Multivariate bias corrections of climate simulations: which benefits for which losses?

Bastien François[1], **Mathieu Vrac**[1], **Alex J. Cannon**[2], **Yoann Robin**[3], **and Denis Allard**[4]

[1]Laboratoire des Sciences du Climat et l'Environnement (LSCE-IPSL) CNRS/CEA/UVSQ, UMR8212, Université Paris-Saclay, Gif-sur-Yvette, France
[2]Climate Research Division, Environment and Climate Change Canada, Victoria, BC, Canada
[3]Centre National de Recherches Météorologiques, Université de Toulouse, CNRS, Météo-France, Toulouse, France
[4]INRAE, BioSP, 84914, Avignon, France

Correspondence: Bastien François (bastien.francois@lsce.ipsl.fr)

Received: 2 March 2020 – Discussion started: 6 March 2020
Revised: 8 May 2020 – Accepted: 11 May 2020 – Published: 15 June 2020

Abstract. Climate models are the major tools to study the climate system and its evolutions in the future. However, climate simulations often present statistical biases and have to be corrected against observations before being used in impact assessments. Several bias correction (BC) methods have therefore been developed in the literature over the last 2 decades, in order to adjust simulations according to historical records and obtain climate projections with appropriate statistical attributes. Most of the existing and popular BC methods are univariate, i.e., correcting one physical variable and one location at a time and, thus, can fail to reconstruct inter-variable, spatial or temporal dependencies of the observations. These remaining biases in the correction can then affect the subsequent analyses. This has led to further research on multivariate aspects for statistical postprocessing BC methods. Recently, some multivariate bias correction (MBC) methods have been proposed, with different approaches to restore multidimensional dependencies. However, these methods are not yet fully apprehended by researchers and practitioners due to differences in their applicability and assumptions, therefore leading potentially to different results. This study is intended to intercompare four existing MBCs to provide end users with aid in choosing such methods for their applications. For evaluation and illustration purposes, these methods are applied to correct simulation outputs from one climate model through a cross-validation method, which allows for the assessment of inter-variable, spatial and temporal criteria. Then, a second cross-validation method is performed for assessing the ability of the MBC methods to account for the multidimensional evolutions of the climate model. Additionally, two reference datasets are used to assess the influence of their spatial resolution on (M)BC results. Most of the methods reasonably correct inter-variable and intersite correlations. However, none of them adjust correctly the temporal structure as they generate bias-corrected data with usually weak temporal dependencies compared to observations. Major differences are found concerning the applicability and stability of the methods in high-dimensional contexts and in their capability to reproduce the multidimensional changes in the model. Based on these conclusions, perspectives for MBC developments are suggested, such as methods to adjust not only multivariate correlations but also temporal structures and allowing multidimensional evolutions of the model to be accounted for in the correction.

Published by Copernicus Publications on behalf of the European Geosciences Union.

1 Introduction

Representing precisely the climate system and the interactions between its components is a major challenge not only for climate modellers but also for scientists working on impact, mitigation and adaptation issues relating to climate change. Indeed, it is now common that climate change impact studies, e.g., in hydrology, environmental science or economics, use global and regional climate model (GCM and RCM) simulations as inputs into impact models. However, in spite of continued scientific progress in climate modeling, climate simulations often remain biased compared to observations (Christensen et al., 2008). This means that their statistical attributes such as mean, variance, extreme or even dependence structures between several variables and/or sites can differ from those calculated based on historical records. Therefore, using plain simulations can significantly affect the results of impact studies.

To solve this issue, many statistical bias correction (BC) methods have been developed, in order to correct the statistical discrepancies of the simulations before climate change assessment studies. Most of the BC methods in use are designed to adjust univariate distribution features of climate variables, such as the mean (e.g., Delta method, Xu, 1999), the variance (e.g., simple scaling adjustment, Berg et al., 2012) or quantiles (e.g., "quantile-mapping", Haddad and Rosenfeld, 1997). The last technique received notable success, since it permits the adjustment of the mean, the variance and any quantile of the climate variables. Its theoretical framework has been conducive to the development of multiple versions of quantile-based methods (e.g., Panofsky and Brier, 1958; Déqué, 2007; Gudmundsson et al., 2012; Vrac et al., 2012). However, all these univariate BC methods are designed to correct variables independently, i.e., are applied separately for each physical variable at each specific location (e.g., grid cell). Although univariate distribution features are adjusted according to references, it can generate inappropriate multivariate situations where the dependence structure between variables and sites is not corrected from the model and misrepresented (Maraun, 2013) or even modified. Ignoring the observed inter-variable and intersite dependencies in the correction procedure can result in obtaining corrected outputs with inappropriate physical laws and, thereby, distorting the results of impact studies (Zscheischler et al., 2019). It is therefore of paramount importance to adjust the dependence structures of climate simulations, in addition to 1-dimensional characteristics, before using it in subsequent studies.

These methodological issues have led up to the recent development of a few multivariate bias correction (MBC) methods. Not only do these methods adjust univariate distribution features, they also are aimed at correcting the dependence structure of climate simulations. Recent studies have shown that univariate BC methods can already provide adequate results for certain specific regional impact studies (Yang et al., 2015; Casanueva et al., 2018) and that using MBC methods does not necessarily present substantial benefits (Räty et al., 2018). However, this does not call into question the interest of MBC methods as these specific results cannot be generalized to each method and application. In particular, MBC methods could be valuable in larger-scale impact modeling frameworks such as compound events, where the combination of physical processes across multiple spatial and temporal scales leads to significant impacts (Zscheischler et al., 2018). As mentioned by Vrac (2018) and completed by Robin et al. (2019), MBC methods may be grouped into three main categories of approaches: the "marginal/dependence" correction approach, the "successive conditional" correction approach and the "all-in-one" correction approach. The marginal/dependence category is made up of multivariate bias adjustment methods correcting separately the marginal distributions and the dependence relationships of climate simulations (e.g., Bárdossy and Pegram, 2012; Mehrotra and Sharma, 2016; Vrac, 2018; Nahar et al., 2018; Cannon, 2018a). In the all-in-one category, multivariate BC methods correct the 1-dimensional marginal properties and dependence structures altogether at the same time (e.g., Robin et al., 2019). At last, successive conditional MBC methods perform successive corrections, conditionally on the variables already corrected (e.g., Bárdossy and Pegram, 2012; Dekens et al., 2017). In particular, this last category has two major limitations. First, the quality of the correction can change depending on the ordering of the variables to correct (see, e.g., Piani and Haerter, 2012). Second, the number of variables already corrected increases at each iteration step, which progressively reduces the number of data available for the correction, making it less and less robust. Accordingly, these methodological limits call into question the applicability of successive conditional BC methods for multivariate bias adjustment of high-dimensional climate simulations.

Additionally to the methodological distinction described above, the few existing multivariate BC methods are based on the use of different statistical techniques. They may also present differences in terms of assumptions and philosophical features, e.g., deterministic versus stochastic. Consequently, the quality of the correction outputs can vary largely from one method to another, depending on their characteristics. It is hence crucial, in particular for end users, to carefully evaluate the suitability of these multivariate BC methods and identify their advantages and limits, not only between the different categories of methodological approaches but also between the different statistical techniques and assumptions. In this study, we present an analysis of four multivariate BC methods and assess their performances in terms of adjustment of dependence structures for temperature and precipitation time series. We focus in particular our intercomparison on methods belonging to the marginal/dependence and the all-in-one categories. Due to the previously mentioned limitations of the successive conditional approach, no methods

belonging to this category are investigated. The selected four MBC methods present differences in terms of conceptual features, statistical techniques used and assumptions. In particular, MBCs with different assumptions about nonstationarity are selected, i.e., differing in how they consider the simulated multidimensional changes between present (i.e calibration) and future (i.e., projection) periods in the correction procedure. Moreover, in order to assess the potential benefits of using multivariate BC methods relative to univariate ones, one univariate quantile-mapping-based BC method is included in the study as a benchmark. It provides a more extensive intercomparison framework in which quality of BC outputs can be assessed and compared by evaluating univariate, inter-variable, spatial and temporal properties, as well as multidimensional changes.

In addition, each BC method is applied to correct climate model outputs over France and three subregions according to two distinct reference datasets with different spatial resolutions. This permits one to assess the potential influence of the reference spatial resolution on bias correction results and to delineate guidance on relevant good practices for end users concerning this aspect.

This paper is organized as follows: Sect. 2 describes the model and reference data used, and Sect. 3 presents the BC methods intercompared. Then, Sect. 4 presents the experimental setup used in this study, while Sect. 5 displays the results of the intercomparison. Finally, our findings are summarized, discussions are given and perspectives for future research are proposed in Sect. 6.

2 Model and reference data

Institut Pierre-Simon Laplace (IPSL) coupled model (Marti et al., 2010; Dufresne et al., 2013) daily data with a 1.25° by 2.5° spatial resolution are used in this study as model data to be corrected. Simulations of the scenario of atmospheric CO_2 concentration pathway associated with a radiative forcing of $+8.5\,W\,m^{-2}$ (RCP 8.5 scenario, i.e., the scenario with highest CO_2 concentration) are selected. Daily temperature (T2) and precipitation (PR) time series from 1 January 1979 to 31 December 2016 are extracted over the geographical area of France ([42, 51° N] × [−5, 10° E]), which corresponds to 321 continental grid cells.

As BC methods require a reference dataset to adjust the simulations, daily temperature and precipitation time series with a 0.5° by 0.5° spatial resolution are first used from the "WATCH Forcing Data methodology applied to ERA-Interim data" (WFDEI) from the EU WATCH project (Weedon et al., 2014) over the same geographical area of France. Note that, as spatial resolution between WFDEI and IPSL-CM5 are different, IPSL model data are regridded by a nearest-neighbor technique to associate each IPSL grid cell with its nearest WFDEI grid cell center. Hence, in the following, the IPSL data will be used at the 0.5° spatial resolution corresponding to that of the WFDEI reference dataset.

To assess the influence of the reference spatial resolution on BC results, we use another reference gridded dataset for France with finer resolution: the "Systeme d'Analyze Fournissant des Renseignements Atmosphériques à la Neige" (SAFRAN) reanalysis dataset (Vidal et al., 2010). Daily T2 and PR time series from SAFRAN have a 8 km × 8 km spatial resolution and divide France into 8981 continental grid cells. IPSL data are regridded to the 8 km × 8 km SAFRAN resolution using the nearest-neighbor technique. Once regridded IPSL simulations are obtained, each MBC method can be applied. However, as some MBC algorithms have difficulties in practice in very high-dimensional contexts (here for 8981 grid cells), we restrict the application of MBCs with SAFRAN reference dataset over the Brittany region, located in the northwest part of France ([47, 49° N] × [−5, 2° E]), which corresponds to 345 continental grid cells. Note that we selected this subregion of Brittany for SAFRAN, i.e., at fine resolution, in order to have a similar number of grid cells as for France selected with the WFDEI reference dataset, i.e., at coarser resolution. MBC methods have also been applied and evaluated over two others subregions of 345 grid cells located, respectively, around the Paris area and in southeast France. For the sake of clarity, as same results were obtained for each of the subregions, we will only present the results for Brittany in the rest of this study.

3 Multivariate bias correction methods

This section presents a brief description of the univariate BC method and the four multivariate BC methods implemented in this study. As a reminder, results from the univariate CDF-t method serve as a benchmark to measure the benefits of considering multivariate aspects in the correction procedure instead of using univariate BC methods. For the sake of clarity, Table 1 provides a concise summary of the different attributes that make the BC methods distinct.

3.1 Cumulative Distribution Function – Transform (CDF-t)

The "Cumulative Distribution Function – Transform" (CDF-t) method is a univariate BC method initially proposed by Michelangeli et al. (2009) to correct the univariate distribution of a modeled climate variable. Since then, CDF-t has been applied for various studies (e.g., Tramblay et al., 2013; Tobin et al., 2015; Defrance et al., 2017; Famien et al., 2018; Guo et al., 2018) and specific variants have been developed (e.g., Kallache et al., 2011; Vrac et al., 2016). The CDF-t approach applies, independently to each variable, a univariate transfer function T, which permits one to link the cumulative distribution function (CDF) of a variable of interest in the model simulations to that of the reference dataset. By assuming that T is valid in a climate different from that of the

Table 1. Summary of attributes of the different bias correction methods. Not-applicable (n/a) is indicated when the statement in rows does not apply.

Characteristics	CDF-t	R^2D^2	dOTC	MBCn	MRec
Type of BC	1d-BC	MBC	MBC	MBC	MBC
Category of MBC	n/a	Marginal/ dependence	All-in-one	Marginal/ dependence	All-in-one
Statistical technique	Nonstationary quantile mapping	Conditional resampling	Optimal transport	Iterative partial matrix recorrelation	Matrix recorrelation
Dependence structure	$\sim$ same as the model	$\sim$ same as the reference	Allows changes in the dep. struct.	Allows changes in the dep. struct.	Allows changes in the *Gaussian* dep. struct.
Conceptual feature	Deterministic	Deterministic and stochastic	Stochastic	Deterministic and stochastic	Deterministic

calibration period, a new CDF for the bias-corrected variable over the projection period is generated. Then, a quantile–quantile approach is performed between the new (reference) CDF and the CDF from the model simulations during the projection period to derive bias-corrected data. This two-step procedure permits one to take into account potential changes (between calibration and projection periods) of the univariate distribution in the correction procedure. For the special case of precipitation, the "Singularity Stochastic Removal" version of CDF-t (Vrac et al., 2016) is applied to correct both precipitation occurrences and intensities. More details about CDF-t can be found in Appendix A. In the following subsections, the four MBC methods are presented.

3.2 Rank Resampling For Distributions and Dependences (R^2D^2)

The "Rank Resampling For Distributions And Dependences" (R^2D^2) method, developed by Vrac (2018) in the context of marginal/dependence category, is an extension of the "Empirical Copula – Bias Correction" (EC-BC; Vrac and Friederichs, 2015) method. R^2D^2 is based on a reordering technique named the Schaake Shuffle. Originally described by John C. Schaake in 2002, it was introduced in the scientific literature by Clark et al. (2004) to postprocess temperature and precipitation forecasts from numerical weather prediction models. This shuffling technique consists of reordering a sample such that its rank structure corresponds to the rank structure of a reference sample and, thus, allows the reconstruction of multivariate dependence structures. The Schaake Shuffle has already been applied for various applications in climate science, such as ensemble postprocessing (e.g., Möller et al., 2013; Schefzik et al., 2013), and in numerous studies (e.g., Voisin et al., 2010; Verkade et al., 2013). According to the marginal/dependence category to which it belongs, the R^2D^2 method performs first a univariate correction to adjust the marginal distribution of each climate variable. In this study, CDF-t is used for this first step, but it has to be noted that other univariate methods can be employed. Instead of directly applying the Schaake Shuffle and reproducing the temporal structure of the reference (as in Vrac and Friederichs, 2015), the method introduces some variability to the timing of the events, by allowing for the possibility to select a "reference dimension" for the Schaake Shuffle, i.e., one physical variable at one given site, for which rank chronology remains unchanged. Reconstruction of inter-variable and spatial correlations of the reference is then performed using the Schaake Shuffle with the constraint of preserving the rank structure for the reference dimension. Note that the R^2D^2 method can generate as many corrections as the number of variables to be corrected and all with identical inter-variable and spatial dependencies but with different temporal structures, depending on the selected reference dimension. Hence, R^2D^2 introduces some stochasticity in the bias correction. For practical reasons, in the following, we will reduce the number of corrected outputs: only R^2D^2 corrections with reference dimensions located either in Paris or in the center of Brittany (respectively, for France and Brittany regions) will be analyzed in Sect. 5. It must also be noted that by using the Schaake Shuffle technique, R^2D^2 assumes by construction the inter-variable and spatial dependence structures (i.e., the rank correlations, or copulas) to be stable in time. Some more mathematical details about R^2D^2 are expressed in Appendix B.

3.3 Dynamical Optimal Transport Correction (dOTC)

The "Dynamical Optimal Transport Correction" (dOTC) method was developed by Robin et al. (2019), in the all-in-one category, i.e., correcting the marginal distributions and dependence structures altogether at the same time. Based on optimal transport theory, it is a generalization of the univariate quantile mapping techniques to the multivariate case. dOTC is aimed at constructing a multivariate transfer function, called a transport plan, to perform bias correction by minimizing a cost function associated with the transforma-

tion of a multivariate distribution to another. Multivariate distribution of a biased random variable and its correction are linked together through this particular transfer function, where for any value of the variable to correct is associated a conditional law linking the biased value and its correction. Corrections are then picked (partially) randomly from these conditional laws, introducing some stochasticity into the bias correction procedure.

As for univariate quantile mapping, the way the transfer function is constructed for dOTC plays a decisive role in the obtained bias correction outputs. As explained before, the univariate method CDF-t is able to learn the change in univariate distributions between the calibration and the projection periods for the climate model and transfers this change to the observational world. Following this philosophy in a multivariate context, dOTC is designed to learn not only the change in univariate distributions but also the change in multidimensional properties of the model and allows them to be transferred the corrections. Contrary to R^2D^2, it assumes nonstationarity of the dependence (copula) structure between the calibration and the projection periods and permits the evolution of the model (e.g., induced by climate change) to be taken into account in the bias correction procedure. More explanations about dOTC are given in Appendix C.

3.4 Multivariate Bias Correction with *N*-dimensional probability density function transform (MBCn)

The "Multivariate Bias Correction with *N*-dimensional probability density function transform" (MBCn) was developed by Cannon (2018a) in the context of the marginal/dependence category. Based on an adaptation of an image processing algorithm used to transfer color information, MBCn permits one to transfer statistical characteristics of a reference multivariate distribution to the multivariate distribution of climate model variables. Being part of the marginal/dependence category, univariate distributions of climate variables are first adjusted using a 1-dimensional BC (1d-BC) method. For this step, MBCn uses the quantile-delta mapping method (QDM; Cannon et al., 2015) that preserves absolute or relative changes in quantiles, e.g., for, respectively, variables like temperature or ratio variables like precipitation. Once univariate distributions are corrected, the dependence structure is adjusted by using an iterative process. At each step, data are multiplied by random orthogonal rotation matrices to partially decorrelate the climate variables to correct. QDM corrections are then applied on (partially) decorrelated data before the recorrelation step with the inverse random matrices. This step (i.e., including rotation, QDM corrections and back rotation) is repeated iteratively until convergence is reached between the multivariate distributions of reference and climate simulations during the calibration period. Indeed, those iterations permit correcting the dependence structure of the model. Moreover, by doing so – and similarly to dOTC – MBCn allows changes in the dependence structure to be in accordance with the model changes. More details about MBCn can be found in Appendix D.

3.5 Matrix recorrelation (MRec)

Bárdossy and Pegram (2012) presented an MBC, hereafter referred to as "matrix recorrelation" (MRec). The latter lies in the all-in-one category and relies on a matrix recorrelation technique. The MRec method consists of first transforming separately each variable of both model and references to the univariate normal distribution with Gaussian quantile–quantile method. This transformation step is particularly appropriate for variables with mixed distributions (e.g., precipitation composed of wet and dry days), for which computing a Pearson correlation matrix on Gaussianized data instead of raw data permits their dependence structure to be better described. Then, a combination of "decorrelation" and "recorrelation" steps using decompositions of correlation matrices through singular value decomposition (SVD, Beltrami, 1873; Jordan, 1874a, b; Stewart, 1993) is applied on the Gaussianized model data, forcing its Pearson correlation matrix to match that of the Gaussianized observed data during the calibration period. For the projection period, the same "decorrelation–recorrelation" matrix is directly applied on Gaussianized model data, which permits the preservation of, for the projection period, the potential changes in correlations as simulated by the model. Finally, for both periods, a quantile–quantile back transformation is applied separately for each variable between recorrelated variables and references to correct marginal distributions. See Appendix E for more details.

Contrary to the R2D2, dOTC and MBCn methods presented previously, MRec differs in being designed to correct only a particular feature of the multivariate dependence structure, here Pearson correlations. Implicitly, it makes the assumption that Pearson correlation values are sufficient to determine the full multivariate dependence structure, which can be called into question for variables with skewed and heavy tailed distributions (like precipitation) and with potentially complex interactions that Pearson correlation cannot capture as a whole. For this reason, implementing the MRec algorithm in the present intercomparison study permits the comparison of the performances of an MBC method based on such an assumption relative to methods intended to correct the non-Gaussian dependence structure of climate simulations.

4 Design of Experiments

4.1 Settings of MBCs

Multivariate BC methods can be implemented in different dimensional configurations, depending on the need of the users to correct inter-variable and/or spatial correlations. However, in most cases, multivariate BC methods are applied grid cell

by grid cell by practitioners to correct inter-variable properties of climate simulations, disregarding spatial structures (e.g., in Meyer et al., 2019; Guo et al., 2019). We not only tested and assessed this approach for each method but also expanded the study to include high-dimensional configurations of MBC to adjust spatial and full (i.e., spatial and inter-variable jointly) dependence structures of climate simulations. Depending on the dimensional configurations, the objectives of corrections for multivariate properties differ. Including different dimensional versions in the study will permit one to better highlight the potential losses and benefits associated with them. Therefore, in the following each of the four MBC methods is applied according to the three following configurations:

- a 2-dimensional (hereinafter referred to as "2d-") version, for which the MBC method is applied independently at each grid cell but jointly corrects both temperature and precipitation time series. For example, to correct a climate dataset of 321 grid cells, the MBC method is performed 321 times, i.e., for each grid cell across the whole grid. By doing so, 2d- versions are aiming to correct inter-variable correlations within each grid cell.

- a spatial-dimensional (hereinafter referred to as "Spatial-") version, where all time series for a particular physical variable are corrected jointly but independently from the other physical variable. Hence, for this version, the MBC method is performed twice, adjusting, on the one hand, all time series for temperature and, on the other hand, all time series for precipitation. Thus, Spatial- versions are designed to adjust spatial correlations of climate models for each physical variable separately.

- a full-dimensional (hereinafter referred to as "Full-") version, where all time series are corrected jointly over the entire grid for both temperature and precipitation. The MBC method is hence applied only once and is intended to correct together the inter-variable and spatial correlations of the simulations.

Regarding the initial settings for MBCn, preliminary tests have been conducted with different dimensional settings to find the number of iterations ensuring the convergence of the algorithm depending on the dimensional configuration. With respect to the results of these tests (not shown), the number of iterations has been chosen to be equal to 50 for 2d- configurations and 200 for both Spatial- and Full- versions.

4.2 Protocols of bias correction

In this study, the BC methods presented above are applied to correct IPSL GCM simulations with either the WFDEI ($0.5° \times 0.5°$) or the SAFRAN ($8\,\text{km} \times 8\,\text{km}$) data as references. Data are available for the period 1979–2016, i.e., 38 years, and are divided into two intervals of 19 years: 1979–1997 and 1998–2016. As a reminder, daily temperature and precipitation times are corrected on 321 and 345 grid cells for France and Brittany regions, respectively. For each method, bias correction is performed separately for the 12 months in order to preserve seasonal properties.

The first protocol in this study takes advantage of the cross-validation technique to generate bias-corrected outputs for the period 1979–2016. Dividing the time period into two parts permits one to perform a 2-fold cross-validation procedure: the 1979–1997 period is first defined as the calibration period, and the 1998–2016 portion, called the projection period, is used for out-of-sample validation. Swapping of the two periods is then done, so that each period has been used once for calibration and once for validation. Bias correction for the period 1979–2016 is then achieved by assembling the adjusted outputs for the projection periods obtained at each step. This 2-fold protocol, largely used in the climate science literature (e.g., in Cannon, 2018a), allows one to reduce overfitting by using two distinct subperiods and is hence well suited to evaluate our results. However, by adjusting the period 1979–1997 according to the 1998–2016 period, this protocol presents the drawback of potentially hiding the climate change signal present in the model. Thus, proper assessment of the multidimensional properties evolutions cannot be conducted via this procedure.

Hence, to evaluate the nonstationary behavior of BC methods, a second protocol is defined. Similarly to the first protocol, the 1998–2016 period is corrected by using the 1979–1997 portion as calibration period. However, here, 1979–1997 simulations are corrected directly with respect to the 1979–1997 references, i.e., without cross-validation. Hence, the potential climate change signal is not distorted by undesirable effects resulting from the protocol procedure, allowing for the appropriate assessment of change aspects of the BC methods between the two periods.

In accordance with common practice, thresholding of 1 mm for precipitation time series is applied before evaluation to replace values lower than 1 mm by 0 after correction.

5 Results

The correction outputs are evaluated according to different characteristics designed to focus on (i) marginal, (ii) inter-variable, (iii) spatial, (iv) temporal and (v) nonstationary properties. Characteristics (i)–(iv) are evaluated on the 1979–2016 period for the adjusted outputs obtained according to the 2-fold protocol and are compared to those from the reference dataset. However, regarding nonstationary properties, corrected outputs from the second protocol are used, and results are compared to the simulations to highlight the performances of the MBC methods regarding their capability to reproduce (or not) the multidimensional changes in the model between the 1979–1997 and 1998–2016 periods.

In the following, evaluation is presented for the winter season (December–January–February) only, as conclusions remain generally the same for the other seasons. However, in order to provide nuances, additional results for the summer season (June–July–August) are displayed in the Supplement when needed.

5.1 Univariate distributions properties

First, bias-corrected data are evaluated relative to univariate statistics. To do so, for temperature and precipitation, the difference of mean values between the bias correction and the reference at each grid cell is computed. The same computation is also made for standard deviation. Absolute difference is calculated for temperature mean, while relative difference is more appropriate for precipitation mean as well as for standard deviation of both physical variables. Results are shown with boxplots for the plain IPSL simulations and for a selection of BC outputs in Fig. 1 for France during the winter season. The results for Brittany during winter are presented in Fig. S1 of the Supplement. As marginal/dependence MBC methods correct univariate properties independently from the dependence structure, results for their 1-dimensional characteristics are equivalent between the three different dimensional configurations (2d-, Spatial- and Full-). Therefore, to avoid redundancy, results for R^2D^2 and MBCn are presented for only one arbitrary dimensional configuration, the other configurations giving the exact same mean and standard deviation results. Clearly, Fig. 1 shows large differences between the IPSL simulations and the references for both temperature and precipitation and illustrates the necessity to adjust 1-dimensional distributions of the model before using it in subsequent analyses. Multivariate BC methods implemented in this study display different performances in adjusting the univariate properties. In agreement with the properties of the marginal/dependence MBC methods, R^2D^2 and MBCn present exactly the same results as the 1d-BC methods they use, i.e., respectively, CDF-t (shown) and QDM (not shown). With regard to the performances of dOTC and MRec, some instabilities are found relative to the dimensional configuration. For dOTC, increasing the number of dimensions to correct from 2d- to Full- seems to have a slight but non-negligible cost on the correction of mean and standard deviation (Fig. 1b and c). However, depending on both the climate variable and the statistical feature, the increasing deterioration with respect to the dimensional setting is not systematically observed, as it can be seen in Fig. 1a and d. Concerning MRec, a slight deterioration of correction is often observed from 2d- to Spatial- versions (Fig. 1b, c and d). Regarding the Full- version, the MRec algorithm produces results that are clearly unsatisfactory. Instead of improving the simulations, Full-MRec corrections strongly degrade the univariate statistics. This underperformance of the MRec method over France appears in a context of high-dimensional correction when the number of available data is not large enough compared to the number of dimensions. In this case, the inverses of high-dimensional sample covariance matrices are a highly biased estimator of the inverse of covariance matrices, which consequently largely affects the quality of the Full-MRec corrections. Anyhow, the increasing degradation, whether it is slight or not, of univariate distribution corrections in high-dimensional contexts is one (undesirable) feature of all-in-one methods, here observed for dOTC and MRec. Indeed, all-in-one methods are designed to adjust both univariate distributions and dependence structure of climate simulations at the same time, which involves a possible deterioration of 1-dimensional marginal distributions during the combined correction process.

For Brittany, the same conclusions hold for R^2D^2, dOTC and MBCn, indicating no particular influence of spatial resolution on the results of the marginal statistics adjustment for these methods. Nevertheless, quite interestingly, for the Full-MRec outputs, the underperformance observed for France is not obtained for Brittany (Fig. S1). A possible reason explaining why Full-MRec version is presenting adequate results on this particular region (and the two other subregions, not shown) concerns the size of its geographical area and will be discussed in more detail in Sect. 6.2.

5.2 Inter-variable correlations

To evaluate inter-variable dependence structure, Spearman correlations between temperature and precipitation are computed at each grid cell to measure the monotonic relationship between the two physical variables. Using rank correlation presents the particularity of not being value dependent; i.e., it measures the dependence between two variables rid of their univariate distributions. As the goal when applying MBC is to adjust not only the univariate distributions but also the dependence structure between the variables of interest, Spearman's correlation is appropriate for this latter aspect. Moreover, this measure does not require any assumption about the distribution of the variables or their statistical relationships. It is hence appropriate for temperature and precipitation studies presenting extreme values and/or a lower bound (Vrac and Friederichs, 2015). The maps of the Spearman correlation differences with respect to the reference – for the IPSL model and the bias-corrected data – are displayed in Fig. 2 for both France and Brittany. Initial maps of Spearman correlations, i.e., without differences with respect to the reference, are also provided in Fig. S2.

For France, the map for the IPSL simulations (Fig. 2b1) indicates strong differences with respect to the WFDEI map (Fig. 2a1). As the univariate CDF-t method does not modify rank sequence of temperature and precipitation time series, it globally conserves both the rank correlation intensities and structures of the IPSL model for each region and does not provide any correction of this aspect (Fig. 2c1). By construction, clear improvements of the inter-variable correlation structure are provided by 2d- versions (Fig. 2d1, g1, j1

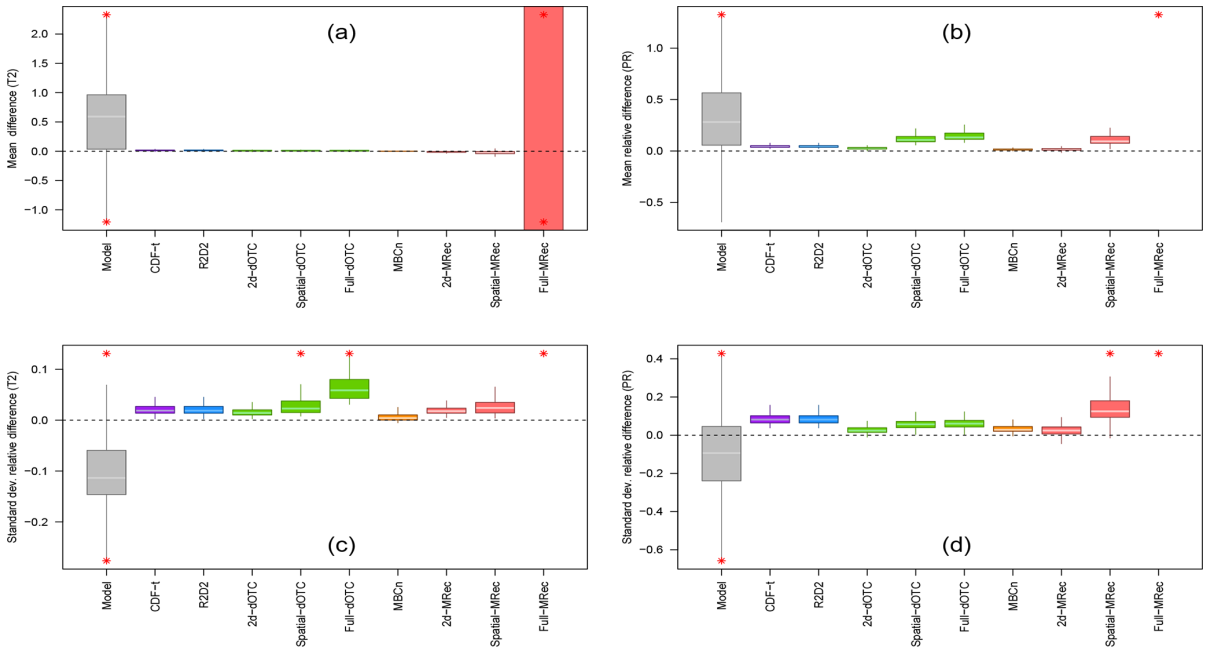

Figure 1. Boxplots of **(a, b)** mean and **(c, d)** standard deviation differences for **(a, c)** temperature (T2) and **(b, d)** precipitation (PR) during winter over the 1979–2016 period for France (WFDEI reference). Results are shown for plain IPSL, CDF-t, R^2D^2, dOTC (2d-, Spatial- and Full- versions), MBC-n and MRec (2d-, Spatial- and Full- versions) outputs. Red asterisks indicate values lying outside the plotted range.

and m1). This is also the case for most of the Full- configurations of MBCs (respectively, Fig. 2f1, i1, l1) despite possible differences in intensities. Note that maps of correlation differences for 2d-R^2D^2 (Fig. 2d1) and Full-R^2D^2 (Fig. 2f1) are identical. Indeed, for the inter-variable aspect, the 2d-version is nested within the Full- configuration (see Vrac, 2018), due to the use of the reordering technique in R^2D^2. Also, for R^2D^2, the choice of the reference dimension does not have any impact on results in the inter-variable context, as it only modifies the rank chronology of time series. As expected from previous explanations, the map for the Full-version of MRec (Fig. 2o1) indicates a strong deterioration of the inter-variable correlation structure. It highlights again the inability of the method to work properly for France in this dimensional setting. Concerning Spatial- versions of MBCs (Fig. 2e1, h1, k1 and n1), as they adjust the whole simulated field of temperature and precipitation separately, they disregard inter-variable relationships. It results in BC outputs with strongly weakened inter-variable correlations structures.

Regarding Brittany, the same conclusions can be drawn for R^2D^2 and dOTC, for which spatial resolution does not affect the results of inter-variable properties adjustment. As noted previously, Full-MRec over Brittany provides more satisfactory results than those obtained over France, which are in line with those obtained for R^2D^2 and dOTC. However, for MBCn outputs, a degrading effect from 2d- (Fig. 2j2) to Full- (Fig. 2l2) is observed, in providing a corrected correlations' structure but with underestimated intensities in the high-dimensional context.

5.3 Spatial correlations

To assess the quality of the corrections in terms of spatial correlations, mean correlograms, i.e., mean Spearman correlation in function of distance, are computed for temperature and precipitation separately after removing daily areal mean. Indeed, climate variables can present a high day-to-day variability that can affect the evaluation of spatial criteria if not removed (e.g., Vrac, 2018).

Figure 3 and S3 show the results obtained for, respectively, precipitation and temperature for the different climate datasets. Note that the choice of the reference dimension for R^2D^2- versions modifies results for temporal criteria and, consequently, for some of the spatial criteria. Hence, in the rest of this work, results from R^2D^2- versions are presented with the reference dimension corresponding to the variable under interest. For the sake of brevity, results for precipitation are mainly discussed in this subsection, and nuances are made when different results are obtained for temperature.

For France, the IPSL precipitation correlogram is fairly distinct from the WFDEI one. The univariate method CDF-t, by simply adjusting univariate distributions, gets closer to the reference dataset (Fig. 3a1), which may be here confusing. Indeed, although CDF-t adjusts the univariate distributions, it

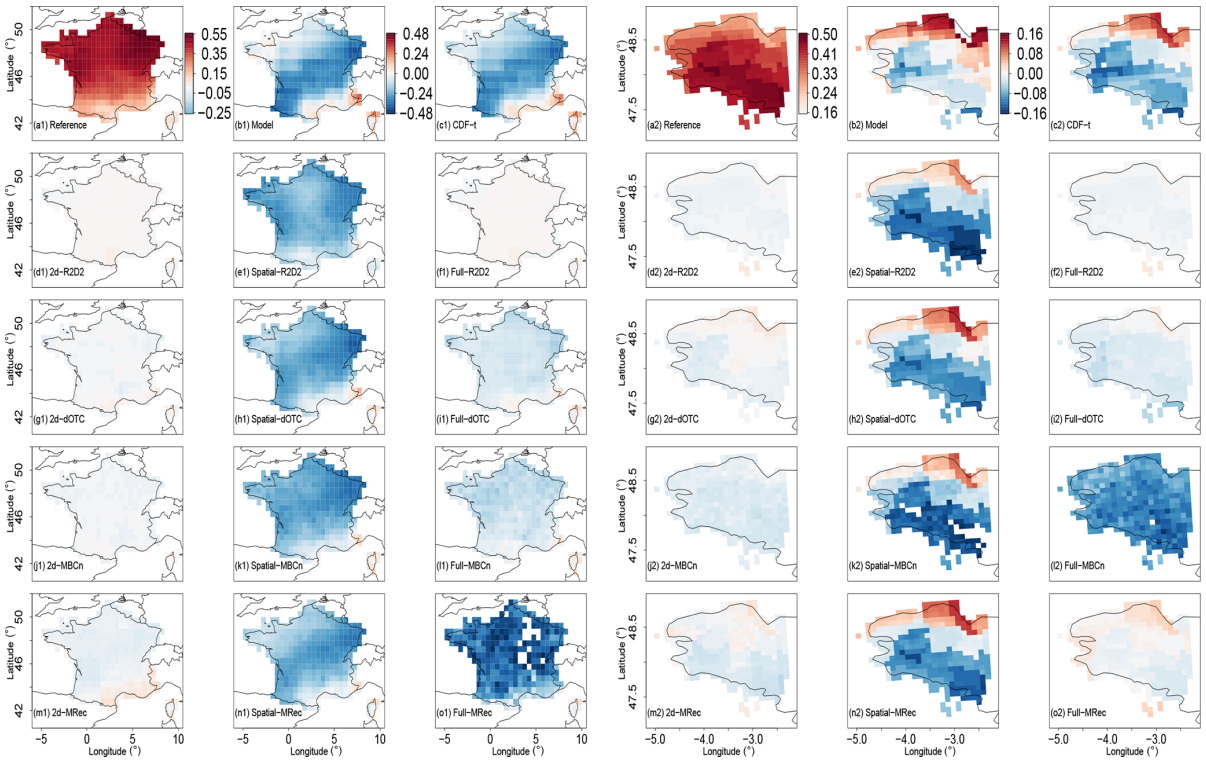

Figure 2. Differences of temperature vs. precipitation Spearman correlation computed at each grid cell for BC methods using **(a1–o1)** WFDEI reference and **(a2–o2)** SAFRAN reference during winter over the 1979–2016 period. Results are shown for reference, plain IPSL, CDF-t, R^2D^2, dOTC, MBC-n and MRec outputs for 2d-, Spatial- and Full- versions. Note that the color scales between panels **(a1)**–**(o1)** and **(a2)**–**(o2)** are not the same to better emphasize intensities of values in the two regions.

is supposed to preserve the rank sequence of the simulations, and therefore spatial correlations are disregarded during the BC procedure. But, as the Singularity Stochastic Removal version of CDF-t (Vrac et al., 2016) is explicitly designed to improve dry days frequency, the method consequently modifies rank correlations, which results here in an improvement of spatial statistics for precipitation. Also, an additional reason is that the correction of the univariate distributions provided by CDF-t associated with the removing of daily areal means modifies ranks of the data, resulting in getting a correlogram closer to that from the reference dataset, and so improves intersite variability.

Correlograms of 2d- versions (dotted) for the four MBC methods (Fig. 3b1, c1, d1 and e1) show results equivalent to CDF-t. Indeed, 2d-configuration MBCs adjust univariate distributions and inter-variable correlations without modifying spatial correlations. The improvements of correlograms for 2d- versions thereby illustrate again that the correction of univariate distributions improves spatial statistics for France. Particularly, 2d-R^2D^2 results (Fig. 3b1) are, by construction, exactly the same as those from CDF-t (Vrac, 2018). Indeed, by construction, 2d-R^2D^2 driven by precipitation preserves Spearman spatial correlations from CDF-t for the precipitation variable. Note that, however, it is definitely not the case for temperature spatial structure (not shown) when 2d-R^2D^2 is driven by precipitation. Indeed, for 2d-R^2D^2 outputs driven by a specific physical variable, spatial structures of the "other" variable are strongly degraded by the reordering step.

Correlograms associated with outputs of Spatial- and Full-versions for R^2D^2 (Fig. 3b1) nicely fit the one from the reference dataset – even at long distances – and provide major improvements in adjusting the spatial properties of the simulations. However, for similar reasons as those explained for 2d-R^2D^2, undesirable degradation effects on spatial cross-correlation between temperature and precipitation are obtained for Spatial-R^2D^2 outputs (not shown). Therefore, it indicates that practitioners must favor the use of Full-R^2D^2 for their applications. With regard to Spatial- and Full-dOTC (Fig. 3c1) and Spatial-MRec (Fig. 3e1), although correlograms are very close to those from the reference dataset, they provide slightly less pronounced improvements compared to the 2d- versions, suggesting a slight degrading effect on results for these methods by considering more variables in the correction. As expected, the correlogram associated with

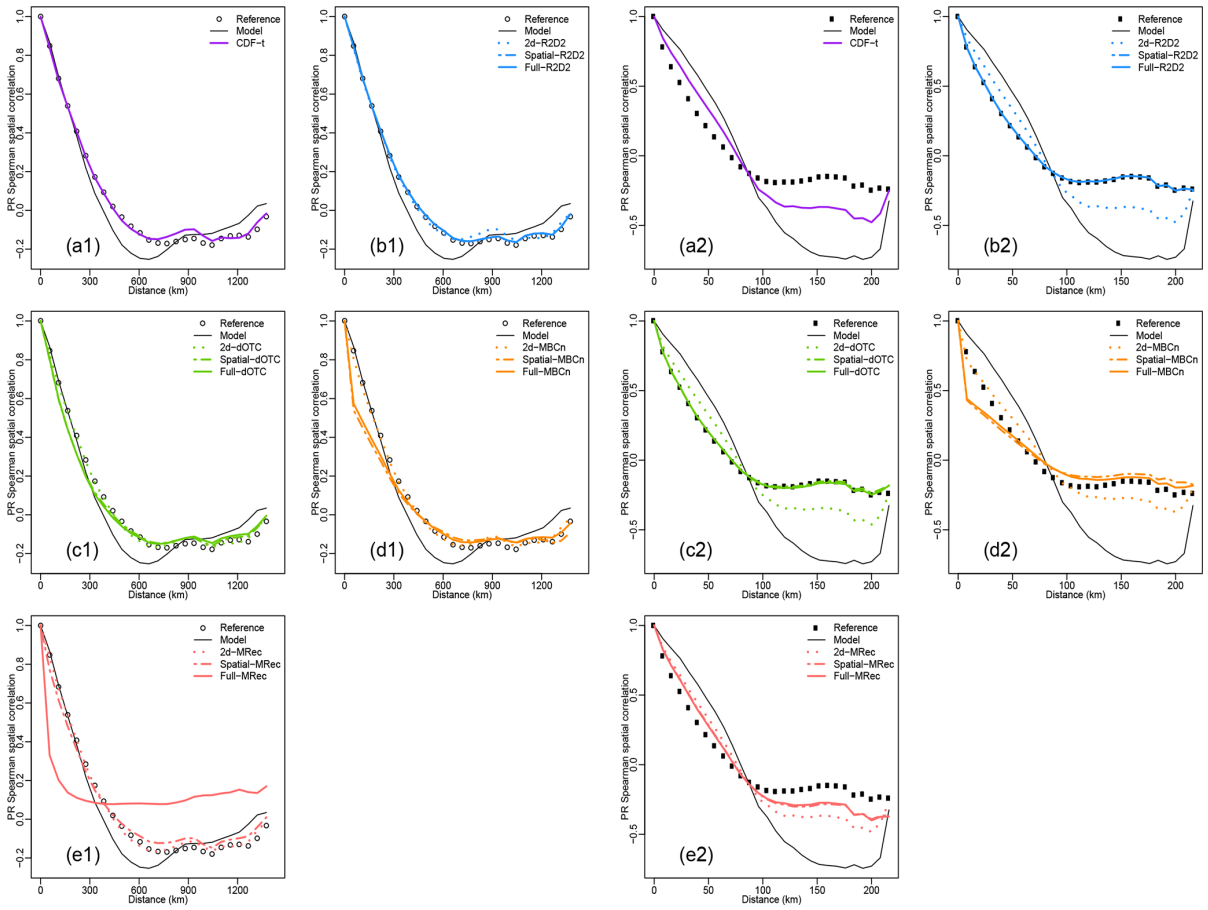

Figure 3. Correlograms for precipitation using **(a1–e1)** WFDEI reference for France and **(a2–e2)** SAFRAN reference for Brittany during winter over the 1979–2016 period. Results are shown for reference (circles) and plain IPSL (black line). Results are displayed for CDF-t, R^2D^2, dOTC, MBC-n and MRec outputs for 2d- (dotted), Spatial- (dashed) and Full- versions (solid lines).

Full-MRec outputs is away from reference data, indicating once again the dysfunction of the MRec method for France. For Spatial- and Full-MBCn (Fig. 3d1), at long distances, similar improvement of spatial correlations are provided as those from dOTC. However, large deviations between correlograms are found for short distances, suggesting a failure for the MBCn method to adjust local spatial properties in a high-dimensional context.

For Brittany, same conclusions hold for R^2D^2 (Fig. 3b2), presenting again a stability of results regardless of both the spatial resolution and the geographical area considered. For dOTC (Fig. 3c2), Spatial- and Full- versions now provide major improvements of spatial correlations compared to their 2d- versions and present results similar to Spatial- and Full-R^2D^2. With regard to MRec (Fig. 3e2), the dysfunction of the Full- version is no longer observed. It now provides results similar to Spatial-MRec and better than 2d-MRec. However, it is worth mentioning that, for Brittany, different results are obtained with MRec between precipitation and temperature spatial corrections. While, for temperature, Spatial-MRec outputs (Fig. S3e2) provide reasonable results with a correlogram relatively close to the one of the reference data, a more moderate improvement of intersite variability is obtained for precipitation (Fig. 3e2). Explanations for these results will be provided in Sect. 6.2. Regarding MBCn (Fig. 3d2), large deviations between correlograms are found for both short and large distances, underlining some instability of the algorithm to adjust for spatial correlations.

5.4 Temporal structure

The different MBC methods implemented here are not intended to adjust temporal structures. Indeed, these multivariate procedures adjust multivariate distributions without accounting for any temporal information. However, although the temporal structures are not adjusted according to the ref-

erence, MBCs necessarily modify the rank sequences of the simulations (Vrac, 2018). This modification is not performed in the same way depending on the MBC or the dimensional configuration used and remains therefore to evaluate. To do so, 1 d lag Pearson autocorrelations are computed at each grid cell for temperature and precipitation. The resulting maps of differences with respect to the reference for the different climate datasets are displayed in Fig. 4 (resp. Fig. S4) for temperature (resp. precipitation).

For France, IPSL temperature autocorrelations differences (Fig. 4b1) are small, indicating a relative agreement of IPSL with the WFDEI reference dataset (Fig. 4a1), showing equivalent high values. A similar differences map is provided by CDF-t outputs (Fig. 4c1). It is however not the case for precipitation (Fig. S4c1), for which a decrease of autocorrelation values is observed over France with respect to the reference and to the model. Although not observed for temperature, it highlights that the univariate correction could have a non-negligible effect on Pearson autocorrelation. Interestingly, 2d- versions (Fig. 4d1, g1, j1 and m1) do not lead to a strong modification of temporal properties with respect to CDF-t. However, from one method to another, temporal structure modifications are not equivalent for Spatial- and Full- versions. For dOTC and MBCn (Fig. 4h1, i1, k1 and l1), as the number of dimensions increases, the temperature autocorrelations seem to be more and more modified, with intensities of values decreasing slightly from Spatial- to Full- versions. This result can also be seen for precipitation in Fig. S4. With regard to MRec, its Spatial- version (Fig. 4n1) presents similar results as those obtained from Spatial-dOTC and Spatial-MBCn. Also, as expected, Full-MRec outputs (Fig. 4o1) do not provide sensible results due to the inability of the method to work properly over the whole of France. Concerning R^2D^2, as the reference dimension driving the rank sequence is the same between Spatial- and Full- configurations, same differences of autocorrelation maps are obtained for these two versions (Fig. 4e1 and f1). Moreover, the autocorrelation value in the grid cell of the reference dimension, i.e., located over Paris for France, is exactly equal to the corresponding one in the CDF-t outputs, by construction. Remarkably, as mentioned by Vrac (2018), autocorrelations of the CDF-t outputs are partially reproduced around the specific locations of the reference dimensions for Spatial-R^2D^2 and Full-R^2D^2 versions, as evidenced by the lightly shaded area around Paris. This reflects the existing spatial correlations between the reference dimension and its local neighborhood, which results in partially reproducing the temporal properties of the model over this area. However, for precipitation (Fig. S4e1 and f1), this result is not as clear-cut as it is for temperature, probably due to weaker spatial correlations around Paris for this physical variable.

In a general way, the same conclusions can be drawn for Brittany, sometimes even better illustrated due to a narrower color scale. The results for Full-MRec are easier to interpret. They present results similar to those from 2d- and Spatial-MRec (Fig. 4o2). In particular, this indicates that, contrary to dOTC and MBCn, MRec does not present an increasing modification of temperature autocorrelations from 2d- to Full- versions.

To better understand the results obtained from Fig. 4, further explanations are required. The relative agreement of Pearson autocorrelation values between the reference and IPSL dataset shown in Fig. 4 might lead one to believe that temporal properties of the model are quite correct for temperature, which is in reality misleading for two main reasons. First, 1 d lag Pearson autocorrelation permits one to assess only a particular feature of the temporal properties, which is obviously insufficient to draw any general conclusions about the quality of simulations concerning these aspects. For example, by simply computing Pearson temperature autocorrelations for higher lag values, a discrepancy of results is obtained between the reference and the simulations (not shown). Second, Pearson autocorrelations depend on two statistical characteristics of time series: their variability and their temporal rank structures. As implemented in Fig. 4, the Pearson autocorrelation metric is hence not able to dissociate them. The similarity between reference and model autocorrelations can then potentially be the combined result of errors stemming from both biased univariate distributions and wrong rank structures of the model.

To better assess temporal structure changes brought by MBCs, the calculation of rank correlations between the bias-corrected time series and the raw climate model simulations is performed for each physical variable and at each grid cell. Results for temperature and precipitation are displayed with boxplots, respectively, in Figs. 5 and S5. The closer the values of the boxplots are to 1, the closer the rank chronologies of the MBC outputs are to the rank chronologies of the model. For France, as expected, similar temperature rank structures are observed between the model and CDF-t/2d-R^2D^2 outputs (Fig. 5a). For the other 2d- versions, rank correlation values are quite close to 1 as well, suggesting that dOTC, MBCn and MRec methods in their 2d- configuration modify only slightly the rank structure of the initial simulations. For Spatial- and Full- configurations, dOTC and MBCn change moderately the rank structures even though they consider more dimensions in the correction. Concerning MRec, without analyzing the Full- outputs, the increasing modification with dimensionality is also observed between 2d- and Spatial-MRec outputs, although less pronounced. In contrast, for Spatial- and Full-R^2D^2 outputs, the changes in the rank structures for France are substantially larger than those discussed until now. This result is also obtained for precipitation in Fig. S5a with an even larger range. The principal reason lies in the fact that, as already explained, R^2D^2 partially preserves rank sequences of the CDF-t outputs – and therefore of the IPSL model – in the direct neighborhood of the reference dimensions but strongly modifies the rank structures outside this neighborhood, which results in

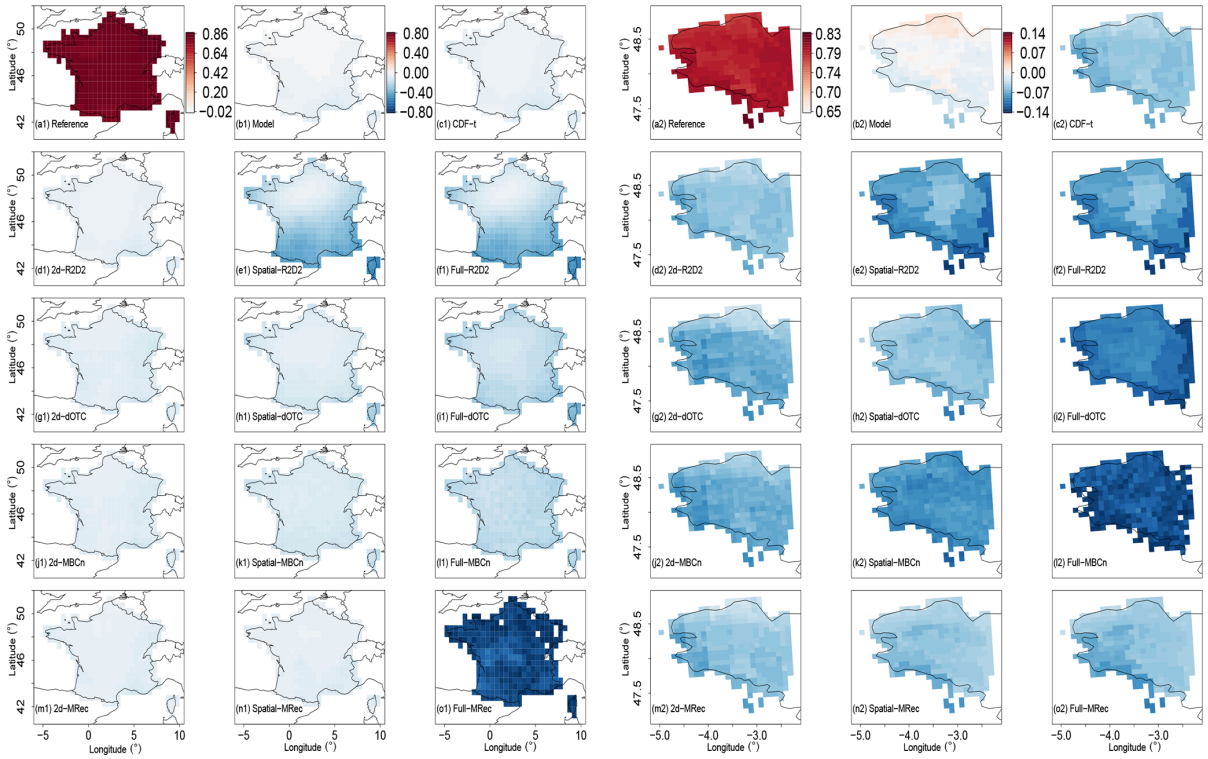

Figure 4. Differences of order 1 Pearson autocorrelation for temperature using **(a1–o1)** WFDEI reference and **(a2–o2)** SAFRAN reference during winter over the 1979–2016 period. Results are shown for reference, plain IPSL, CDF-t, R^2D^2, dOTC, MBC-n and MRec outputs for 2d-, Spatial- and Full- versions. Note that the color scales between panels **(a1)**–**(o1)** and **(a2)**–**(o2)** are not the same to better emphasize intensities of values of the two regions.

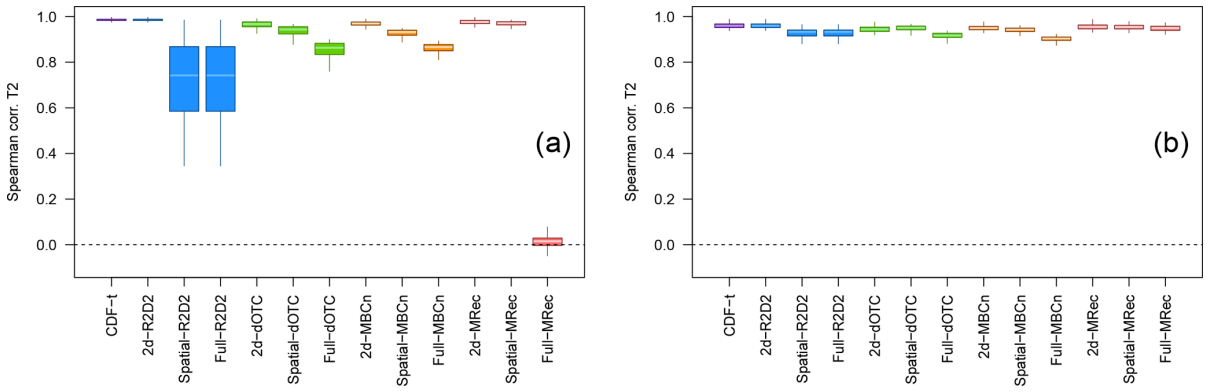

Figure 5. Boxplots of rank correlations computed at each grid cell between the bias-corrected and the raw climate model time series, for temperature, using **(a)** WFDEI for France and **(b)** SAFRAN for Brittany region during winter over the 1979–2016 period. Results are shown for CDF-t, R^2D^2, dOTC, MBC-n and MRec outputs for 2d-, Spatial- and Full- versions.

obtaining some low Spearman correlation values in Figs. 5a and S5a.

For Brittany, results show a less pronounced modification of rank structure for both temperature (Fig. 5b) and precipitation (Fig. S5b) than those observed for France. In particular for temperature, similar rank correlations are obtained for all versions of the methods, even for Spatial- and Full-R^2D^2 outputs, indicating that the number of dimensions has potentially a nonsignificant effect on this criterion over a smaller area. The differences of results between France and Brittany highlight that the size of the region of interest seems to have a non-negligible influence on the temporal properties of BC outputs.

5.5 Multidimensional changes analysis

When correcting climate simulations, in practice, while climate simulations for the present period are adjusted with respect to observations, no reference data are available for the correction of future periods. Assumptions of either stationarity or nonstationarity of copula are then made within the MBCs concerning the change in the multidimensional features between present and future periods. This has then consequences on how MBCs can account for the changes in the multidimensional properties of the climate simulations. Therefore, using the second protocol defined in Sect. 4.2, we now focus on how the different MBC methods reproduce the change in inter-variable and intersite structures, as given by the model to be corrected between two different periods.

5.5.1 Analysis of change in inter-variable correlations

Figure 6 shows, for the bias-corrected outputs, the maps of the difference between the Spearman correlation between temperature and precipitation, computed for the calibration (1979–1997) and the projection (1998-2016) period, respectively. It permits one to visually assess part of the change in the inter-variable dependence structure. Over France, inter-variable change in the IPSL simulations (Fig. 6b1) seems to be distinct from those of WFDEI (Fig. 6a1). CDF-t outputs (Fig. 6c1) reproduce globally the change in the simulations, as they present similar maps. Concerning results for the 2d- (Fig. 6d1) and Full- versions (Fig. 6f1) of R^2D^2, they present inter-variable rank correlation values close to 0. This illustrates the stationarity assumption in R^2D^2: the copula function (i.e., dependence structure) of the observations during the calibration period is reproduced for both calibration and projection, resulting in having no change in inter-variable rank correlations. For their part, 2d-dOTC, 2d-MBCn and 2d-MRec maps (resp. Fig. 6g1, j1 and m1) present roughly the same spatial structures for the differences of Spearman correlations, which indicates that the evolution of the simulations is somehow taken into account in the correction procedures. It must be remarked that, contrary to dOTC and MRec, the stochastic generation of random rotation matrices within the MBCn algorithm leads to get a non-negligible variability in the estimation of the evolution (not shown). This highlights a particular aspect of MBCn: contrary to other methods, MBCn is based on a stochastic procedure, which has a significant impact on its adjustments. Consequently, the quality of MBC data obtained from MBCn can differ from a correction to another for the same climate simulation, depending on the random rotation matrices generated in the algorithm and on the stopping rule (i.e., number of iterations). Interestingly, concerning the method's Spatial- versions (Fig. 6e1, h1, k1 and n1), outputs show changes in inter-variable rank structure similar to those from the model. Indeed, as for CDF-t, rank inter-variable correlations are not adjusted with Spatial- versions. Consequently, the change in inter-variable rank structure of the model is somehow preserved in outputs of Spatial- versions.

For the Full-configuration maps of dOTC and MBCn (Fig. 6i1 and l1), changes simulated by the model are not reproduced at all, which might be due to the failure of these methods to handle the change in time of this statistical feature in high dimensions. As expected, the Full-MRec map (Fig. 6o1) does not provide adequate results due to its inability to adjust the simulated data for France in this dimensional setting.

Concerning the results for Brittany, conclusions similar to those obtained for France can be drawn for R^2D^2 outputs. However, conclusions are quite different for CDF-t, 2d-dOTC, 2d-MBCn and 2d-MRec. Indeed, the changes in rank correlations obtained for these outputs (Fig. 6c2, g2, j2 and m2) are not in agreement at all with the simulated ones (Fig. 6b2). In fact, changes from 2d- outputs are in line with those from CDF-t, illustrating the importance of the correction of 1-dimensional characteristics for inter-variable changes. It is also the case for the Full-MRec map (Fig. 6o2), providing more sensible results than those obtained for France.

Generally speaking, for 2d- and Spatial- versions of MBCs making the assumptions of copula nonstationarity, similar results as those brought by their univariate BC outputs are obtained, suggesting the importance of the correction of univariate distributions for changes in inter-variable rank correlations. Additional results in agreement with these conclusions are obtained for summer and are displayed in Fig. S6.

5.5.2 Analysis of change in spatial correlations

In order to assess changes in spatial structures in bias-corrected outputs, p-Wasserstein distance (see, e.g., Villani, 2008, chap. 6) is computed. This metric measures the distance between two multivariate probability distributions μ and υ and is defined as follows:

$$W_p(\mu,\upsilon) := \left(\inf_{\gamma\in\tau(\mu,\upsilon)} \int_{\mathbb{R}^d\times\mathbb{R}^d} ||x-y||^p \mathrm{d}\gamma(x,y) \right)^{\frac{1}{p}}, \qquad (1)$$

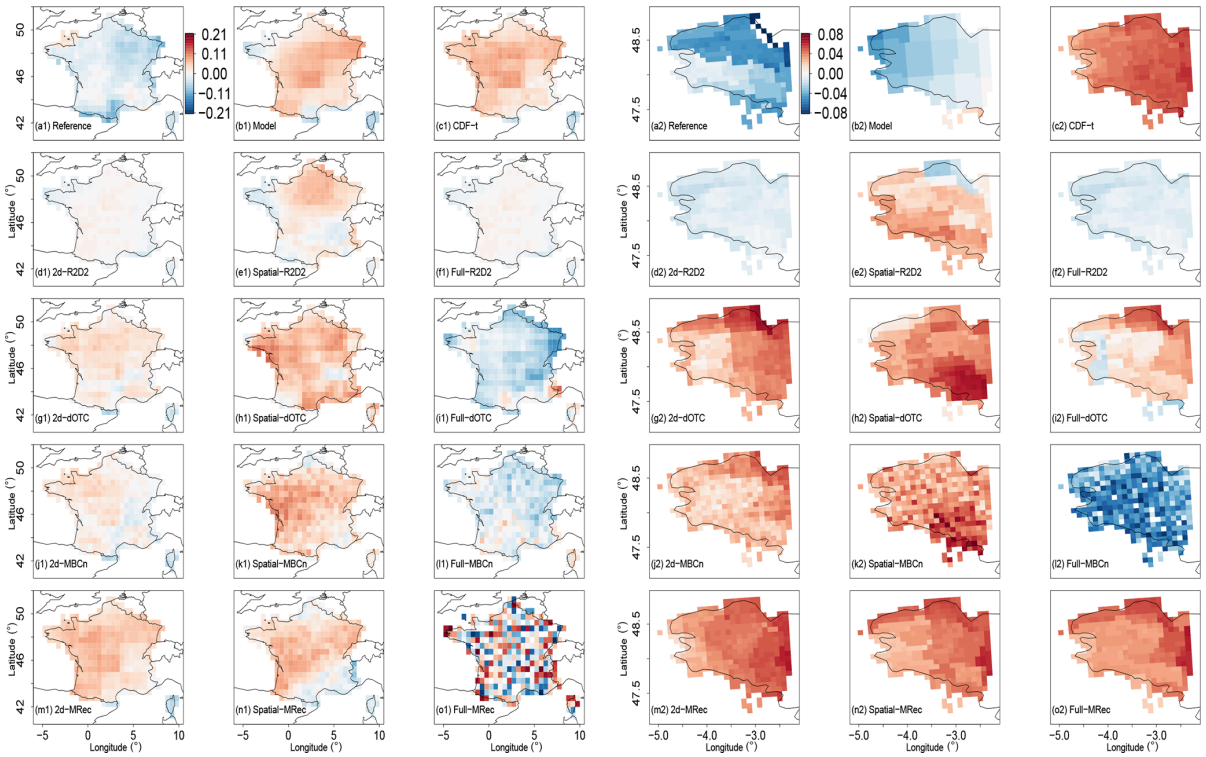

Figure 6. Differences of temperature vs. precipitation Spearman correlations computed at each grid cell between the 1979–1997 and 1998–2016 periods during winter. **(a1–o1)** WFDEI and **(a2–o2)** SAFRAN data are used as references for the bias correction. Note that the color scales between panels **(a1)**–**(o1)** and **(a2)**–**(o2)** are not the same to better emphasize intensities of values of the two regions.

with $\tau(\mu, \upsilon)$ denoting the set of probability measures on $\mathbb{R}^d \times \mathbb{R}^d$ with, respectively, μ and υ as first and second margins and $||.||$ the Euclidean distance. In the present study, p is taken equal to 2, as it ensures the uniqueness of the minimization problem (Santambrogio, 2015). The Wasserstein distance can be seen as the minimum "cost" for transforming a multivariate probability distribution μ into another, here υ. In particular, computing Eq. (1) between a distribution characterizing a sample during the calibration period and another distribution characterizing a sample during the projection period, permits one to provide information on its change across time, whether it represents a univariate, multivariable or multi-site (or both) distribution. More details on how to compute in practice this distance are provided in Appendix C. The resulting metric, denoted Wd, is calculated using the R package "transport" (Schuhmacher et al., 2019) over the region of interest according to three different multivariate distributions:

- on ranks of temperature only over the whole region to assess change in the spatial dependence structure of temperature;
- on ranks for precipitation only over the whole region to assess change in the spatial dependence structure of precipitation;
- on ranks for both temperature and precipitation over the whole region to assess change in the inter-variable and spatial dependence structures of the two variables.

In particular, computing Wd using ranks instead of raw values allows the removal of the change in the univariate distributions from that in spatial and inter-variable relationships. However, comparing Wd values of climate datasets must be made with caution. Indeed, similar values of Wd for different climate datasets do not necessarily imply that their changes in spatial structure are similar. Results for the three Wasserstein distances on ranks are displayed in Fig. 7 for both France and Brittany. Additional results for Wd on raw values are displayed in Fig. S7 for information purposes only.

For France (Fig. 7a), the three Wd are slightly higher for the reference than for the model data (represented by straight lines). Although the differences are quite small, it cannot be concluded directly that changes in spatial structure are identical, as there is no particular reason for this. For CDF-t out-

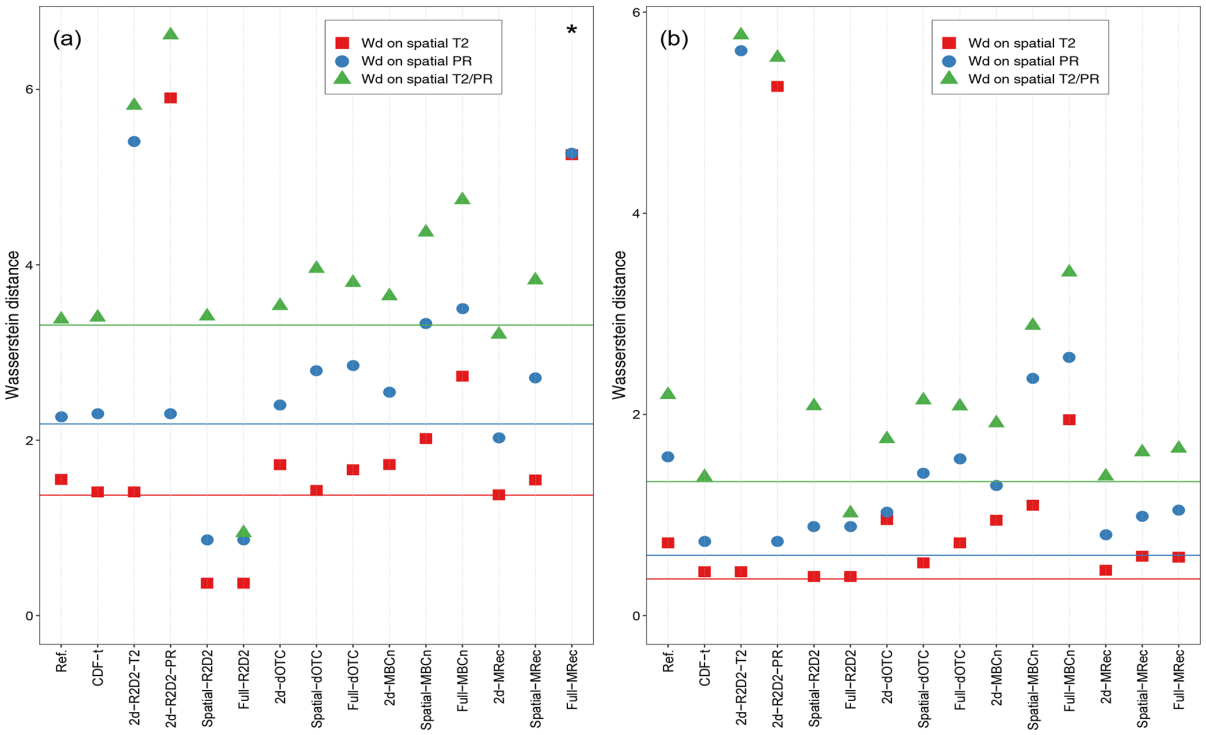

Figure 7. Values of the three Wasserstein distances on ranks between 1979–1997 and 1998–2016 periods during winter for temperature (square), precipitation (circle) and both temperature and precipitation (triangle) for the region of **(a)** France and **(b)** Brittany. Results are presented for the reference, plain IPSL (lines), CDF-t and the different MBCs. 2d-R^2D^2-T2 (resp. 2d-R^2D^2-PR) indicates results for 2d-R^2D^2 with temperature (resp. precipitation) used as reference dimension. Black asterisks indicate values lying outside the plotted range.

puts, similar Wd are obtained as those from the model. However, as the 1d-BC method does not modify (too much) rank sequence of temperature and precipitation time series, it can be deduced that CDF-t outputs globally reproduce/preserve the spatial structure change in the model.

For 2d-R^2D^2 outputs, two results are presented, corresponding to those obtained with either temperature or precipitation used as reference dimension. For the reasons already given (see, e.g., Sect. 5.3), results for 2d-R^2D^2 driven by temperature (resp. precipitation) for the change in spatial structure of temperature (resp. precipitation) are by construction identical to those from CDF-t. Nevertheless, for the spatial structure of temperature and precipitation jointly (triangle), Wd for 2d-R^2D^2 outputs are quite high. Indeed, when the 2d-R^2D^2 version uses either temperature or precipitation rank sequence to drive the other physical variable at each grid cell, the method is likely to degrade the spatial structures of the other variable in a different way for calibration and projection periods. Consequently, the Wasserstein distance captures a "change" in the spatial structure of the two variables between these two periods, but it is in fact due to its deterioration. Concerning Spatial-R^2D^2, low Wd are observed for the change in the spatial structures for temperature and precipitation separately, illustrating the stationarity copula assumption used. However, for the Wd computed for the whole multivariate distribution (triangle in Fig. 7a), Spatial-R^2D^2 presents a higher value, close to that of the IPSL simulations. Indeed, as already explained in Sect. 5.5.1, within Spatial-R^2D^2, copula functions of temperature and precipitation are adjusted separately without correcting inter-variable rank correlations, which results in partially preserving the changes in inter-variable rank structure of the model between calibration and projection period. With regard to Full-R^2D^2, the three Wd are all quite low, in agreement with the stationarity copula assumption it uses. However, it should be noted that the Wd are not equal to 0, whereas, theoretically, no change in spatial structure is performed by Full-R^2D^2. In addition to the reason already cited concerning dry days frequency correction, this is also due to the fact that, in the present study, bias corrections have been performed on a monthly basis, while the evaluation is done at a seasonal scale.

For both dOTC and MBCn outputs, Wd are higher than those from the model. Although the changes in spatial correlations derived by these two methods are too strong, it nevertheless highlights their ability to capture such a change

from the model and to use it in their bias correction procedure. Moreover, as explained in Sect. 5.4, dOTC and MBCn methods modify only slightly the rank structure of the initial simulations. It can then be deduced that the changes in spatial correlations measured for the two methods are (partially) in agreement with those from the model. However, for MBCn, the three Wasserstein distances increase according to the number of dimensions considered in the bias correction, from 2d- to Full- versions. It can be linked with the deterioration of the quality of results already observed for spatial features for very high-dimensional bias correction. Regarding MRec, without speaking about its Full- version, similar observations can be made for 2d- and Spatial- outputs as well. In a general way, the Wd associated with the different configurations for dOTC, MBCn and MRec are always above the Wasserstein distances for R^2D^2, illustrating somehow the assumptions made by these methods about the stationary or nonstationary copula functions.

For Brittany (Fig. 7b), the Wd values computed for the model are quite low, indicating little simulated change in spatial structures for this region. Consequently, the differences of Wd between methods assuming stationarity and nonstationarity of copula functions are less pronounced, but the same conclusions as those drawn for France hold. However, for Full-MRec outputs, Wd values are in relative agreement with those from the model, highlighting the ability of the method to preserve (partially) the simulated changes in spatial structure between the calibration and the projection periods, for a smaller region.

6 Conclusion, discussion and future work

6.1 Conclusion

In this study, we have presented a global picture of the performances of four multivariate bias correction (MBC) methods designed to adjust various multivariate properties of climate simulations. These MBC methods were carefully selected for their differences in terms of methodologies, statistical techniques used, assumptions and philosophical features. For each method, three different dimensional configurations have been tested to correct climate simulations from the IPSL model: a 2d- version to adjust temperature and precipitation time series together but separately for each grid cell, a Spatial- version aiming to correct the simulated fields of temperature and precipitation separately, and a Full- version designed to adjust the two physical variables jointly over the entire domain. Depending on the versions, the objectives of adjustments for multivariate properties are not the same: whereas 2d- and Spatial- versions are designed to correct, respectively, inter-variable and intersite dependence structures, it is expected that the Full- versions adjust both the inter-variable and intersite relationships together. In addition, the univariate CDF-t bias correction method has been implemented and used as a benchmark to assess the benefits of considering multivariate aspects in the correction procedure. A wide range of metrics has been developed to compare bias correction outputs with observations and model data and analyze the adjustments of univariate distributions, inter-variable correlations, intersite correlations and temporal structure. Multidimensional change, i.e., nonstationary, properties have been assessed, providing a comprehensive framework to compare the performance of the methods. The IPSL simulations have been corrected with respect to two distinct reference datasets, i.e., WFDEI and SAFRAN, for, respectively, France and Brittany to attempt to measure the potential influence of the reference spatial resolution on MBC results.

6.2 Discussion and recommendations

General recommendations can be drawn to help practitioners in the choice of BC methods for their applications. For the sake of clarity, Table 2 provides a concise summary of the different recommendations made below. If the univariate CDF-t method corrects the univariate distributions well, it replicates the dependence properties of the model, i.e., inter-variable, intersite or temporal structures, and preserves its multidimensional change across time. Hence, if the multivariate properties of raw climate simulations are not relevant, using 1d-BC methods is not appropriate to get adequate dependence properties. Concerning MBC methods, in general, R^2D^2, dOTC, MBCn and MRec algorithms showed a great ability to adjust the statistical properties associated with the corresponding objectives of the dimensional configurations. Indeed, in addition to correcting univariate distributions, the 2d-, Spatial- and Full- versions of each multivariate method adjust, respectively, inter-variable, spatial and inter-variable/spatial correlations of climate simulations reasonably well. However, caution has to be taken before applying multivariate methods and conducting analysis studies. It has been noted that, depending on the dimensional configuration, instability of some methods can possibly affect corrected outputs, and practitioners have to make sure that no degradation of the desired statistical features is made by the multivariate BC method. In particular, for MBCn and MRec, increasing the number of variables to be corrected jointly in the dimensional configuration is often accompanied by a potentially strong deterioration of spatial properties (see orange tildes in the row "Capacity to correct spatial prop." in Table 2). However, for MBCn, it must be recalled that the number of iterations for the algorithm was fixed to 200 for Full- versions. Although this choice is a good compromise between computation time and fitting the multivariate distribution in the calibration period, this might be suboptimal for some regions. Indeed, early stopping of the procedure could be necessary to avoid overfitting in high dimension, as discussed in Cannon (2018a). Therefore, more research is needed to improve the global performances of MBCn, such as early stopping, optimizing the sequence of random rotation ma-

trices to speed up convergence or, for spatial downscaling problems, adding a conservation step to provide more physical constraints to the bias correction (as proposed in Lange, 2019). Moreover, it has been shown that the characteristics of the climate data to correct can influence the results of the MBCs. In particular, as noted in Sect. 5.3, a distinction of results between temperature and precipitation has been identified for the MRec method (e.g., in Figs. 1, S1, 3 and S3). This might be caused by the way the MRec method performs the correction: only the Pearson correlation structure is adjusted, since it is assumed to be sufficient to correct the full multivariate dependence structure. Although correcting only Pearson spatial correlations for temperature seems reasonable as temperature has traditionally a multivariate Gaussian dependence structure, it appears to be not enough for precipitation, presenting more complex spatial interactions. In that sense, to adjust non-Gaussian climate variables as precipitation, MBCs correcting the full multivariate dependence structure (e.g., R^2D^2, dOTC or MBCn) must be preferred by practitioners.

Also, the ability of the MRec method to adjust Brittany in a very high-dimensional context strongly suggests that the size of the geographical area under study is an important feature for multivariate bias correction. Indeed, a small region like Brittany is likely to present a homogeneous climate or at least to be spatially second-order stationary and, consequently, strong statistical dependencies between locations. Dimensions are then somehow redundant, and spatial correlations for each physical variable are strong, which potentially reduces the number of effective dimensions, also called "spatial degrees of freedom" (e.g., in der Megreditchian, 1990; Bretherton et al., 1999). For MRec, it results in consequently reducing the errors in the computation of the inverse covariance matrices and providing more adequate results. For larger regions presenting a high number of effective dimensions such as France, MRec is however able to provide appropriate results if enough data are provided. For illustration purposes, the MRec method has been additionally applied on a seasonal basis instead of on a monthly one, i.e., correcting 642 dimensions with at least $90\,\mathrm{d} \times 19$ years $= 1710$ time steps. By increasing the number of time steps used in the procedure, high-dimensional sample covariance matrices within MRec are estimated in a more "robust" way, permitting a more suitable correction of the simulations using FullMRec. Results for some criteria are presented in the Supplement (Figs. S8, S9, S10, S11 and S12) but are not commented on in the present study. Also, within MRec, more robust estimators of inverse covariance matrices could be used to obtain more appropriate corrections in a high-dimensional context (e.g., as presented in Levina et al., 2008). More generally, for most MBCs, for a given number of statistical dimensions (e.g., number of grid cells), as going from a large (e.g., France) to a smaller (e.g., Brittany) area reduces the effective dimension, it facilitates the multivariate corrections and therefore improves the results (e.g., compare Figs. 1, S1, 4, S4, 5 and S5). This raises the question of whether applying MBC to climate simulations over large geographical areas is justified, i.e., if it is worth striving for the correction of correlation structures between distant sites presenting weak statistical relationships, and, by doing so, taking the risk of losing global effectiveness of the BC methods. It also highlights the importance of choosing parsimoniously the variables to correct, in order to adjust dependence structures that are relevant without potential quality loss induced by additional (and unneeded) variables.

Regarding the temporal structure, none of the presented multivariate BC methods are designed to adjust this specific statistical aspect (red crosses in Table 2). Moreover, as highlighted by Vrac (2018), any multivariate BC method will necessarily modify the rank sequence of the simulated variables. Results from the present study allow adding nuances to this statement: modification of rank chronologies of the simulations depends on both the multivariate BC methods and the dimensional configurations. In particular, for dOTC, MBCn and MRec methods, a similar behavior was observed: the higher the number of dimensions to correct, the stronger the deterioration of rank chronology of the simulations. However, concerning R^2D^2, depending on the dimensional version, the rank chronology of the model can be reproduced for the specific area around the location of the reference dimension, which could (or not) be desired by practitioners depending on the performance of the simulations.

Finally, we shed light on the nonstationary properties of the multivariate BC methods. While dOTC, MBCn and MRec are designed to transfer some of the multidimensional properties evolution (i.e., change in time) from the model to the bias-corrected data, R^2D^2 assumes the inter-variable and intersite rank correlations – or copula functions – to be stable in time. In a general way, copula nonstationarity for future periods can be reasonably expected, e.g., as documented for rainfall spatial distributions (Wasko et al., 2016), for the dependence between storm surge and rainfall (Wahl et al., 2015), and the dependence between seasonal summer temperature and precipitation (Zscheischler and Seneviratne, 2017). However, on the contrary, it can be argued that inter-variable and spatial dependence structures can be assumed to be stable over time for specific regions, because, to some extent, they can be considered as imposed by physical regional constraints (Vrac, 2018). The differences of Wasserstein distances between the France and the Brittany region for the reference in Fig. 7a and b illustrate well that copula stationarity (or nonstationarity) is not straightforward depending on the geographical domain. The question of the evolution of the copula (i.e., the rank dependence structure) is, therefore, still an open question and needs to be answered on a case-by-case basis. In practice, performances of the methods concerning the multidimensional changes in the different BC outputs are hard to assess precisely, as the potential instability (as in MBCn and MRec) or the stochasticity (as in MBCn) of the methods could affect the quality of the results, making

Table 2. Summary of recommendations for the multivariate BC methods to use with respect to the different assumptions made by practitioners or end users. Green checks and red crosses indicate whether BC methods are recommended for use or not, depending on the statement in rows. Orange tildes indicate when particular caution has to be taken. Not-applicable (n/a) is indicated when the statement in rows does not apply.

Characteristics	CDF-t	R^2D^2	dOTC	MBCn	MRec
Correction of univariate distrib. prop.	✓	✓	✓	✓	✓
Modification of the correlations of the model	✗	✓	✓	✓	✓
Capacity to correct inter-var. prop.	✗	✓	✓	✓	✓
Capacity to correct spatial prop.	✗	✓	✓	∼	∼
Capacity to correct temporal prop.	✗	✗	✗	✗	✗
Preserve the rank structure of the model	✓	∼	∼	∼	∼
Capacity to correct small geographical area	n/a	✓	✓	✓	✓
Capacity to correct large geographical area	n/a	∼	∼	∼	✗
Allow for evolution of the rank dependence	✓	✗	✓	∼	✓

difficult the identification of changes. Moreover, the adjustment of univariate distributions has a non-negligible effect on changes in inter-variable and spatial rank dependences for MBCs assuming non-copula stationarity; in fact, rather than reproducing simulated changes in the correction procedure, these methods are more likely to provide changes in agreement with the ones provided by 1d-BC (e.g., as seen for Brittany in Fig. 6b). Then, in the case where the adjustment of univariate distributions does not modify (too much) the simulated changes in inter-variable and spatial rank dependences, MBCs assuming nonstationary copula would be more likely to present changes in line with those from the model. This result is further confirmed by the results obtained for summer and displayed in Fig. S6 for inter-variable rank dependence changes. The nonstationary property also partly explains the possible differences of results obtained during evaluation (i.e., protocol 1; see Sect. 5) for each criterion. Indeed, as noted in Robin et al. (2019), if the multivariate properties changes provided by the model simulations are incorrect, those of the corrections from methods assuming nonstationarity can be, retrospectively, in disagreement with the changes in the observations.

Therefore, before choosing any multivariate BC method, practitioners have to ask themselves some questions: what are the important statistical properties I want my corrections to provide? Can the evolution of the copula (i.e., rank dependence) in the simulations between calibration and projection be considered as relevant? And should it be reproduced in the correction? If so, according to the results obtained in the present study, dOTC and MRec are good candidates among the presented MBCs. Using these methods, the corrections will be likely to present change in rank dependence similar to the simulations or at least of same sign. It could also be recommended to use these methods if practitioners do not have any idea if the rank dependence changes in the simulations could be considered relevant or not, advocating to let the model express its own dynamic in the absence of relevant judgements. However, if it is assumed that the change in the simulations, in spite of all efforts exerted by climate modellers, is not considered as relevant, R^2D^2 is a good candidate, as it is better to have stationarity of multidimensional rank properties in the correction rather than a non-relevant or wrong one. Moreover, R^2D^2 is also a good candidate for practitioners who do not expect any rank dependence change. The obtained BC outputs from R^2D^2 will not have any change in inter-variable or intersite rank dependence structures, because they are assumed to be imposed by physical constraints and hence stable in time. Concerning MBCn, the global instability of the method in high-dimensional settings, added to the inherent variability due to its stochastic nature, affects significantly the quality of the correction. In practice, therefore, it makes difficult the appropriate preservation of the simulated changes, although the method is specifically designed for that.

6.3 Future work

This intercomparison has been designed such that new BC methods can be easily added. As a result, adding new methods relying on different assumptions, correcting different statistical aspects or using other statistical techniques, is reasonably feasible. Moreover, as mentioned in the Introduction section, bias-adjusted simulations are particularly valuable for impact studies. Despite the challenge of missing impact data, evaluating how the quality of multivariate bias-corrected data influences the results of complex impact models is an important perspective. Providing such an analysis will be useful for the scientific community working on climate change impacts, e.g., in hydrology, agronomy or ecology. In an attempt to answer this question, an appropriate future step could be to apply the presented multivariate BC methods in different dimensional configurations to various

GCM simulations – and not only one as in this study – in order to provide an ensemble of multivariate BC simulations. The obtained datasets would also be useful to carry out scientific studies on other aspects of climate change, such as climate change attribution studies aimed to identify which mechanisms are responsible for changes in the Earth's climate (e.g., Stott et al., 2016; Yiou et al., 2017; Ribes et al., 2020). Indeed, most of these studies use plain simulations, and consequently do not take into account their statistical biases. Conducting attribution studies using plain and bias-corrected simulations will permit one to increase the understanding of the influence of these biases on results, which is essential to provide valuable information to the society concerning the ongoing climate change.

In the present study, it has been highlighted that none of the presented multivariate BC methods were designed to correct or preserve the temporal properties of the simulations. Nevertheless, a few studies have attempted to develop BC methods providing adjustments of some temporal properties of climate variables in addition to the correction of intersite or inter-variable properties (Mehrotra and Sharma, 2015, 2016, 2019). However, considering adjustments for temporal properties will necessarily modify, even slightly, univariate distributions and intersite and/or inter-variable properties. From a more philosophical perspective, striving for the development of MBCs correcting a wide range of statistical features raises also the question of what has been preserved from the simulations in the final BC outputs. By improving the agreement of simulations with observations, this may have the effect of lowering (misleadingly) the uncertainty of the simulated statistical attributes, often without sound physical justifications (Ehret et al., 2012), which puts into question the validity of such methods. Multivariate BC methods developed in the future should, therefore, take into account these issues, in attempting to find a reasonable balance between, on the one hand, the correction of intersite and inter-variable dependences and, on the other hand, the correction or modification of temporal properties, while being able to preserve meaningful simulated characteristics for future periods. To do so, developing new MBC methods including some physical processes to drive the correction procedure is a consistent perspective of development to obtain more realistic bias-corrected simulations. The new developed MBCs could be then included in this intercomparison study, to evaluate and compare their performances with the existing multivariate BC methods.

Appendix A: Details on the CDF-t method

BC methods are applied to correct a simulated fields of S grid cells, each of them described by V physical variables. The total number of statistical dimensions to correct is hence equal to $D = V \times S$, with each of the dimensions composed of N time steps. Let $\mathbf{X}_A$ being a matrix of dimension $N \times D$ and $X_A^d(t)$ the value of the physical variable corresponding to the dth dimension at time t from the matrix $\mathbf{X}_A$. Datasets, i.e., matrices, to correct with BC methods are model outputs during the calibration (denoted $\mathbf{X}_{\mathrm{M_C}}$) and the projection period (denoted $\mathbf{X}_{\mathrm{M_P}}$), according to the data from the reference observed during calibration (denoted $\mathbf{X}_{\mathrm{R_C}}$). Corrected outputs for the calibration and the projection period are denoted $\widehat{\mathbf{X}}_{\mathrm{M_C}}$ and $\widehat{\mathbf{X}}_{\mathrm{M_P}}$, respectively.

CDF-t is a version of quantile–quantile method that takes into account, by defining a transfer function T, the potential evolution of univariate CDFs from the calibration to the projection period. For this subsection, let's assume that $F_{\mathrm{M_C}}^d$ and $F_{\mathrm{R_C}}^d$ are respectively the univariate CDFs of the dth dimension $\boldsymbol{X}_{\mathrm{M_C}}^d$ and $\boldsymbol{X}_{\mathrm{R_C}}^d$ located at the same grid cell for the model and the reference in the calibration period. To simplify the notation, we will denote these CDFs $F_{\mathrm{M_C}}$ and $F_{\mathrm{R_C}}$, respectively. The transfer function T is defined such that it links the two CDFs $F_{\mathrm{M_C}}$ and $F_{\mathrm{R_C}}$ as follows:

$$T\left(F_{\mathrm{M_C}}(x)\right) = F_{\mathrm{R_C}}(x). \tag{A1}$$

A more simple formulation of T is then obtained by replacing x by $F_{\mathrm{M_C}}^{-1}(u)$, with u probabilities in $[0, 1]$.

$$T(u) = F_{\mathrm{R_C}}\left(F_{\mathrm{M_C}}^{-1}(u)\right). \tag{A2}$$

By assuming time-stationarity of the transformation T, it can be applied similarly in the projection period to link CDFs between the model and the reference:

$$T\left(F_{\mathrm{M_P}}(x)\right) = F_{\mathrm{R_P}}(x). \tag{A3}$$

By combining Eqs. (A2) and (A3), we then can generate $F_{\mathrm{R_P}}$, the estimated CDF of the climate variable in the reference during the projection period:

$$F_{\mathrm{R_P}}(x) = F_{\mathrm{R_C}}\left(F_{\mathrm{M_C}}^{-1}\left(F_{\mathrm{M_P}}(x)\right)\right). \tag{A4}$$

Once $F_{\mathrm{R_P}}$ has been estimated, a simple quantile–quantile method is performed between $F_{\mathrm{R_P}}$ and $F_{\mathrm{M_P}}$ to derive the bias-corrected time series $\widehat{X}_{\mathrm{M_P}}^d$ for the projection period as follows:

$$\widehat{X}_{\mathrm{M}_p}^d(t) = F_{\mathrm{R_P}}^{-1}\left(F_{\mathrm{M_P}}\left(X_{\mathrm{M_P}}^d(t)\right)\right). \tag{A5}$$

While a traditional quantile-mapping approach performed to correct a dataset $\mathbf{X}_{\mathrm{M_P}}$ of simulations over the projection period will use the formulation $\widehat{X}_{\mathrm{M}_p}^d(t) = F_{\mathrm{R_C}}^{-1}\left(F_{\mathrm{M_C}}\left(X_{\mathrm{M_P}}^d(t)\right)\right)$ (i.e., based on two distributions characterizing the calibration period), the CDF-t method relies on Eq. (A5) where the two involved distributions characterize projected distributions. By proceeding this way, CDF-t takes into account the potential evolution of CDFs of the model between the calibration and projection periods to adjust the projection period. CDF-t is applied independently for each of the D statistical dimensions and for both calibration and projection period to derive the final bias-corrected outputs $\widehat{\mathbf{X}}_{\mathrm{M_C}}$ and $\widehat{\mathbf{X}}_{\mathrm{M_P}}$.

Appendix B: Details on the R^{2}D^2 method

The R^{2}D^2 method, belonging to the marginal/dependence category, consists of several successive steps that are similar to adjust climate simulations for calibration and projection periods. Hence, to avoid redundancy, the correction procedure for the projection period will only be explained in this subsection. In this appendix, temporary corrected outputs for the projection period are denoted $\widetilde{\mathbf{X}}_{\mathrm{M_P}}$.

- First, an univariate BC method is performed for the projection period to obtain the $N \times D$ matrix output $\widetilde{\mathbf{X}}_{\mathrm{M_P}}$. As a reminder, $\widetilde{\mathbf{X}}_{\mathrm{M_P}} = \left[\left(\widetilde{X}_{\mathrm{M_P}}^1(1), \ldots, \widetilde{X}_{\mathrm{M_P}}^1(N)\right)', \ldots, \left(\widetilde{X}_{\mathrm{M_P}}^D(1), \ldots, \widetilde{X}_{\mathrm{M_P}}^D(N)\right)'\right]$.

- For each dimension d, R^{2}D^2 computes the ranks of the time series within the univariate BC outputs $\widetilde{\mathbf{X}}_{\mathrm{M_P}}$. For example, for the dimension d, the $N \times 1$ vector $\left(\mathrm{rank}\left(\widetilde{X}_{\mathrm{M_P}}^d(1)\right), \ldots, \mathrm{rank}\left(\widetilde{X}_{\mathrm{M_P}}^d(N)\right)\right)'$, denoted $\left(\widetilde{r}_{\mathrm{M_P}}^d(1), \ldots, \widetilde{r}_{\mathrm{M_P}}^d(N)\right)'$, is computed. It results in getting, for each time step t, a D-dimensional vector $\widetilde{\mathbf{R}}_{\mathrm{M_P}}(t) = \left(\widetilde{r}_{\mathrm{M_P}}^1(t), \ldots, \widetilde{r}_{\mathrm{M_P}}^D(t)\right)$, which provides the multivariate rank structure of $\widetilde{\mathbf{X}}_{\mathrm{M_P}}$ at t.

- For each dimension d, R^{2}D^2 computes the ranks of the time series within the reference dataset during calibration $\mathbf{X}_{\mathrm{R_C}}$. For example, for the dimension d, the $N \times 1$ vector $\left(\mathrm{rank}\left(X_{\mathrm{R_C}}^d(1)\right), \ldots, \mathrm{rank}\left(X_{\mathrm{R_C}}^d(N)\right)\right)'$, denoted $\left(r_{\mathrm{R_C}}^d(1), \ldots, r_{\mathrm{R_C}}^d(N)\right)'$, is computed. It results in getting, for each time step t, a D-dimensional vector $\mathbf{R}_{\mathrm{R_C}}(t) = \left(r_{\mathrm{R_C}}^1(t), \ldots, r_{\mathrm{R_C}}^D(t)\right)$, which provides the multivariate rank structure of $\mathbf{X}_{\mathrm{R_C}}$ at t.

- A reference dimension d needs to be selected by the users in $\widetilde{\mathbf{X}}_{\mathrm{M_P}}$. The corresponding univariate time series will be kept untouched in the final R^{2}D^2 outputs as the correction of the multivariate dependence structure is articulated on this dimension "pivot". For each time step t:

- the algorithm R^{2}D^2 finds t^* such that $\widetilde{r}^d_{\mathrm{M_P}}(t) = r^d_{\mathrm{R_C}}(t^*)$. From t^*, R^{2}D^2 deduces the multivariate rank structure of the reference during the calibration period at this specific time step: $\mathbf{R}_{\mathrm{R_C}}(t^*) = \left(r^1_{\mathrm{R_C}}(t^*), \ldots, r^D_{\mathrm{R_C}}(t^*)\right)$;
- R^{2}D^2 forces the D-dimensional vector of ranks of its final outputs $\widehat{\mathbf{X}}_{\mathrm{M_P}}$ to be equal to $\widehat{\mathbf{R}}_{\mathrm{M_P}}(t) = \left(r^1_{\mathrm{R_C}}(t^*), \ldots, \widetilde{r}^d_{\mathrm{M_P}}(t), \ldots, r^D_{\mathrm{R_C}}(t^*)\right)$.
  To do so, the algorithm looks to shuffle the values in each of the dimensions $k \neq d$ of $\widetilde{\mathbf{X}}_{\mathrm{M_P}}$, such that its rank structure at time t matches $\widehat{\mathbf{R}}_{\mathrm{M_P}}(t)$. In a more explicit way, for all $k \neq d$, R^{2}D^2 finds the time steps t_k such that $r^k_{\mathrm{R_C}}(t^*) = \widetilde{r}^k_{\mathrm{M_P}}(t_k)$. The value in $\widetilde{X}^k_{\mathrm{M_P}}$ to shuffle associated with the rank $\widetilde{r}^k_{\mathrm{M_P}}(t_k)$ is then derived and copied in the final outputs $\widehat{X}^k_{\mathrm{M_P}}(t)$.

- By repeating the step 4 until each dimension has been used one time as a reference for the shuffling, R^{2}D^2 is able to derive a collection of D MBC outputs, with exactly the same multivariate dependence structure but differing in temporal properties, describing the possible variability in the different rank structures.

Appendix C: Details on the dOTC method

The dOTC method, belonging to the all-in-one category, relies on optimal transport theory to adjust climate simulations. A slightly different mathematical notation needs to be used here to explain dOTC. Let define $\mathbf{X}_{\mathrm{R_C}}(t)$ the realizations of $\mathbf{X}_{\mathrm{R_C}}$ at each time step t across each of the D dimensions. The collection of the variables $\left(\mathbf{X}_{\mathrm{R_C}}(1), \ldots, \mathbf{X}_{\mathrm{R_C}}(N)\right)$ forms a $D \times N$ matrix and describes $\mathbf{X}_{\mathrm{R_C}}$ in a different way. Similarly, $\left(\mathbf{X}_{\mathrm{M_C}}(1), \ldots, \mathbf{X}_{\mathrm{M_C}}(N)\right)$ and $\left(\mathbf{X}_{\mathrm{M_P}}(1), \ldots, \mathbf{X}_{\mathrm{M_P}}(N)\right)$ are considered for, respectively, $\mathbf{X}_{\mathrm{M_C}}$ and $\mathbf{X}_{\mathrm{M_P}}$. In the following, $\mathbf{c}_i$ denotes a collection of multivariate cells that partition regularly $\mathbb{R}^D$ and fully cover $\left(\mathbf{X}_{\mathrm{M_C}}(1), \ldots, \mathbf{X}_{\mathrm{M_C}}(N)\right)$ and $\left(\mathbf{X}_{\mathrm{M_P}}(1), \ldots, \mathbf{X}_{\mathrm{M_P}}(N)\right)$. To simplify notations, the center of a grid cell $\mathbf{c}_i$ is also denoted $\mathbf{c}_i$. Hereinafter is presented first how dOTC adjusts the calibration period of climate simulations to derive $\widehat{\mathbf{X}}_{\mathrm{M_C}}$. Then, the algorithm procedure will be detailed for the adjustment of the projection period $\widehat{\mathbf{X}}_{\mathrm{M_P}}$.

The "OTC" procedure for the calibration period:

- First, the algorithm estimates $\widetilde{\mathbb{P}}_{\mathbf{X}_{\mathrm{R_C}}}$ and $\widetilde{\mathbb{P}}_{\mathbf{X}_{\mathrm{M_C}}}$ the empirical multivariate distributions of $\mathbf{X}_{\mathrm{R_C}}$ and $\mathbf{X}_{\mathrm{M_C}}$. To do so, dOTC computes a sum of Dirac masses. For example, for $\mathbf{X}_{\mathrm{M_C}}$, we have

$$\widetilde{\mathbb{P}}_{\mathbf{X}_{\mathrm{M_C}}}(A) = \sum_{i=1}^{I} p_{\mathbf{X}_{\mathrm{M_C},i}} \delta_{\mathbf{c}_i}(A),$$

  where $p_{\mathbf{X}_{\mathrm{M_C},i}} = \frac{1}{N} \sum_{t=1}^{N} \mathbf{1}(\mathbf{X}_{\mathrm{M_C}}(t) \in \mathbf{c}_i)$, and $A \subset \mathbb{R}^D$.

- Then, the coefficients γ_{ij} defining the estimator $\widetilde{\gamma}$ of the optimal plan that moves the bin $\mathbf{c}_i$ of $\widetilde{\mathbb{P}}_{\mathbf{X}_{\mathrm{M_C}}}$ to the bin $\mathbf{c}_j$ of $\widetilde{\mathbb{P}}_{\mathbf{X}_{\mathrm{R_C}}}$ are computed. For $A, B \subset \mathbb{R}^D$, $\widetilde{\gamma}$ is defined as follows:

$$\widetilde{\gamma}(A \times B) = \sum_{i,j=1}^{I,J} \gamma_{ij} \delta_{(\mathbf{c}_i, \mathbf{c}_j)}(A \times B).$$

  The coefficient γ_{ij} corresponds to the joint probability of $\mathbf{X}_{\mathrm{M_C}}$ being in $\mathbf{c}_i$ and $\mathbf{X}_{\mathrm{R_C}}$ being in $\mathbf{c}_j$, which is part of the MBC process. They have to respect the following constraints:

$$\sum_{j=1}^{J} \gamma_{ij} = p_{\mathbf{X}_{\mathrm{M_C},i}},$$

$$\sum_{i=1}^{I} \gamma_{ij} = p_{\mathbf{X}_{\mathrm{R_C},j}},$$

  and they have to minimize the following cost function $\widetilde{C}$:

$$\widetilde{C}(\widetilde{\gamma}) = \sum_{i,j=1}^{I,J} \| \mathbf{c}_i - \mathbf{c}_j \|^2 \gamma_{ij}.$$

  To find these coefficients that form the so-called optimal transport plan, the algorithm resolves the linear programming problem by using the procedure developed by Flamary and Courty (2017).

- Then, for each time step t are the following steps:
  - The algorithm finds the cell $\mathbf{c}_i$ containing $\mathbf{X}_{\mathrm{M_C}}(t)$.
  - Using the plan γ_{ij}, it constructs the conditional probability vector $\widetilde{\gamma}_{\mathbf{X}_{\mathrm{M_C}}(t)} = \left(\gamma_{i,1}, \ldots, \gamma_{i,J}\right) / p_{\mathbf{X}_{\mathrm{M_C},i}}$.
  - According to the probability vector $\widetilde{\gamma}$, the algorithm draws a $j^* \in 1, \ldots, J$.
  - The correction $\widehat{\mathbf{X}}_{\mathrm{M_C}}(t)$ is then derived with an uniform draw in $\mathbf{c}_{j^*}$.

- After iterating for each t, the final outputs for the calibration period $\widehat{\mathbf{X}}_{\mathrm{M_C}}$ is obtained.

The "dOTC" procedure for the projection period:

- As explained before, dOTC estimates $\widetilde{\mathbb{P}}_{\mathbf{X}_{\mathrm{R_C}}}$, $\widetilde{\mathbb{P}}_{\mathbf{X}_{\mathrm{M_C}}}$ and $\widetilde{\mathbb{P}}_{\mathbf{X}_{\mathrm{M_P}}}$ the empirical multivariate distributions of $\mathbf{X}_{\mathrm{R_C}}$, $\mathbf{X}_{\mathrm{M_C}}$ and $\mathbf{X}_{\mathrm{M_P}}$.

- Then, the coefficients γ_{ij} defining the estimator $\widetilde{\gamma}$ of the optimal plan that moves the bin $\mathbf{c}_i$ of $\widetilde{\mathbb{P}}_{\mathbf{X}_{\mathrm{M_C}}}$ to the bin $\mathbf{c}_j$ of $\widetilde{\mathbb{P}}_{\mathbf{X}_{\mathrm{R_C}}}$ are computed.

- Similarly, the coefficients φ_{ik} defining the estimator $\widetilde{\varphi}$ of the optimal plan that moves the bin $\mathbf{c}_i$ of $\widetilde{\mathbb{P}}_{\mathbf{X}_{\mathrm{M_C}}}$ to the bin $\mathbf{c}_k$ of $\widetilde{\mathbb{P}}_{\mathbf{X}_{\mathrm{M_P}}}$ are computed.

- By default, the diagonal matrix of the standard deviations $\mathbf{D}$ is computed: $\mathbf{D} = \mathrm{diag}\left(\sigma_{X_{\mathrm{M_C}}} \sigma_{X_{\mathrm{R_C}}}^{-1}\right)$. Others alternatives for the computation of $\mathbf{D}$ are possible and detailed in Robin et al. (2019).
- Then, for each time step t are the following steps:
  - The algorithm finds the cell $\mathbf{c}_j$ containing $\mathbf{X}_{\mathrm{R_C}}(t)$.
  - Using the plan γ_{ij}, it finds the cell $\mathbf{c}_i$ of $\widetilde{\mathbb{P}}_{\mathbf{X}_{\mathrm{M_C}}}$ associated with $\mathbf{c}_j$.
  - Using the plan φ_{ik}, it finds the cell $\mathbf{c}_k$ of $\widetilde{\mathbb{P}}_{\mathbf{X}_{\mathrm{M_P}}}$ associated with $\mathbf{c}_i$.
  - Using $\mathbf{D}$, it computes the vector $\mathbf{v}_{ik} := \mathbf{c}_k - \mathbf{c}_i$ for scaling adjustment of the correction.
  - A preliminary (and temporary) correction of the model during the projection $\check{\mathbf{X}}_{\mathrm{M_P}}(t)$ is then obtained, $\check{\mathbf{X}}_{\mathrm{M_P}}(t) = \mathbf{X}_{\mathrm{M_C}}(t) + \mathbf{D}.\mathbf{v}_{ik}$.
- Then, it estimates $\check{\mathbb{P}}_{\check{\mathbf{X}}_{\mathrm{M_P}}}$ the empirical multivariate distribution of $\check{\mathbf{X}}_{\mathrm{M_P}}$.
- Finally, the OTC procedure (see above for calibration period) is applied between $\left(\mathbf{X}_{\mathrm{M_P}}(1), \ldots, \mathbf{X}_{\mathrm{M_P}}(N)\right)$ and $\left(\check{\mathbf{X}}_{\mathrm{M_P}}(1), \ldots, \check{\mathbf{X}}_{\mathrm{M_P}}(N)\right)$ to produce the final outputs $\left(\widehat{\mathbf{X}}_{\mathrm{M_P}}(1), \ldots, \widehat{\mathbf{X}}_{\mathrm{M_P}}(N)\right)$.

Appendix D: Details on the MBCn method

The MBCn method can be summarized in three steps in the way it corrects climate simulations. As a reminder, MBCn belongs to the marginal/dependence category, i.e., correcting separately marginal distributions and full dependence structure of climate simulations. In this appendix, temporary corrected outputs of a matrix $\mathbf{X}_A$ are denoted with tilde accents ($\widetilde{\mathbf{X}}_A$) or inverted hats ($\check{\mathbf{X}}_A$).

- Step 1: first, marginal distributions are corrected with an univariate BC method. To do so, MBCn uses the Quantile Delta Mapping (QDM from Cannon et al., 2015) algorithm defined as follows:

$$\begin{cases} \widetilde{X}^d_{\mathrm{M_C}}(t) & = F^{-1}_{\mathrm{R_C}}\left(F_{\mathrm{M_C}}\left(X^d_{\mathrm{M_C}}(t)\right)\right) \\ \Delta_t & = X^d_{\mathrm{M_P}}(t) - F^{-1}_{\mathrm{M_C}}\left(F_{\mathrm{M_P}}\left(X^d_{\mathrm{M_P}}(t)\right)\right) \\ \widetilde{X}^d_{\mathrm{M_P}}(t) & = F^{-1}_{\mathrm{R_C}}\left(F_{\mathrm{M_P}}\left(X^d_{\mathrm{M_P}}(t)\right)\right) + \Delta_t \,. \end{cases} \quad \text{(D1)}$$

  This transfer function preserves absolute changes in quantiles and has to be applied for interval variables such as temperature. For ratio variables like precipitation, the addition/substraction operators in the transfer function have to be replaced by multiplication/division operators to define a function that preserves relative changes in quantiles. For both calibration and projection period, the D physical variables are independently adjusted by applying the corresponding transfer function. The resulting matrices $\widetilde{\mathbf{X}}_{\mathrm{M_C}}$ and $\widetilde{\mathbf{X}}_{\mathrm{M_P}}$ with adjusted marginal distributions are stored by the algorithm in, respectively, $\widetilde{\mathbf{X}}^{\mathrm{init}}_{\mathrm{M_C}}$ and $\widetilde{\mathbf{X}}^{\mathrm{init}}_{\mathrm{M_P}}$ before the second step, as it reuses them in the third one.
- Step 2: within the MBCn algorithm, the multivariate dependence structure of the simulations is adjusted through an iterative procedure. At each iteration j, an application of a $D \times D$ random orthogonal rotation matrix $\mathbf{R}^{[j]}$ (Mezzadri, 2007) is performed on the datasets $\mathbf{X}_{\mathrm{R_C}}$, $\widetilde{\mathbf{X}}_{\mathrm{M_C}}$ and $\widetilde{\mathbf{X}}_{\mathrm{M_P}}$ obtained from Step 1:

$$\begin{cases} \check{\mathbf{X}}^{[j]}_{\mathrm{R_C}} & = \mathbf{X}^{[j]}_{\mathrm{R_C}} \mathbf{R}^{[j]} \\ \check{\mathbf{X}}^{[j]}_{\mathrm{M_C}} & = \widetilde{\mathbf{X}}^{[j]}_{\mathrm{M_C}} \mathbf{R}^{[j]} \\ \check{\mathbf{X}}^{[j]}_{\mathrm{M_P}} & = \widetilde{\mathbf{X}}^{[j]}_{\mathrm{M_P}} \mathbf{R}^{[j]}. \end{cases} \quad \text{(D2)}$$

  It permits one to provide linear combinations of the original variables. The QDM transfer function defined in Eq. (D1) for interval variables, i.e., with addition/substraction operators, is then applied on each of the rotated marginal distributions of $\check{\mathbf{X}}^{[j]}_{\mathrm{M_C}}$ and $\check{\mathbf{X}}^{[j]}_{\mathrm{M_P}}$, considering the corresponding rotated marginal distributions in $\check{\mathbf{X}}^{[j]}_{\mathrm{R_C}}$ as the reference. Once marginal distributions have been adjusted in $\check{\mathbf{X}}^{[j]}_{\mathrm{M_C}}$ and $\check{\mathbf{X}}^{[j]}_{\mathrm{M_P}}$, matrices are rotated back to the physical variables ranges:

$$\begin{cases} \mathbf{X}^{[j+1]}_{\mathrm{R_C}} & = \mathbf{X}^{[j]}_{\mathrm{R_C}} \\ \widetilde{\mathbf{X}}^{[j+1]}_{\mathrm{M_C}} & = \check{\mathbf{X}}^{[j]}_{\mathrm{M_C}} \mathbf{R}^{[j]^{-1}} \\ \widetilde{\mathbf{X}}^{[j+1]}_{\mathrm{M_P}} & = \check{\mathbf{X}}^{[j]}_{\mathrm{M_P}} \mathbf{R}^{[j]^{-1}}. \end{cases} \quad \text{(D3)}$$

  These successive steps are applied iteratively until the multivariate distribution of the corrected simulations $\widetilde{\mathbf{X}}^{[j+1]}_{\mathrm{M_C}}$ matches the one of the reference $\mathbf{X}_{\mathrm{R_C}}$.
- Step 3: once the full dependence structure of simulated variables converged to the one of the reference after, let say, the j^*th iteration, MBCn replaces quantiles of each of the variables in $\widetilde{\mathbf{X}}^{[j^*+1]}_{\mathrm{M_C}}$ and $\widetilde{\mathbf{X}}^{[j^*+1]}_{\mathrm{M_P}}$ obtained at the end of Step 2 with those from $\widetilde{\mathbf{X}}^{\mathrm{init}}_{\mathrm{M_C}}$ and $\widetilde{\mathbf{X}}^{\mathrm{init}}_{\mathrm{M_P}}$ obtained during Step 1. This additional step prevents the possible deterioration of the model trend during the correction of the multivariate dependence structure in Step 2. Simulations with corrected marginal distributions features and full dependence structure $\widehat{\mathbf{X}}_{\mathrm{M_C}}$ and $\widehat{\mathbf{X}}_{\mathrm{M_P}}$ are then obtained.

Appendix E: Details on the MRec method

The MRec method, belonging to the all-in-one category, consists of the following steps.

- First, each of the D dimensions in $\mathbf{X}_{\mathrm{R_C}}$ is transformed independently in the Gauss domain. However, the trans-

formation differs between interval variables, i.e., temperature, and ratio variables, i.e., precipitation, and is performed as follows:

- For a dimension d being an interval variable, a distribution $F^d_{\mathrm{R_C}}$ is fitted:

  $$F^d_{\mathrm{R_C}}(x) = \mathbb{P}\left(X^d_{\mathrm{R_C}}(t) < x\right).$$

  Then, the corresponding vector $\boldsymbol{W}^d$ is computed as follows:

  $$\boldsymbol{W}^d(t) = \Phi^{-1}\left(F^d_{\mathrm{R_C}}\left(X^d_{\mathrm{R_C}}(t)\right)\right),$$

  with Φ the distribution function of the standard normal distribution $\mathcal{N}(0,1)$.
- For a dimension k being a ratio variable, a distribution $F^k_{\mathrm{R_C}}$ is fitted:

  $$F^k_{\mathrm{R_C}}(x) = \mathbb{P}\left(X^k_{\mathrm{R_C}}(t) < x | X^k_{\mathrm{R_C}}(t) > 0\right).$$

  Additionally, the frequency P_{k0} of null events in $X^k_{\mathrm{R_C}}$ is computed:

  $$P_{k0} = \mathbb{P}\left(X^k_{\mathrm{R_C}}(t) = 0\right).$$

  Then, the corresponding vector $\boldsymbol{W}^k$ is computed as follows:

  $$\boldsymbol{W}^k(t) = \begin{cases} \Phi^{-1}\left(F^k_{\mathrm{R_C}}\left(X^k_{\mathrm{R_C}}(t)\right)(1-P_{k0}) + P_{k0}\right) \\ \Phi^{-1}\left(\frac{P_{k0}}{2}\right). \end{cases}$$

Doing this step for each dimension permits one to derive the matrix $\mathbf{W}$ of dimension $N \times D$, composed of the Gaussian transformed vectors $\boldsymbol{W}^1, \ldots, \boldsymbol{W}^D$.Following the notation in Bárdossy and Pegram (2012), the same procedure is repeated for $\mathbf{X}_{\mathrm{M_C}}$ and $\mathbf{X}_{\mathrm{M_P}}$ to derive, respectively, the Gaussian transformed data $\mathbf{Y}$ and $\mathbf{Y}'$.

- For both Gaussian transformed data $\mathbf{W}$ and $\mathbf{Y}$, the $N \times N$ Pearson cross-correlation matrices $\mathbf{C}_W$ and $\mathbf{C}_Y$ are computed.
- A singular value decomposition (SVD) is applied on $\mathbf{C}_W$ such that

  $$\mathbf{C}_W = \mathbf{A}_W \mathbf{D}_W \mathbf{B}_W^T,$$

  with $\mathbf{A}_W$ and $\mathbf{B}_W$ having same dimensions as $\mathbf{C}_W$, and $\mathbf{D}_W$ a diagonal matrix of singular values. From this decomposition, the square root matrix of $\mathbf{C}_W$, denoted $\mathbf{S}_W$, can be obtained as follows:

  $$\mathbf{S}_W = \mathbf{A}_W \mathbf{D}_W^{1/2} \mathbf{A}_W^T.$$

- Similarly, a singular value decomposition (SVD) is applied on $\mathbf{C}_Y$ such that

  $$\mathbf{C}_Y = \mathbf{A}_Y \mathbf{D}_Y \mathbf{B}_Y^T.$$

  From this decomposition, its inverse square root matrix $\mathbf{T}_Y$ can be obtained as follows:

  $$\mathbf{T}_Y = \mathbf{A}_Y \mathbf{D}_Y^{-1/2} \mathbf{A}_Y^T.$$

- $\mathbf{Y}$ is decorrelated to $\mathbf{Q}$: $\mathbf{Q} = \mathbf{Y}\mathbf{T}_Y$.
- $\mathbf{Q}$ is then recorrelated to $\mathbf{V}$: $\mathbf{V} = \mathbf{Q}\mathbf{S}_W$. $\mathbf{V}$ is hence the recorrelated transformed model data for the calibration period presenting the same correlation structure as $\mathbf{W}$.
- For the projection period, $\mathbf{V}'$ is computed directly without decorrelation step: $\mathbf{V}' = \mathbf{Y}'\mathbf{T}_Y\mathbf{S}_W$.
- $\mathbf{V}$ and $\mathbf{V}'$ are then transformed back to physical variables using a univariate quantile–quantile method for each dimension d, with $\mathbf{X}^d_{\mathrm{R_C}}$ being the target for the correction. The desired adjusted matrices $\widehat{\mathbf{X}}_{\mathrm{M_C}}$ and $\widehat{\mathbf{X}}_{\mathrm{M_P}}$ are then finally obtained.

Code and data availability. The R package for MBCn is available at https://CRAN.R-project.org/package=MBC (Cannon, 2018b). R^2D^2 is not publicly available yet but only on demand from Mathieu Vrac. dOTC and MRec are publicly available at https://github.com/yrobink/SBCK (Robin, 2019).

Supplement. The supplement related to this article is available online at: https://doi.org/10.5194/esd-11-537-2020-supplement.

Author contributions. MV had the initial idea of the study. MV and BF designed the experiments and protocols. BF made all computations and figures. BF and MV made the analyses and interpretations. BF wrote the first complete draft of the article, iteratively revised by MV. Then AJC, YR and DA all revised each section of the article.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. Bastien François and Mathieu Vrac acknowledge financial support from the EUPHEME project. Mathieu Vrac also acknowledges support from the CoCliServ project and French "Convention de Service Climatique". Both EUPHEME and CoCliServ are part of ERA4CS, an ERA-NET initiated by JPI Climate and cofunded by the European Union. This work was supported by the metaprogram Adaptation of Agriculture and Forest to Climate Change (AAFCC) of the French National Research Institute for Agriculture, Food & Environment (INRAE).

Financial support. This research has been supported by the ERA4CS (grant no. 690462).

Review statement. This paper was edited by Ben Kravitz and reviewed by Jakob Zscheischler and one anonymous referee.

References

Bárdossy, A. and Pegram, G.: Multiscale spatial recorrelation of RCM precipitation to produce unbiased climate change scenarios over large areas and small, Water Resour. Res., 48, W09502, https://doi.org/10.1029/2011WR011524, 2012.

Beltrami, E.: Sulle funzioni bilineari, Giornale di Matematiche ad Uso degli Studenti Delle Universita, 11, 98–106, 1873.

Berg, P., Feldmann, H., and Panitz, H.-J.: Bias correction of high resolution regional climate model data, J. Hydrol., 448–449, 80–92, https://doi.org/10.1016/j.jhydrol.2012.04.026, 2012.

Bretherton, C. S., Widmann, M., Dymnikov, V. P., Wallace, J. M., and Bladé, I.: The Effective Number of Spatial Degrees of Freedom of a Time-Varying Field, J. Climate, 12, 1990–2009, https://doi.org/10.1175/1520-0442(1999)012<1990:TENOSD>2.0.CO;2, 1999.

Cannon, A. J.: Multivariate quantile mapping bias correction: an N-dimensional probability density function transform for climate model simulations of multiple variables, Clim. Dynam., 50, 31–49, https://doi.org/10.1007/s00382-017-3580-6, 2018a.

Cannon, A. J.: Multivariate Bias Correction of Climate Model Outputs, available at: https://CRAN.R-project.org/package=MBC (last access: 20 May 2019), 2018b.

Cannon, A. J., Sobie, S. R., and Murdock, T. Q.: Bias correction of simu- lated precipitation by quantile mapping: how well do methods preserve relative changes in quantiles and extremes?, J. Climate, 28, 6938–6959, https://doi.org/10.1175/JCLI-D-14-00754.1, 2015.

Casanueva, A., Bedia, J., Herrera García, S., Fernández, J., and Gutiérrez, J.: Direct and component-wise bias correction of multi-variate climate indices: the percentile adjustment function diagnostic tool, Climatic Change, 147, 411–425, https://doi.org/10.1007/s10584-018-2167-5, 2018.

Christensen, J. H., Boberg, F., Christensen, O. B., and Lucas-Picher, P.: On the need for bias correction of regional climate change projections of temperature and precipitation, Geophys. Res. Lett., 35, L20709, https://doi.org/10.1029/2008GL035694, 2008.

Clark, M., Gangopadhyay, S., Hay, L., Rajagopalan, B., and Wilby, R.: The Schaake Shuffle: A Method for Reconstructing Space–Time Variability in Forecasted Precipitation and Temperature Fields, J. Hydrometeorol., 5, 243–262, 2004.

Defrance, D., Ramstein, G., Charbit, S., Vrac, M., Famien, A. M., Sultan, B., Swingedouw, D., Dumas, C., Gemenne, F., Alvarez-Solas, J., and Vanderlinden, J.-P.: Consequences of rapid ice sheet melting on the Sahelian population vulnerability, P. Natl. Acad. Sci. USA, 114, 6533–6538, https://doi.org/10.1073/pnas.1619358114, 2017.

Dekens, L., Parey, S., Grandjacques, M., and Dacunha-Castelle, D.: Multivariate distribution correction of climate model outputs: A generalization of quantile mapping approaches: Multivariate distribution correction of climate model outputs, Environmetrics, 28, e2454, https://doi.org/10.1002/env.2454, 2017.

Déqué, M.: Frequency of precipitation and temperature extremes over France in an anthropogenic scenario: Model results and statistical correction according to observed values, Global Planet. Change, 57, 16–26, https://doi.org/10.1016/j.gloplacha.2006.11.030, 2007.

der Megreditchian, G.: Meteorological networks optimization from a statistical point of view, Comput. Stat. Data An., 9, 57–75, https://doi.org/10.1016/0167-9473(90)90071-O, 1990.

Dufresne, J.-L., Foujols, M.-A., Denvil, S., Caubel, A., Marti, O., Aumont, O., Balkanski, Y., Bekki, S., Bellenger, H., Benshila, R., Bony, S., Bopp, L., Braconnot, P., Brockmann, P., Cadule, P., Cheruy, F., Codron, F., Cozic, A., Cugnet, D., de Noblet, N., Duvel, J.-P., Ethé, C., Fairhead, L., Fichefet, T., Flavoni, S., Friedlingstein, P., Grandpeix, J.-Y., Guez, L., Guilyardi, E., Hauglustaine, D., Hourdin, F., Idelkadi, A., Ghattas, J., Joussaume, S., Kageyama, M., Krinner, G., Labetoulle, S., Lahellec, A., Lefebvre, M.-P., Lefevre, F., Levy, C., Li, Z. X., Lloyd, J., Lott, F., Madec, G., Mancip, M., Marchand, M., Masson, S., Meurdesoif, Y., Mignot, J., Musat, I., Parouty, S., Polcher, J., Rio, C., Schulz, M., Swingedouw, D., Szopa, S., Talandier, C., Terray, P., Viovy, N., and Vuichard, N.: Climate change projections using the IPSL-CM5 Earth System Model: from CMIP3 to CMIP5,

Clim. Dynam., 40, 2123–2165, https://doi.org/10.1007/s00382-012-1636-1, 2013.

Ehret, U., Zehe, E., Wulfmeyer, V., Warrach-Sagi, K., and Liebert, J.: HESS Opinions "Should we apply bias correction to global and regional climate model data?", Hydrol. Earth Syst. Sci., 16, 3391–3404, https://doi.org/10.5194/hess-16-3391-2012, 2012.

Famien, A. M., Janicot, S., Ochou, A. D., Vrac, M., Defrance, D., Sultan, B., and Noël, T.: A bias-corrected CMIP5 dataset for Africa using the CDF-t method – a contribution to agricultural impact studies, Earth Syst. Dynam., 9, 313–338, https://doi.org/10.5194/esd-9-313-2018, 2018.

Flamary, R. and Courty, N.: POT Python Optimal Transport library, available at: https://pythonot.github.io/ (last access: 8 June 2019), 2017.

Gudmundsson, L., Bremnes, J. B., Haugen, J. E., and Engen-Skaugen, T.: Technical Note: Downscaling RCM precipitation to the station scale using statistical transformations – a comparison of methods, Hydrol. Earth Syst. Sci., 16, 3383–3390, https://doi.org/10.5194/hess-16-3383-2012, 2012.

Guo, L.-Y., Gao, Q., Jiang, Z.-H., and Li, L.: Bias correction and projection of surface air temperature in LMDZ multiple simulation over central and eastern China, Adv. Clim. Change Res., 9, 81–92, https://doi.org/10.1016/j.accre.2018.02.003, 2018.

Guo, Q., Chen, J., Zhang, X., Shen, M., Chen, H., and Guo, S.: A new two-stage multivariate quantile mapping method for bias correcting climate model outputs, Clim. Dynam., 53, 3603–3623, https://doi.org/10.1007/s00382-019-04729-w, 2019.

Haddad, Z. and Rosenfeld, D.: Optimality of empirical Z-R relations, Q. J. Roy. Meteor. Soc., 123, 1283–1293, https://doi.org/10.1002/qj.49712354107, 1997.

Jordan, C.: Mémoire sur les formes bilinéaires, J. Math. Pures Appl., 19, 35–54, 1874a.

Jordan, C.: Sur la réduction des formes bilinéaires, C. R. Acad. Sci., Paris, France, 614–617, 1874b.

Kallache, M., Vrac, M., Naveau, P., and Michelangeli, P.-A.: Non-stationary probabilistic downscaling of extreme precipitation, J. Geophys. Res.-Atmos., 116, D05113, https://doi.org/10.1029/2010JD014892, 2011.

Lange, S.: Trend-preserving bias adjustment and statistical downscaling with ISIMIP3BASD (v1.0), Geosci. Model Dev., 12, 3055–3070, https://doi.org/10.5194/gmd-12-3055-2019, 2019.

Levina, E., Rothman, A., and Zhu, J.: Sparse estimation of large covariance matrices via a nested Lasso penalty, Ann. Appl. Stat., 2, 245–263, https://doi.org/10.1214/07-aoas139, 2008.

Maraun, D.: Bias Correction, Quantile Mapping, and Downscaling: Revisiting the Inflation Issue, J. Climate, 26, 2137–2143, https://doi.org/10.1175/JCLI-D-12-00821.1, 2013.

Marti, O., Braconnot, P., Dufresne, J.-L., Bellier, J., Benshila, R., Bony, S., Brockmann, P., Cadule, P., Caubel, A., Codron, F., de NOBLET, N., Denvil, S., Fairhead, L., Fichefet, T., Foujols, M.-A., Friedlingstein, P., Goosse, H., Grandpeix, J., Guilyardi, E., and Talandier, C.: Key features of the IPSL ocean atmosphere model and its sensitivity to atmospheric resolution, Clim. Dynam., 34, 1–26, https://doi.org/10.1007/S00382-009-0640-6, 2010.

Mehrotra, R. and Sharma, A.: Correcting for systematic biases in multiple raw GCM variables across a range of timescales, J. Hydrol., 520, 214–223, https://doi.org/10.1016/j.jhydrol.2014.11.037, 2015.

Mehrotra, R. and Sharma, A.: A Multivariate Quantile-Matching Bias Correction Approach with Auto- and Cross-Dependence across Multiple Time Scales: Implications for Downscaling, J. Climate, 29, 3519–3539, https://doi.org/10.1175/JCLI-D-15-0356.1, 2016.

Mehrotra, R. and Sharma, A.: A Resampling Approach for Correcting Systematic Spatiotemporal Biases for Multiple Variables in a Changing Climate, Water Resour. Res., 55, 754–770, https://doi.org/10.1029/2018WR023270, 2019.

Meyer, J., Kohn, I., Stahl, K., Hakala, K., Seibert, J., and Cannon, A. J.: Effects of univariate and multivariate bias correction on hydrological impact projections in alpine catchments, Hydrol. Earth Syst. Sci., 23, 1339–1354, https://doi.org/10.5194/hess-23-1339-2019, 2019.

Mezzadri, F.: How to generate random matrices from the classical compact groups, Not. Am. Math. Soc., 54, 592–604, 2007.

Michelangeli, P.-A., Vrac, M., and Loukos, H.: Probabilistic downscaling approaches: Application to wind cumulative distribution functions, Geophys. Res. Lett., 36, L11708, https://doi.org/10.1029/2009GL038401, 2009.

Möller, A., Lenkoski, A., and Thorarinsdottir, T. L.: Multivariate probabilistic forecasting using ensemble Bayesian model averaging and copulas, Q. J. Roy. Meteor. Soc., 139, 982–991, https://doi.org/10.1002/qj.2009, 2013.

Nahar, J., Johnson, F., and Sharma, A.: Addressing Spatial Dependence Bias in Climate Model Simulations—An Independent Component Analysis Approach, Water Resour. Res., 54, 827–841, https://doi.org/10.1002/2017WR021293, 2018.

Panofsky, H. and Brier, G.: Some applications of statistics to meteorology, Earth and Mineral Sciences Continuing Education, College of Earth and Mineral Sciences, The Pennsylvania State University, University Park, Pennsylvania, USA, 103 pp., 1958.

Piani, C. and Haerter, J.: Two dimensional bias correction of temperature and precipitation copulas in climate models, Geophys. Res. Lett., 39, L20401, https://doi.org/10.1029/2012GL053839, 2012.

Räty, O., Räisänen, J., Bosshard, T., and Donnelly, C.: Intercomparison of Univariate and Joint Bias Correction Methods in Changing Climate From a Hydrological Perspective, Climate, 6, 33, https://doi.org/10.3390/cli6020033, 2018.

Ribes, A., Thao, S., and Cattiaux, J.: Describing the relationship between a weather event and climate change: a new statistical approach, J. Climate, https://doi.org/10.1175/JCLI-D-19-0217.1, online first, 2020.

Robin, Y.: SBCK (Statistical Bias Correction Kit), GitHub, available at: https://github.com/yrobink/SBCK, last access: 20 May 2019.

Robin, Y., Vrac, M., Naveau, P., and Yiou, P.: Multivariate stochastic bias corrections with optimal transport, Hydrol. Earth Syst. Sci., 23, 773–786, https://doi.org/10.5194/hess-23-773-2019, 2019.

Santambrogio, F.: Optimal Transport for Applied Mathematicians, Birkhaüser, Basel, Switzerland, vol. 87, 2015.

Schefzik, R., Thorarinsdottir, T. L., and Gneiting, T.: Uncertainty Quantification in Complex Simulation Models Using Ensemble Copula Coupling, Stat. Sci., 28, 616–640, https://doi.org/10.1214/13-STS443, 2013.

Schuhmacher, D., Bähre, B., Gottschlich, C., Hartmann, V., Heinemann, F., and Schmitzer, B.: transport: Computation of Optimal

Transport Plans and Wasserstein Distances, r package version 0.11-1, available at: https://cran.r-project.org/package=transport (last access: 11 March 2020), 2019.

Stewart, G. W.: On the Early History of the Singular Value Decomposition, SIAM Rev., 35, 551–566, https://doi.org/10.1137/1035134, 1993.

Stott, P. A., Christidis, N., Otto, F. E. L., Sun, Y., Vanderlinden, J.-P., van Oldenborgh, G. J., Vautard, R., von Storch, H., Walton, P., Yiou, P., and Zwiers, F. W.: Attribution of extreme weather and climate-related events, WIRES Clim. Change, 7, 23–41, https://doi.org/10.1002/wcc.380, 2016.

Tobin, I., Vautard, R., Balog, I., Bréon, F.-M., Jerez, S., Ruti, P. M., Thais, F., Vrac, M., and Yiou, P.: Assessing climate change impacts on European wind energy from ENSEMBLES high-resolution climate projections, Climatic Change, 128, 99–112, https://doi.org/10.1007/s10584-014-1291-0, 2015.

Tramblay, Y., Ruelland, D., Somot, S., Bouaicha, R., and Servat, E.: High-resolution Med-CORDEX regional climate model simulations for hydrological impact studies: a first evaluation of the ALADIN-Climate model in Morocco, Hydrol. Earth Syst. Sci., 17, 3721–3739, https://doi.org/10.5194/hess-17-3721-2013, 2013.

Verkade, J., Brown, J., Reggiani, P., and Weerts, A.: Post-processing ECMWF precipitation and temperature ensemble reforecasts for operational hydrologic forecasting at various spatial scales, J. Hydrol., 501, 73–91, https://doi.org/10.1016/j.jhydrol.2013.07.039, 2013.

Vidal, J.-P., Martin, E., Franchistéguy, L., Baillon, M., and Soubeyroux, J.-M.: A 50-year high-resolution atmospheric reanalysis over France with the Safran system, Int. J. Climatol., 30, 1627–1644, https://doi.org/10.1002/joc.2003, 2010.

Villani, C.: Optimal transport – Old and new, in: Grundlehren der mathematischen Wissenschaften, Springer-Verlag, Berlin, Heidelberg, Germany, 992 pp., 2008.

Voisin, N., Schaake, J. C., and Lettenmaier, D. P.: Calibration and Downscaling Methods for Quantitative Ensemble Precipitation Forecasts, Weather Forecast., 25, 1603–1627, https://doi.org/10.1175/2010WAF2222367.1, 2010.

Vrac, M.: Multivariate bias adjustment of high-dimensional climate simulations: the Rank Resampling for Distributions and Dependences (R2D2) bias correction, Hydrol. Earth Syst. Sci., 22, 3175–3196, https://doi.org/10.5194/hess-22-3175-2018, 2018.

Vrac, M. and Friederichs, P.: Multivariate–Intervariable, Spatial, and Temporal–Bias Correction, J. Climate, 28, 218–237, https://doi.org/10.1175/JCLI-D-14-00059.1, 2015.

Vrac, M., Drobinski, P., Merlo, A., Herrmann, M., Lavaysse, C., Li, L., and Somot, S.: Dynamical and statistical downscaling of the French Mediterranean climate: uncertainty assessment, Nat. Hazards Earth Syst. Sci., 12, 2769–2784, https://doi.org/10.5194/nhess-12-2769-2012, 2012.

Vrac, M., Noël, T., and Vautard, R.: Bias correction of precipitation through Singularity Stochastic Removal: Because Occurrences matter, J. Geophys. Res.-Atmos., 121, 5237–5258, https://doi.org/10.1002/2015JD024511, 2016.

Wahl, T., Jain, S., Bender, J., Meyers, S., and Luther, M.: Increasing risk of compound flooding from storm surge and rainfall for major US cities, Nat. Clim. Chang., 5, 1093–1097, https://doi.org/10.1038/nclimate2736, 2015.

Wasko, C., Sharma, A., and Westra, S.: Reduced spatial extent of extreme storms at higher temperatures, Geophys. Res. Lett., 43, 4026–4032, https://doi.org/10.1002/2016GL068509, 2016.

Weedon, G. P., Balsamo, G., Bellouin, N., Gomes, S., Best, M. J., and Viterbo, P.: The WFDEI meteorological forcing data set: WATCH Forcing Data methodology applied to ERA-Interim reanalysis data, Water Resour. Res., 50, 7505–7514, https://doi.org/10.1002/2014WR015638, 2014.

Xu, C.-Y.: From GCMs to river flow: A review of downscaling methods and hydrologic modelling approaches, Prog. Phys. Geog., 23, 229–249, https://doi.org/10.1177/030913339902300204, 1999.

Yang, W., Gardelin, M., Olsson, J., and Bosshard, T.: Multi-variable bias correction: application of forest fire risk in present and future climate in Sweden, Nat. Hazards Earth Syst. Sci., 15, 2037–2057, https://doi.org/10.5194/nhess-15-2037-2015, 2015.

Yiou, P., Jézéquel, A., Naveau, P., Otto, F. E. L., Vautard, R., and Vrac, M.: A statistical framework for conditional extreme event attribution, Adv. Stat. Clim. Meteorol. Oceanogr., 3, 17–31, https://doi.org/10.5194/ascmo-3-17-2017, 2017.

Zscheischler, J. and Seneviratne, S.: Dependence of drivers affects risks associated with compound events, Sci. Adv, 3, e1700263, https://doi.org/10.1126/sciadv.1700263, 2017.

Zscheischler, J., Westra, S., Hurk, B., Seneviratne, S., Ward, P., Pitman, A., AghaKouchak, A., Bresch, D., Leonard, M., Wahl, T., and Zhang, X.: Future climate risk from compound events, Nat. Clim. Chang, 8, 469–477, https://doi.org/10.1038/s41558-018-0156-3, 2018.

Zscheischler, J., Fischer, E. M., and Lange, S.: The effect of univariate bias adjustment on multivariate hazard estimates, Earth Syst. Dynam., 10, 31–43, https://doi.org/10.5194/esd-10-31-2019, 2019.

2.3 Summary and reminder of the conclusions

In this chapter, an intercomparison study of four MBC methods (R^2D^2, dOTC, MBC-n and MRec) designed to adjust multivariate properties of climate simulations has been carried out. I selected these MBC methods for their differences in terms of methodologies, statistical techniques and assumptions to present a representative and comprehensive picture of their different characteristics. A univariate bias correction method (CDF-t) has also been implemented to assess the benefits of considering multivariate bias correction methods.

I applied each multivariate method in three dimensional configurations to adjust simulated temperature and precipitation time series from the IPSL model: a 2d-version to adjust temperature and precipitation inter-variable dependencies, a Spatial-version to adjust the spatial temperature and precipitation dependence separately and a Full-version to adjust both inter-variable and spatial relationships together. A wide range of metrics has been used to evaluate bias adjustment outputs with respect to references and initial climate model data and analyse the corrections of univariate distributions, inter-variable correlations, spatial correlations and temporal properties by using a first protocol of bias correction. In general, R^2D^2, dOTC, MBCn and MRec algorithms showed a great ability to adjust the statistical properties associated with the corresponding objectives of the dimensional configurations. Indeed, in addition to adjusting univariate distributions, applying each MBC method using the 2d-, Spatial- and Full- versions permitted to adjust, respectively, inter-variable, spatial and inter-variable/spatial dependence of climate simulations reasonably well. However, caution has to taken when applying these methods in high-dimensional contexts. Indeed, for MBC-n and MRec, some instabilities have been found when adjusting a very high number of variables at the same time (i.e., when applying these methods using the Spatial- and Full-versions). These instabilities involved strong deterioration of inter-variable and spatial properties, thus affecting the quality of corrected outputs.

Concerning temporal properties, none of the MBC methods used in this study were designed to adjust this statistical aspect. By not taking into account this aspect in their bias correction procedure, it resulted in generating corrected outputs with unexpected temporal behaviors or weak temporal dependencies compared to observations. Thus, for applications where temporal properties are crucial, the use of the presented MBC methods is not recommended and the adding of some additional constraints on temporality is probably necessary (see section 2.4).

The ability of BC methods to reproduce simulated multidimensional change, i.e., nonstationary, properties has also been assessed by using a second protocol. While dOTC, MBCn and MRec are designed by construction to take some of the multidimensional properties changes into account in their bias correction procedure, R^2D^2 assumes multivariate properties (i.e., inter-variable and spatial dependence) to be stable in time. Hence, if simulated dependence changes are considered as relevant to be reproduced in the corrections, dOTC, MBCn and MRec must be preferred by practitioners. However, if no dependence changes are expected, or if changes in the simulations are not considered as relevant, R^2D^2 would have to be preferred to produce corrected outputs with stationarity of multivariate properties.

The study permits to draw general recommendations for end users in order to provide them with help in the choice of BC methods depending on their needs. This work not only provides an intercomparison framework to which other MBC methods can be easily added but also permits to identify future avenues for the development of new multivariate BC methods.

2.4 Perspectives

The work developed in this chapter leads to several perspectives. I will provide more details to some of the perspectives mentioned in the conclusion of the article (subsection 6.3). Other perspectives that are not mentioned in the article will be also discussed and some will be partly addressed in the rest of the manuscript.

2.4.1 A look back at some of the perspectives of the article

Influences of MBC on impact studies

Having assessed the performance of multivariate BC methods for adjusting the statistical properties of climate models, an important perspective of this study is to evaluate the potential influence of these adjustments on the results of impact studies. However, quantifying the effect of bias adjustment on impact modelling can be difficult (Papadimitriou et al., 2017). While the quality of bias correction of climate variables can be assessed against observations before being used in impact models, a lack of observed climate impacts data makes it hard to assess the modelled impacts (Cramer et al., 2014). In addition, many impact models rely on non-linear interactions between multiple climate variables at various spatial and temporal scales, where small variations in inputs can lead to large differences in results. Thus, in addition to initial uncertainties in the choice of GCMs, forcing scenario and impact models, bias correction introduces a level of uncertainty in the modelling chain that needs to be explored (e.g., Tao et al., 2018). The appropriateness of bias adjustment for impact modelling requires to be answered on a case-by-case basis. Laux et al. (2021) found that univariate BC approaches can improve the performance of climate projections in agricultural impact studies, despite large BC-inherent uncertainties. Concerning multivariate bias correction, some studies have shown that multivariate BC methods do not necessarily lead to improve results for specific regional impact studies (Yang et al., 2015; Räty et al., 2018). However, other studies demonstrated the effectiveness of the adjustment of inter-variable relationships on multivariate hazard indicators (e.g., Cannon, 2018; Zscheischler et al., 2019) or hydrological impact projections (e.g., Meyer et al., 2019; Guo et al., 2020; Van de Velde et al., 2022), although less pronounced in a non-stationary context (Guo et al., 2020; Van de Velde et al., 2022). Conducting additional impact studies with corrected inputs provided by different MBCs in several dimensional configurations (i.e., not only adjusting inter-variable relationships but also spatial or inter-variable/spatial relationships) is an important perspective that would permit to further cover uncertainties in the choice and the application of BC methods.

Are simulated dependence changes reliable?

Although results concerning multidimensional changes properties from MBC methods are sometimes difficult to evaluate, the intercomparison study concludes that, as expected, MBC methods assuming nonstationarity are more likely to present changes in agreement with the climate models (see subsection 5.5 of the article). The development and the use of these methods assuming nonstationarity such as MBCn (Cannon, 2018) or dOTC (Robin et al., 2019) is justified by the fact that, despite climate simulations have biases in multivariate properties, simulated changes of statistical properties for future periods are supposed to be driven by relevant physical processes providing key information concerning climate changes. However, the relevance of such simulated multidimensional changes can still be discussed, which consequently calls into

question the use of these MBC methods. Of course, determining whether future simulated changes are relevant is difficult in practice. However, work can be carried out to evaluate at least if climate models show appropriate evolution of multivariate properties with respect to reanalyses or observations over historical periods. For example, by investigating temperature vs. precipitation correlation changes over Europe, Vrac et al. (2021) showed the inability of a climate multi-model ensemble (a subset of CMIP6 ensemble comprising 11 models) to reproduce the ERA5 historical changes in correlations. Thus, for future periods, there is little confidence that the simulated changes of temperature-precipitation dependence are realistically simulated by models from CMIP6. In this context, MBC methods assuming stationarity of inter-variable structure such as R^2D^2 (Vrac, 2018; Vrac and Thao, 2020) might be preferred. An interesting perspective could be to further investigate the relevance of simulated changes of other inter-variable relationships (e.g. wind and precipitation) over historical periods, but also to extend this assessment framework to spatial and temporal changes provided by ensembles of climate simulations.

Considering temporal adjustments

An other important perspective from the intercomparison study presented in this chapter is the development of new multivariate BC methods providing adjustments of temporal properties of simulated climate variables in addition to the correction of inter-variable and spatial properties. Based on the results of the study, several multivariate BC methods able to adjust some of the temporal properties have been developed (e.g., Vrac and Thao, 2020; Robin and Vrac, 2021), complementing the few existing methods already available (Nguyen et al., 2019; Mehrotra and Sharma, 2019). These new methods show an ability to largely reduce biases in temporal properties, while still correctly adjusting inter-variable and spatial dependence structures. However, developing MBCs methods that are able to adjust a wide range of statistical properties (such as inter-variable, spatial and temporal properties) raises the question of what has been preserved from the simulations in the final corrections. At the end, what is left of climate models if so many properties are corrected? This point will be further discussed later in Chapter 5.

2.4.2 Additional perspectives

Additional perspectives that are not mentioned in the article can be envisaged.

Influence of atmospheric circulation on local-scale dependencies

It is well known that statistical properties of climate variables are largely influenced by large-scale atmospheric circulations (e.g., Yiou et al., 2018; Jézéquel et al., 2020; Faranda et al., 2020; Rust et al., 2013). For example, atmospheric blocking events are linked to the occurrence of extreme temperature events such as heat waves (e.g., Lenggenhager and Martius, 2020). Hence, if circulation dynamics, as well as their physical relationships with local-scale phenomena, is misrepresented within climate models, it can lead to major biases in marginal and dependence properties of climate variables. For instance, Maraun et al. (2021) showed that biases in the occurrence of synoptic-scale weather types can propagate to univariate properties of temperature and precipitation. In addition, Vrac et al. (2022) found that the dependence of temperature and precipitation conditionally on atmospheric regimes is misrepresented within climate models.

Developing new MBC methods that are able to take into account large-scale information into their correction procedure is an important perspective to explore. By driving the correction using some physical processes, the bias-corrected outputs obtained from these MBC methods would be more realistic. However, before designing these new MBC methods, it will be necessary to carefully define the statistical features of atmospheric circulations that need to be adjusted.

Machine Learning: promising for multivariable?

Other bias correction and downscaling techniques, based on less statistical but more empirical tools, have also emerged in recent years (e.g., Moghim and Bras, 2017; Sachindra et al., 2018; Vandal et al., 2019; Baño-Medina et al., 2020). Machine Learning tools have become an important part of climate modelling and simulation processing as they proved to be very effective in modelling complex relationships between statistical variables for a wide variety of scientific problems. In particular, an interesting application of Machine Learning tools made newspapers headlines in 2018: the first auction of a painting produced by neural networks[1]. It demonstrated that these statistical tools are able to generate images such as master paintings quite realistically. As explained by Cannon (2018), multivariate bias correction can be seen as an image problem that aims to transform inappropriate images from model simulations to more realistic images similar to a reference. Consequently, one can wonder if these Machine Learning tools able to produce realistic paintings are relevant in a context of bias correction of climate simulations. In particular, are they able to reproduce spatial properties of climate variables as they reproduce colour information and structures from painting masters? Favouring the development of new multivariate BC methods based on Machine Learning seems to be a relevant research direction to explore. During my thesis, I have chosen to explore this direction. The work carried out is presented in the following chapter.

[1] https://www.nytimes.com/2018/10/25/arts/design/ai-art-sold-christies.html

Résumé

Contexte et objectifs

Les méthodes de correction de biais multivariés (MBC) sont des méthodes statistiques permettant d'ajuster les propriétés statistiques multivariés des modèles climatiques par rapport à des observations ou des réanalyses. Ces méthodes n'ont cependant été développées que récemment, et sont parfois encore mal appréhendées par les utilisateurs. Dans ce chapitre, une étude d'intercomparaison de quatre méthodes MBC a été réalisée pour évaluer leurs avantages et inconvénients et ainsi aider les utilisateurs à choisir quelles méthodes utiliser selon leurs besoins.

Méthodes

Quatre méthodes multivariées (R^2D^2, dOTC, MBC-n et MBC-n) ayant des différences en termes de méthodologies, de techniques statistiques et d'hypothèses sont comparées. Nous avons appliqué chaque méthode dans différentes configurations dimensionnelles pour ajuster les données de température et de précipitation simulées par le modèle climatique IPSL. Un large panel de métriques statistiques a été utilisé pour évaluer les performances de corrections des méthodes par rapport à des données de références. Ainsi, les performances d'ajustements en terme de distributions univariées, de corrélations inter-variables, de corrélations spatiales et de propriétés temporelles ont pu être évaluées. De surcroît, une méthode de correction de biais univariée (CDF-t) a aussi été implémentée pour évaluer les avantages de la prise en compte de l'aspect multivarié dans la correction de biais.

Résultats

De manière général, les algorithmes R^2D^2, dOTC, MBCn et MRec ont montré une grande capacité à ajuster les propriétés univariées, inter-variables et spatiales des simulations selon les différentes configurations dimensionnelles. Cependant, l'application de ces méthodes dans des contextes hautement dimensionnels doit se faire avec prudence. En effet, pour MBC-n et MRec, certaines instabilités ont été constatées lors de l'ajustement simultané d'un très grand nombre de variables, impliquant une forte détérioration de la qualité des sorties corrigées. Aussi, en ce qui concerne les propriétés temporelles, aucune des méthodes utilisées dans cette étude n'a été conçue pour ajuster cet aspect. Cela résulte à obtenir des données corrigées avec de faibles dépendances temporelles par rapport aux observations. Ainsi, utiliser ces méthodes MBCs n'est pas recommandé pour des applications où les propriétés temporelles sont essentielles.

L'étude ne cherchait pas à établir la "meilleure" méthode MBC, mais a plutôt permis de formuler des recommandations générales pour les utilisateurs afin de les aider à choisir quelles méthodes de corrections utiliser en fonction de leurs besoins. Ce travail fournit non seulement un cadre d'intercomparaison auquel d'autres méthodes MBC peuvent être facilement ajoutées, mais permet également d'identifier des pistes de développements futures pour de nouvelles méthodes multivariées.

Chapter 3

Development of a new multivariate bias correction method using Machine Learning

Developing MBC methods able to adjust temporal properties of climate simulations is an interesting perspective from the previous chapter. However, when the intercomparison study of MBC methods was completed, new methods for the adjustments of temporal properties were being developed by colleagues, which have since been published (Vrac and Thao, 2020; Robin and Vrac, 2021). Thus, I rather decided to explore the applicability of Machine Learning tools for multivariate bias correction in high-dimensional context, i.e., not necessarily for temporal properties adjustments. This resulted in developing a new multivariate bias correction method based on a particular neural network-based algorithm named CycleGAN (Zhu et al., 2017). This new MBC method, named MBC-CycleGAN, is designed to adjust the spatial properties of climate simulations. The core of this chapter is an article published in the scientific journal *Climate Dynamics* presenting the new method MBC-CycleGAN. The article is preceded by a brief introduction on neural networks. It is then completed by a section presenting some details on the different failures encountered when developing the new method. Finally, a section will recall the main conclusions of this work and will discuss some perspectives.

3.1 A brief overview of neural networks

Inspired by how information flows through the brain, McCulloch and Pitts (1943) designed the first artificial neural network to perform logic computations. The evolution of this first — and naive — neural network has then progressed through several stages, e.g., by improving training effectiveness (Rosenblatt, 1958; Werbos, 1974; Rumelhart et al., 1986). Neural networks have become today a powerful statistical tool able to reproduce complex and non-linear relationships between variables.

A neural network is a computational learning system that uses a network of neurons, or nodes. The nodes are organised in layers and connections between nodes are modelled as weights. Through these connections, the nodes perform successive operations that result in a non-linear mapping between a set of explanatory variables (input layer) and response variables (output

layer). During training, the goal of the neural network is to find the optimal set of weights needed to explain the response variables from the explanatory variables. Neural networks have been used for a variety of applications, including regression analysis (e.g., Specht, 1991), pattern and sequence recognition (e.g., Lawrence et al., 1997) and data processing such as filtering (e.g., Xue et al., 1992).

To illustrate the different elements involved in neural networks architecture, I propose to repeat the explanation for a simple application: animal recognition in pictures. Fig. 3.1 shows a possible neural network architecture for such an application. First, pixels information from an image of an animal (here, chameleon) is given to the input layer and is allocated among the nodes (green dots). Then, input data get multiplied by weight values (black lines) and the results is passed to the next layer (hidden layer, orange dots). This process is repeated within the successive hidden layers until the output layer (purple dots). The output layer is responsible for producing the final result. For a classification problem like the one presented, each node of the output layer can correspond for example to the probability of having a specific category of animal (such as dog, cat or chameleon) on the picture. When the neural network is trained, probability that a chameleon is present within the image is derived. This probability output is then compared with the "truth" (i.e., the value 1 as this is a picture of a chameleon). Depending on whether the predicted probability the value was close to the "true value", weight values are adjusted so that the neural network will better predict when encountering another chameleon picture.

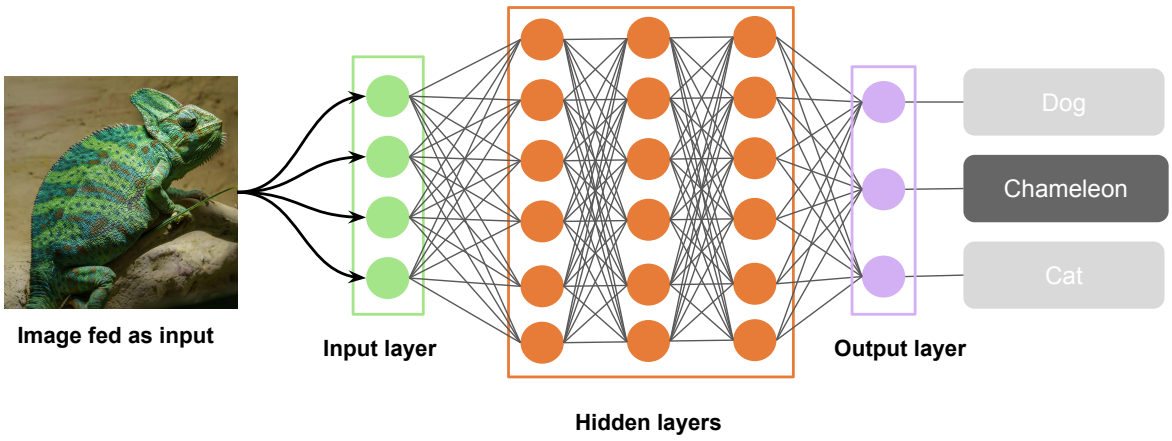

Figure 3.1: Illustration of a simple neural network topology for an animal recognition application. Illustration inspired by the website https://www.analyticsvidhya.com/. Use and modification of the photography "Chameleon Chamäleon" taken by Matthias, CC BY 3.0.

Due to computational limitations, the use of neural networks was limited for many years to simple architectures with no more than two hidden layers, and were thus not able to model very complex data relationships. A major breakthrough that has allowed neural networks achieving better performances is linked to the development of backward propagation algorithms (e.g., Rumelhart et al., 1986). These algorithms permit the information for updating weights during training to flow effectively through the network. Also, the major advances in computing technology in the 2000s allowed to define more complex neural network architectures with more layers. These complex neural networks also named deep neural networks were then successfully trained (e.g., Hinton et al., 2006; Ranzato et al., 2006; Bengio et al., 2007), showing their

ability to extract highly complex features from datasets. The promise of deep learning has been definitively demonstrated to the world when the deep neural network AlexNet (Krizhevsky et al., 2017) composed of eight layers won in 2012 the most challenging competition in visual recognition by surpassing a wide variety of other machine learning methods. In 2016, the deep learning AlphaGo has attracted worldwide attention when it beat the legendary Go master Lee Sedol[1]. The board game of Go is one of the most complex games ever designed, with an astronomical amount of possible game configurations. By playing against himself during training ("self-taught"), AlphaGo showed the exceptional ability of neural networks to solve very complex problems. An overview of major improvements obtained with Deep Learning models is available in Schmidhuber (2015).

3.2 CNNs in climate science

Among the different types of neural networks, convolutional neural networks (CNNs, see e.g., Lecun and Bengio, 1995) showed that they can learn with great performances complex spatial structures. CNNs are a specific type of neural networks in which convolutional operations through the application of filters are performed to extract features from images. Fig. 3.2 shows a possible CNN architecture for the same animal recognition problem. Filters are matrices of weights (red squares in Fig. 3.2) that scan the input image by performing matrix multiplications. The results from these operations are then passed to the next convolutional layer. By updating matrices of weights during training, CNNs are able to extract *local* spatial features from an image and to combine them to learn higher-level image features. The extracted feature signals can then be utilised, e.g., for classification using a fully connected neural network (orange and purple nodes in Fig. 3.2). The ability of CNNs to learn spatial structures comes from the matrix operations they perform over the dimensions of the input (e.g., over the 2 dimensions of an image).

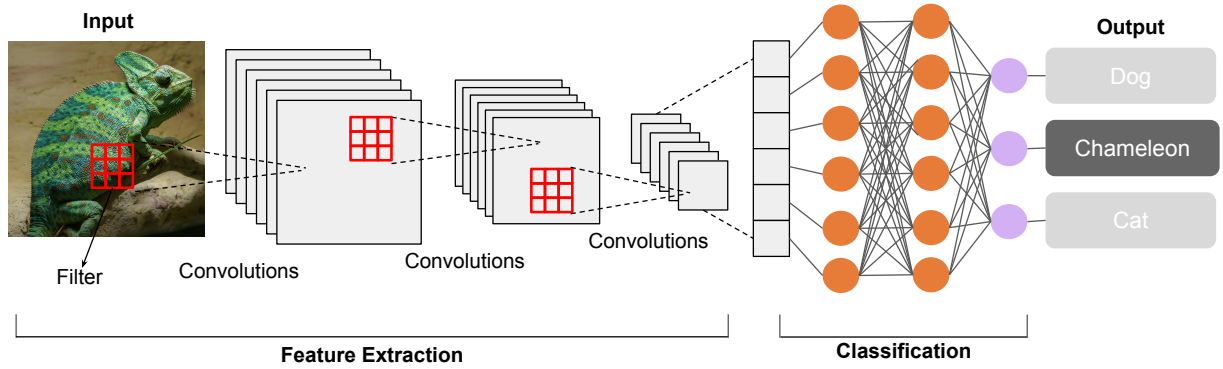

Figure 3.2: Illustration of a simple CNN topology for an animal recognition application. Use and modification of the photography "Chameleon Chamäleon" taken by Matthias, CC BY 3.0.

Initially developed for computer vision applications to consider spatial dependencies (e.g., Lecun and Bengio, 1995; Szegedy et al., 2015; He et al., 2016), CNNs have been applied with suc-

[1] https://www.theatlantic.com/technology/archive/2016/03/the-invisible-opponent/475611/

cess in various fields of climate science to complement physical knowledge or post-process climate datasets. For instance, CNNs have already shown promising results in replacing computationally-demanding physical parameterizations for climate models (Han et al., 2020). They also permitted to provide accurate weather forecasts (Scher and Messori, 2018, 2019; Weyn et al., 2020; Hewage et al., 2021; Li et al., 2022) or to emulate atmospheric dynamics (Shi et al., 2015; Chapman et al., 2019; Babaousmail et al., 2021; Doury et al., 2022) through the modelling of climate variables relationships. CNNs have also been applied to downscale climate data (Vandal et al., 2017; Baño-Medina et al., 2020), providing climate information at higher spatial resolution. A recent overview of Deep Learning applications including CNNs for Earth system science can be found in Reichstein et al. (2019).

3.3 CNNs and GANs for MBC: how did I get there?

Although the problem of bias correction can be seen as an image problem (Cannon, 2018), CNNs were not used to adjust multivariate properties of climate simulations despite showing great performances for the modelling of spatial relationships in other applications. Indeed, one can think about applying CNNs to learn the transfer of simulated daily maps (seen as images) with inappropriate spatial features to more realistic images with spatial properties resembling those observed. As explained in Chapter 1, this transformation using CNNs models could be determined based on either perfect prognosis (PP) or model output statistics (MOS) approaches (Maraun et al., 2010). The PP approach consists in learning the synchronous relationships between a variable of interest from references (predictand) and one or several observed variables (predictors) before applying these relationships to predictors from climate simulations. Translated into the language of machine learning, the PP approach determines the statistical transformation in a *supervised* context. As PP approach assumes that predictors are perfectly simulated by the climate model, which, in practice, is often not met (Maraun, 2016), the MOS approach has been developed. For MOS approach, observed and simulated variables are not required to be in synchrony to learn to adjust statistical biases. Translated into the language of machine learning, the MOS approach determines the statistical transformation in a *unsupervised* context. Our objective being to develop a MBC approach (i.e. linking simulations and observations in a statistical way and without synchronicity in time), I was interested in unsupervised CNNs algorithms.

Designing a new MBC method using CNNs in an unsupervised context adds an additional complexity. Such MBC methods would consist in linking daily maps (or images) from a source dataset (climate simulations) to a target dataset (references) without one-to-one correspondence. Algorithms able to perform image-to-image translations in unsupervised context are not many and the most famous one is the Generative Adversarial Networks-based algorithm. In their most basic formulation, Generative Adversarial Networks (GANs, Goodfellow et al., 2014) based on CNNs are made up of two neural networks set-up in competition and trained together, hence the name "adversarial". The first neural network is named the "generator", and aims at learning to produce new samples that resemble those in the target dataset. In the original formulation of GANs, the generator takes Gaussian random vectors as inputs to generate new images. The second neural network, named the "discriminator" is a classifier that returns the probability that a given image comes from the target dataset. By putting them in competition during training, the generator will try to improve itself to "fool" the discriminator with its generated images, while the discriminator will try not to be fooled by the generator and to better discriminate

images from the target dataset. CNN-based GANs demonstrated impressive results in computer vision problems, e.g., to produce realistic paintings as discussed briefly in Chapter 2, and more generally to learn representations from various image dataset (Radford et al., 2016).

The formulation of GANs explained above is unconditional: by taking Gaussian random vectors as inputs, the generator is not particularly constrained to produce a certain type of images. Conditional formulations of GANs have then been developed, which offer the possibility to condition the generation of new images with additional information such as class labels (e.g., Mirza and Osindero, 2014; Gauthier, 2014; Denton et al., 2015) or images (e.g., Yoo et al., 2016; Isola et al., 2017; Kim et al., 2017; Zhu et al., 2017). In particular, Zhu et al. (2017) designed a specific formulation of image-conditional GANs for unsupervised image-to-image translation problems named CycleGAN. By defining a specific optimisation problem for the training of the generator and the discriminator that will be explained in the following article, generators from CycleGAN architectures are able to transform images from the source dataset so that they resemble to images from the target dataset, even when no one-to-one correspondence between source and target images exist. Thus, CycleGAN meets all the requirements of bias correction using the MOS approach. For these reasons, it was natural that I looked at adapting CycleGAN for the development of a new multivariate BC using neural networks.

In the following article, I explore the potential of Generative Adversarial Networks combined with CNNs to adjust climate simulations. Based on these innovative tools, I provide a new method named MBC-CycleGAN for the adjustments of spatial properties of climate simulations. I test my new method for the adjustment of temperature and precipitation time series from IPSL simulations over the region of Paris in both the PP and MOS approaches and compare the results obtained with other state-of-the-art MBC methods.

3.4 Article published in *Climate Dynamics:* Adjusting spatial dependence of climate model outputs with cycle-consistent adversarial networks

Climate Dynamics
https://doi.org/10.1007/s00382-021-05869-8

Adjusting spatial dependence of climate model outputs with cycle-consistent adversarial networks

Bastien François[1] · **Soulivanh Thao**[1] · **Mathieu Vrac**[1]

Received: 4 March 2021 / Accepted: 25 June 2021

Abstract
Climate model outputs are commonly corrected using statistical univariate bias correction methods. Most of the time, those 1d-corrections do not modify the ranks of the time series to be corrected. This implies that biases in the spatial or inter-variable dependences of the simulated variables are not adjusted. Hence, over the last few years, some multivariate bias correction (MBC) methods have been developed to account for inter-variable structures, inter-site ones, or both. As proof-of-concept, we propose to adapt a computer vision technique used for Image-to-Image translation tasks (CycleGAN) for the adjustment of spatial dependence structures of climate model projections. The proposed algorithm, named MBC-CycleGAN, aims to transfer simulated maps (seen as images) with inappropriate spatial dependence structure from climate model outputs to more realistic images with spatial properties similar to the observed ones. For evaluation purposes, the method is applied to adjust maps of temperature and precipitation from climate simulations through two cross-validation approaches. The first one is designed to assess two different post-processing schemes (Perfect Prognosis and Model Output Statistics). The second one assesses the influence of nonstationary properties of climate simulations on the performance of MBC-CycleGAN to adjust spatial dependences. Results are compared against a popular univariate bias correction method, a "quantile-mapping" method, which ignores inter-site dependencies in the correction procedure, and two state-of-the-art multivariate bias correction algorithms aiming to adjust spatial correlation structure. In comparison with these alternatives, the MBC-CycleGAN algorithm reasonably corrects spatial correlations of climate simulations for both temperature and precipitation, encouraging further research on the improvement of this approach for multivariate bias correction of climate model projections.

Keywords Bias correction · Spatial dependence · Post-processing · Climate simulations · Generative adversarial networks · Model output statistics

1 Introduction

With ongoing climate change, mitigation and adaptation strategies have to be anticipated by decision makers in order to reduce potential future consequences of climate change on human societies and activities (IPCC 2014). Such consequences are commonly assessed through climate change impact studies, for instance in hydrology (e.g., Bates et al. 2008), agronomy (e.g., Wheeler and von Braun 2013) or epidemiology (e.g., Caminade et al. 2014). They rely on impact model simulations, the quality of which highly depends on the reliability of the climate information used as inputs (e.g., Muerth et al. 2013; Ramirez-Villegas et al. 2013). Besides observations, global and regional climate models (GCM and RCM) are the major tools to understand the climate system and its evolutions in the future (Randall et al. 2007; Reichler and Kim 2008). However, despite considerable improvements in climate modelling, climate simulations often remain biased compared to observations: even for the current climate, key statistical features such as mean, variance or the dependence structures between physical variables or between sites can differ from those calculated for observational references (e.g., Eden et al. 2012; Cattiaux et al. 2013; Mueller and Seneviratne 2014). Consequently, biases are expected to be present in climate projections for future periods, making bias correction an often unavoidable

✉ Bastien François
bastien.francois@lsce.ipsl.fr

[1] Laboratoire des Sciences du Climat et l'Environnement (LSCE-IPSL) CNRS/CEA/UVSQ, UMR8212, Université Paris-Saclay, Gif-sur-Yvette, France

Published online: 30 July 2021

data pre-processing step for impact studies (e.g., Christensen et al. 2008; Maraun et al. 2010; Teutschbein and Seibert 2012).

In the recent years, many statistical bias correction (BC) methods have been developed that aim to correct (selected features of) the distribution of climate variables. The idea of statistical bias correction is to find a mathematical transformation that makes climate simulations have similar statistical properties as a reference dataset over the historical period, and then apply this transformation for the modeled projection. Such transformations may be determined with statistical models based on either perfect prognosis (PP) or model output statistics (MOS) approaches (Maraun et al. 2010). The PP approach consists in determining the statistical link between a variable of interest from references (predictand) and one or several observed variables (predictors) occurring at the same time. Simultaneous values of predictand and predictors are indeed required to implement the PP approach and learn the (synchronous) relationships between them. By applying these relationships to predictors from climate simulations, this approach implicitly makes the assumption that these predictors are realistically simulated (Wilks 2006). In the MOS approach, observed and simulated variables are not considered to be synchronized in time, and biases relate to differences in some statistics (such as means or variances) or in distributions between references and modeled climate variables. Adjustments can be made to the simulated mean (e.g., Delta method, Xu 1999), variance (e.g., simple scaling adjustment, Berg et al. 2012) and also all moments of higher order and percentiles (e.g.,"quantile-mapping", Haddad and Rosenfeld 1997; Déqué 2007; Gudmundsson et al. 2012). In particular, quantile-mapping technique has received a keen interest since it permits for adjusting not only the mean and variance but also the whole distribution of climate variables. It has been conducive to the development of many variants (e.g., Vrac et al. 2012, 2016; Tramblay et al. 2013; Cannon et al. 2015), and applied for various studies (e.g., Vigaud et al. 2013; Defrance et al. 2017; Bartok et al. 2019; Tong et al. 2020). However, such BC methods are designed to only correct statistical aspects of univariate distributions. Simulated variables are indeed adjusted separately for each physical variable at each specific location. Thus, potential biases in the spatial dependence structure of modeled variables are not corrected (e.g., Wilcke et al. 2013), which can generate corrections with inappropriate multivariate situations and can affect subsequent analyses that depend on spatial characteristics of climate variables (e.g., Zscheischler et al. 2019). For instance, this can occur with flood risk assessment, that depends on spatial (and temporal) properties of precipitation, soil moisture and river flow (Vorogushyn et al. 2018) or with drought-related impacts, that depend on complex interaction of natural and anthropogenic processes (Van Loon et al. 2016). It is hence crucial to provide end users with bias corrections of climate simulations that present not only relevant 1-dimensional information at each individual site but also appropriate spatial representation.

Over the last years, a few multivariate bias correction (MBC) methods have been developed to address the issues of biases in multivariate dependencies. Not only do these methods correct marginal properties of simulated variables, they are also designed to adjust statistical dependencies between variables. Although it has been found for specific cases that MBC methods do not particularly outperform univariate ones for the adjustment of dependencies between multiple variables (Räty et al. 2018), this finding cannot be generalized to all applications and methods. For instance, François et al. (2020) showed the added value of MBC to improve inter-variable dependence and spatial structures for temperature and precipitation over Europe. More generally, MBCs could be of great interest for compound events studies, where dependencies between drivers of extreme events with large impacts are crucial to evaluate their risks (Zscheischler et al. 2018).

A categorization of MBC methods in three main families of approaches has been proposed in the literature (e.g., Vrac 2018; François et al. 2020):

- the "marginal/dependence" correction approach, that consists of MBC methods adjusting in two distinct steps, i.e. separately, marginal distributions and multivariate dependencies of climate simulations (e.g., Bárdossy and Pegram 2012; Mehrotra and Sharma 2016; Hnilica et al. 2017; Nahar et al. 2018; Cannon 2018; Nguyen et al. 2019; Guo et al. 2019; Vrac and Thao 2020).
- the "successive conditional" category, made up of MBC methods performing successive univariate corrections of climate variables conditionally on the previously adjusted ones (e.g., Piani and Haerter 2012; Dekens et al. 2017).
- the "all-in-one" correction approach, that adjusts directly the whole statistical distribution (i.e. both univariate and multivariate properties) of climate simulations at the same time (e.g., Robin et al. 2019).

Based on this categorization, François et al. (2020) performed an intercomparison and critical review of MBC methods. It presents a global picture of the performances of MBCs in terms of multivariate adjustments of climate simulations, as well as the different assumptions and statistical techniques used.

In parallel, i.e., in contexts other than bias correction, over the last decades, machine learning techniques have emerged as a promising approach to model highly nonlinear and complex relationships between statistical variables. Major improvements have been obtained with Deep Learning models (see the overview of Schmidhuber 2015),

which have proved to be efficient to extract high-level feature information from various datasets. In particular, convolutional neural networks (CNNs, see e.g., Lecun and Bengio 1995) showed that they can capture with great performances complex spatial structures. Initially developed for computer vision problems (e.g., Szegedy et al. 2015; He et al. 2016), they found numerous applications in climate sciences: for instance for weather forecast prediction uncertainty (Scher and Messori 2018), emulations of atmospheric dynamics (Shi et al. 2015; Scher and Messori 2019; Chapman et al. 2019), detection of extreme weather events (Liu et al. 2016; Racah et al. 2017) and statistical downscaling (Vandal et al. 2017; Rodrigues et al. 2018; Baño-Medina et al. 2020). A recent overview of Deep Learning applications for Earth system science is offered by Reichstein et al. (2019).

Recently, a new class of artificial neural networks, named Generative Adversarial Networks (GANs; Goodfellow et al. 2014), has led to tremendous interests due to their ability to infer high dimensional probability distributions. Initially, this machine-learning method has been developed for estimating the distribution of images from a target dataset, with the aim of sampling new (and unseen) images from this distribution. GANs, implemented with deep convolutional neural networks, have achieved impressive results in computer vision problems (e.g., Radford et al. 2016) and are a subject of active research to improve computing architectures (e.g., Salimans et al. 2016; Karras et al. 2018; Menick and Kalchbrenner 2018) and optimization techniques (e.g., Mao et al. 2017; Arjovsky et al. 2017; Roth et al. 2017). Conditional formulations of GANs have also been developed, for which additional information, such as class labels or images, can serve as inputs to condition the generation of the new images (e.g., Mirza and Osindero 2014; Gauthier 2014; Denton et al. 2015; Kim et al. 2017; Isola et al. 2017). In particular, image-conditional GANs permit to perform image-to-image translation tasks by learning how to map the statistical distribution of one set of images (source dataset) to the statistical distribution of another set (target dataset). Depending on the correspondence between images of the source and target datasets, different versions of image-conditional GANs have been developed. When all the images are paired (i.e., there is a known one-to-one correspondence between every images of the source and target datasets), conditional GANs are trained by supervised learning (Yoo et al. 2016; Isola et al. 2017). When only a few images are paired, semi-supervised is used (Gan et al. 2017) and when all points are unpaired, only unsupervised learning can be applied (Kim et al. 2017; Yi et al. 2017; Zhu et al. 2017). Due to the stochastic and high-dimensionality nature of many physical processes of the Earth system, GANs and conditional GANs are particularly appealing for atmospheric science problems. Recently, they have been used for various Earth-science related applications: for instance for statistical downscaling (Leinonen et al. 2020; Wang et al. 2021), temporal disaggregation of spatial rainfall fields (Scher and Peßenteiner 2020), sampling of extreme values (Bhatia et al. 2020), modelling of chaotic dynamical systems (e.g., Xie et al. 2018; Wu et al. 2020), classification of snowflake images (Leinonen and Berne 2020), weather forecasting (Bihlo 2020) and stochastic parameterization in geophysical models (Gagne II et al. 2020).

In climate modelling context, no one-to-one correspondence exists between observations and model simulations as they have different internal variabilities and thus are not synchronized in time. Biases refer to differences in distributional properties between references and simulated climate variables. Hence, in this context, bias correction can be seen as an unsupervised image-to-image problem that aims to map daily images from model simulations to daily images from historical observational references in order to adjust the distributional properties of the climate model.

In this study, we adapt a specific formulation of conditional GANs, initially used for unsupervised image-to-image translation problems (CycleGAN, Zhu et al. 2017), for multi-site corrections of climate simulations. The new MBC method, referred to as MBC-CycleGAN in the following, is introduced and applied in a proof-of-concept context for the correction of daily temperature and precipitation fields with a simple neural network architecture. In order to investigate and evaluate the proposed methodology, applications and comparisons of MBC-CycleGAN based on PP (corresponding to a supervised context) and MOS (unsupervised context) approaches are performed through a cross-validation method. In addition, a second cross-validation method is used in this study to assess the performances of MBC-CycleGAN in a context of different degrees of nonstationarity of the climate model between present (i.e., calibration) and future (i.e., projection) periods. One univariate quantile-mapping-based BC method and two MBC algorithms are included in the study in order to gain a better understanding of the performances of MBC-CycleGAN concerning univariate, spatial and temporal properties.

The paper is organized as follows: Section 2 presents the model and reference data used, and Sect. 3 describes the MBC-CycleGAN algorithm. Then, Sect. 4 displays the experimental setup used in this study, and results are provided in Sect. 5. Conclusions, discussions and perspectives for future research are finally proposed in Sect. 6.

2 Reference and model data

In this study, the dataset employed as reference for the bias correction is the "Système d'Analyse Fournissant des Renseignements Atmosphèriques á la Neige" (SAFRAN) reanalysis (Vidal et al. 2010) with an approximate 8 km × 8

km spatial resolution. Daily temperature and precipitation time series from 1 January 1979 to 31 December 2016 are extracted over the region of Paris, France ([47.878, 49.830° N] × [0.949,3.947° E]), which corresponds to a domain with $28 \times 28 = 784$ continental grid cells.

For the climate simulations data to be corrected, daily temperature and precipitation time series are taken from runs of the IPSL-CM5A-MR Earth system model (Marti et al. 2010; Dufresne et al. 2013) with a $1.25° \times 2.5°$ spatial resolution over the same region of Paris. For the 1979–2005 period, a historical run is extracted and concatenated with a run under RCP 8.5 scenario (i.e., the scenario with highest CO_2 concentration) for the 2006–2016 period, to obtain the desired 1979–2016 period. To perform a bias correction, a one-to-one correspondence between model and reference grid cells is needed, i.e., spatial resolutions between reference and model data have to be the same. Hence, IPSL data are regridded to the SAFRAN spatial resolution with a bilinear interpolation for both temperature and precipitation.

More data are required for this study, in particular for the implementation of the PP approach and to assess the influence of nonstationary properties of climate simulations on the performance of the proposed MBC method. For sake of clarity and make reading easier, these data will be introduced thereafter in the appropriate sections.

For illustration purpose, Fig. 1a displays the topographic map of France with the region of Paris in a box, as well as the mean daily temperature (Fig. 1b, c) and precipitation (Fig. 1d, e) maps for SAFRAN and IPSL datasets during winter over the 1979–2016 period for Paris.

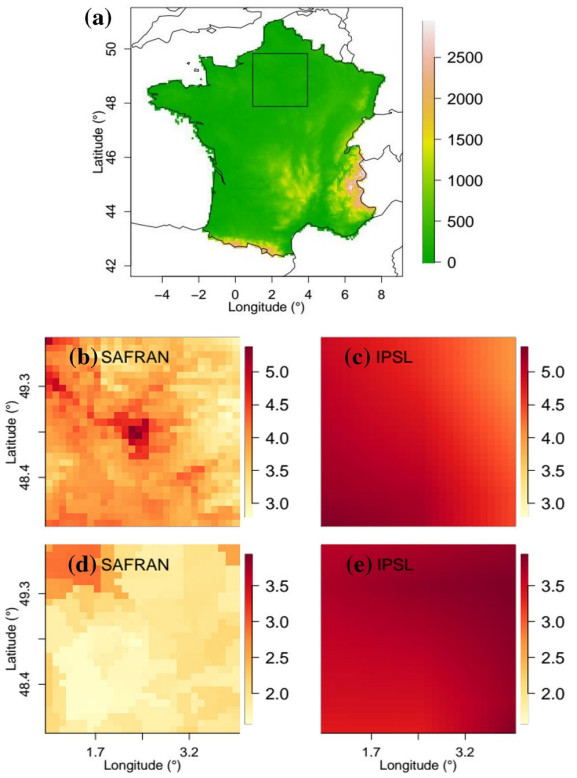

Fig. 1 **a** Topographic map of France with the selected region over Paris in a box, **b**, **c** temperature and **d**–**e** precipitation daily mean computed at each grid cell during winter over the 1979–2016 period for Paris. Results are shown for SAFRAN reference and plain IPSL outputs

3 Methodology

3.1 GAN

In its most basic formulation, a generative adversarial network consists of two neural networks that are trained conjointly: a generator and a discriminator. We first consider one random variable $\mathbf{Y}$ living in $\mathbb{R}^d$, with a probability distribution denoted $\mathbb{P}_{\mathbf{Y}}$. This random variable characterizes the available data, such as images of the target dataset (i.e., references), and hence takes its values in a high-dimensional space. We assume to have at hand samples $\mathbf{y}_1, \ldots, \mathbf{y}_n$ drawn according to the density $\mathbb{P}_{\mathbf{Y}}$ on $\mathbb{R}^d$. The generator, denoted G, is a function from $\mathbb{R}^{d'}$ to $\mathbb{R}^d$ and is intended to be applied to a d'-dimensional random variable $\mathbf{W}$, usually multivariate Gaussian random noise (with $d' \ll d$), such that the random variable $G(\mathbf{W})$ follows the law of $\mathbf{Y}$, i.e. $\mathbb{P}_{\mathbf{Y}} = \mathbb{P}_{\mathbf{G}(\mathbf{W})}$. Let $\mathbf{w}_1, \ldots, \mathbf{w}_n$ be a sample drawn from the distribution of $\mathbf{W}$. To train the generator G, the discriminator $D_{\mathbf{Y}}$, that is a function from $\mathbb{R}^d$ to [0, 1], is used as complex loss function (Goodfellow et al. 2014). This neural network is a binary classifier that returns the probability that a given observation, or image, comes from $\mathbb{P}_{\mathbf{Y}}$. The discriminator is trained in a supervised way to return maximal probability values on the reference images $\mathbf{y}_i$ and minimal values on the artificially generated images $G(\mathbf{w}_i)$. Conversely, the goal of the generator is to "fool" the discriminator by making the distribution of $G(\mathbf{w}_i)$ as indistinguishable as possible from that of $\mathbf{y}_i$, i.e., making difficult for the discriminator to determine that a sample $G(\mathbf{w}_i)$ comes from a distribution different from $\mathbb{P}_{\mathbf{Y}}$. Generator and discriminator are trained in turns and are in competition (i.e. "adversarial training") to improve themselves until it reaches an optimal equilibrium state.

The original formulation of GANs explained above is unconditional: the generator G only takes as input noise vectors $\mathbf{w}_i$ to produce new samples that are drawn from the target distribution $\mathbb{P}_{\mathbf{Y}}$. The idea of conditional GANs (e.g., Goodfellow et al. 2014; Mirza and Osindero 2014) is to add some information as inputs to direct the generation. By conditioning the generation on an input image, the generator is able to generate a corresponding output image, rendering the

conditional GANs appropriate for image-to-image translation tasks (e.g., Isola et al. 2017).

3.2 CycleGAN for unsupervised image-to-image translation

CycleGAN (Zhu et al. 2017) is a particular image-conditional GANs that is commonly used for unsupervised image-to-image translation. In the original application, CycleGAN has been applied with great success to transform photographs into the styles of master paintings by modifying colour information (i.e., RGB colour channels and/or spatial features of colours) of the photographs. Instead of the random noise $\mathbf{W}$, we introduce another random variable $\mathbf{X}$, with probability distribution $\mathbb{P}_{\mathbf{X}}$, living in the same dimensional space as $\mathbf{Y}$ (i.e., $\mathbb{R}^d$). This random variable $\mathbf{X}$ characterizes the images of the source dataset (i.e., biased simulations to correct). The CycleGAN approach consists in learning a mapping (i.e., a generator) $G_{\mathbf{X}\to\mathbf{Y}} : \mathbb{R}^d \to \mathbb{R}^d$ such that the random variable $G_{\mathbf{X}\to\mathbf{Y}}(\mathbf{X})$ follows the law of $\mathbf{Y}$ (i.e., $\mathbb{P}_{\mathbf{Y}} = \mathbb{P}_{G_{\mathbf{X}\to\mathbf{Y}}(\mathbf{X})}$). In addition to samples $\mathbf{y}_1, \dots, \mathbf{y}_n$, we assume to have at hand image samples $\mathbf{x}_1, \dots, \mathbf{x}_n$ drawn according to density $\mathbb{P}_{\mathbf{X}}$ on $\mathbb{R}^d$. Similarly as unconditional GANs, the mapping $G_{\mathbf{X}\to\mathbf{Y}}$ is learned using an adversarial loss, i.e. with a discriminator $D_{\mathbf{Y}}$ which forces the generator $G_{\mathbf{X}\to\mathbf{Y}}$ to generate images from a distribution close to the target distribution $\mathbb{P}_{\mathbf{Y}}$. The adversarial loss is defined as:

$$L_{GAN}(G_{\mathbf{X}\to\mathbf{Y}}, D_{\mathbf{Y}}) = \frac{1}{n}\sum_{i=1}^{n} \ln D_{\mathbf{Y}}(\mathbf{y}_i) + \frac{1}{n}\sum_{i=1}^{n} \ln\big(1 - D_{\mathbf{Y}} \circ G_{\mathbf{X}\to\mathbf{Y}}(\mathbf{x}_i)\big). \tag{1}$$

$G_{\mathbf{X}\to\mathbf{Y}}$ aims to minimize this adversarial objective against $D_{\mathbf{Y}}$, that means, tries to fool the discriminator with its generated images (i.e., maximizing the probability $D_{\mathbf{Y}}(G_{\mathbf{X}\to\mathbf{Y}}(\mathbf{x}_i))$). On the contrary, the discriminator $D_{\mathbf{Y}}$ aims to maximize the adversarial loss by distinguishing between transferred samples $G_{\mathbf{X}\to\mathbf{Y}}(\mathbf{x}_i)$ and samples $\mathbf{y}_i$ from the distribution $\mathbb{P}_{\mathbf{Y}}$. A perfect discriminator $D_{\mathbf{Y}}$ would return probability values equal to 1 for samples drawn from $\mathbb{P}_{\mathbf{Y}}$ and equal to 0 for samples generated by $G_{\mathbf{X}\to\mathbf{Y}}$. Hence, $G_{\mathbf{X}\to\mathbf{Y}}$ is designed to solve the optimization problem against $D_{\mathbf{Y}}$:

$$G_{\mathbf{X}\to\mathbf{Y}} = \arg\min_{G_{\mathbf{X}\to\mathbf{Y}}} \max_{D_{\mathbf{Y}}} L_{GAN}\big(G_{\mathbf{X}\to\mathbf{Y}}, D_{\mathbf{Y}}\big). \tag{2}$$

As highlighted by Zhu et al. (2017), this adversarial objective for unsupervised problems is under-constrained: there is no guarantee that "an individual input $\mathbf{x}_i$ and output $\mathbf{y}_i$ are paired up in a meaningful way" with such a mapping $G_{\mathbf{X}\to\mathbf{Y}}$. In fact, without further constraints, several different mappings can optimize similarly the adversarial loss by transferring the same set of images from $\mathbb{P}_{\mathbf{X}}$ to any random permutation of a same set of images from the distribution $\mathbb{P}_{\mathbf{Y}}$. Moreover, optimizing in practice this under-constrained adversarial objective alone has been found to be difficult for unsupervised problems, often leading to a well-known problem called "mode collapse". Mode collapse appears when a generator fails to model the complete range of input images. This results in a lack of diversity in the generated outputs. To address these issues, Zhu et al. (2017) propose to reduce the number of possible mapping functions by adding more constraints to the optimization problem. To do so, they introduce the inverse mapping $G_{\mathbf{Y}\to\mathbf{X}} : \mathbb{R}^d \to \mathbb{R}^d$, as well as a second discriminator $D_{\mathbf{X}}$ aimed to recognize images from the distribution $\mathbb{P}_{\mathbf{X}}$. Similarly to the mapping $G_{\mathbf{X}\to\mathbf{Y}}$, an equivalent adversarial loss can be used to learn the mapping $G_{\mathbf{Y}\to\mathbf{X}}$ by solving $\arg\min_{G_{\mathbf{Y}\to\mathbf{X}}} \max_{D_{\mathbf{X}}} L_{GAN}(G_{\mathbf{Y}\to\mathbf{X}}, D_{\mathbf{X}})$. Zhu et al. (2017) proposed to use $G_{\mathbf{Y}\to\mathbf{X}}$ to enforce the learned mappings to be cycle-consistent. That means that, for each input image $\mathbf{x}_i$, the mappings $G_{\mathbf{X}\to\mathbf{Y}}$ and $G_{\mathbf{Y}\to\mathbf{X}}$ can be constrained such that it learns to translate $\mathbf{x}_i$ back to the initial image, i.e. $G_{\mathbf{Y}\to\mathbf{X}} \circ G_{\mathbf{X}\to\mathbf{Y}}(\mathbf{x}_i) \approx \mathbf{x}_i$ (and similarly for image $\mathbf{y}_i$, such that $G_{\mathbf{X}\to\mathbf{Y}} \circ G_{\mathbf{Y}\to\mathbf{X}}(\mathbf{y}_i) \approx \mathbf{y}_i$). This property can be enforced by using a "cycle-consistency" loss which is defined as:

$$\begin{aligned} L_{cyc}\big(G_{\mathbf{X}\to\mathbf{Y}}, G_{\mathbf{Y}\to\mathbf{X}}\big) =& \frac{1}{n}\sum_{i=1}^{n} \big|\big|G_{\mathbf{Y}\to\mathbf{X}}(G_{\mathbf{X}\to\mathbf{Y}}(\mathbf{x}_i)) - \mathbf{x}_i\big|\big|_1 \\ &+ \frac{1}{n}\sum_{i=1}^{n} \big|\big|G_{\mathbf{X}\to\mathbf{Y}}(G_{\mathbf{Y}\to\mathbf{X}}(\mathbf{y}_i)) - \mathbf{y}_i\big|\big|_1. \end{aligned} \tag{3}$$

Finally, to ensure that images in $\mathbf{x}_1, \dots, \mathbf{x}_n$ that already seem to be draw from the distribution $\mathbb{P}_{\mathbf{Y}}$ (and vice-versa) are not mapped to another images, an identity mapping loss can also be defined as:

$$\begin{aligned} L_{id}\big(G_{\mathbf{X}\to\mathbf{Y}}, G_{\mathbf{Y}\to\mathbf{X}}\big) =& \frac{1}{n}\sum_{i=1}^{n} \big|\big|G_{\mathbf{Y}\to\mathbf{X}}(\mathbf{x}_i) - \mathbf{x}_i\big|\big|_1 \\ &+ \frac{1}{n}\sum_{i=1}^{n} \big|\big|G_{\mathbf{X}\to\mathbf{Y}}(\mathbf{y}_i) - \mathbf{y}_i\big|\big|_1, \end{aligned} \tag{4}$$

which further reduces the solution space of mapping functions and prevents even more the optimization problem from being under-constrained. The full objective function of the CycleGAN architecture can be expressed as follows:

$$\begin{aligned} L\big(G_{\mathbf{X}\to\mathbf{Y}}, G_{\mathbf{Y}\to\mathbf{X}}, D_{\mathbf{X}}, D_{\mathbf{Y}}\big) =& L_{GAN}\big(G_{\mathbf{X}\to\mathbf{Y}}, D_{\mathbf{Y}}\big) + L_{GAN}\big(G_{\mathbf{Y}\to\mathbf{X}}, D_{\mathbf{X}}\big) \\ &+ \lambda_{cyc} L_{cyc}(G_{\mathbf{X}\to\mathbf{Y}}, G_{\mathbf{Y}\to\mathbf{X}}) \\ &+ \lambda_{id} L_{id}\big(G_{\mathbf{X}\to\mathbf{Y}}, G_{\mathbf{Y}\to\mathbf{X}}\big), \end{aligned} \tag{5}$$

where λ_{cyc} and λ_{id} control the relative importance of both cycle-consistency and identity losses. Finally, the CycleGAN aims to solve:

$$\big(G_{\mathbf{X}\to\mathbf{Y}}, G_{\mathbf{Y}\to\mathbf{X}}\big) = \arg\min_{G_{\mathbf{X}\to\mathbf{Y}}, G_{\mathbf{Y}\to\mathbf{X}}} \max_{D_{\mathbf{X}}, D_{\mathbf{Y}}} L\big(G_{\mathbf{X}\to\mathbf{Y}}, G_{\mathbf{Y}\to\mathbf{X}}, D_{\mathbf{X}}, D_{\mathbf{Y}}\big). \tag{6}$$

Although estimating the inverse mapping $G_{\mathbf{Y}\to\mathbf{X}}$ is not necessarily the initial goal of many image-to-image translation problems, its use to constrain the optimization problem has been found to be crucial in an unsupervised context for the convergence of the algorithm and the estimation of the desired mapping $G_{\mathbf{X}\to\mathbf{Y}}$. Illustrations of the adversarial, cycle-consistent and identity losses within the CycleGAN architecture are given in Fig. 2.

3.3 The MBC-CycleGAN approach

3.3.1 Adaptation of CycleGAN for MBC

The main idea of the proposed methodology, named MBC-CycleGAN, is to adapt the CycleGAN approach so that it turns daily maps of a simulated variable with spatial features inappropriate compared to a reference dataset, to more realistic maps. Here, MBC-CycleGAN is developed in the context of the "marginal/dependence" MBC category, i.e., correcting separately marginal distributions and dependence relationships. In addition to marginal distributions, we consider the adjustment of spatial dependence structures. The algorithm is trained on a historical period (i.e., calibration) for which both climate simulations and reference datasets are available. Once the adversarial neural network has converged, adjustment of climate simulations over a projection period (e.g., a future time period) is performed using the pretrained algorithm. The MBC-CycleGAN proceeds as follows:

1. As MBC-CycleGAN belongs to the marginal/dependence category, univariate distributions of modeled climate variables are first corrected independently using a univariate BC method for both calibration and projection periods. In this study, the quantile-quantile (QQ) mapping method is used (Déqué 2007).
2. Then, quantile-quantile and reference data over the calibration time period are transformed to belong to [0, 1] using a pointwise min-max normalization. For each grid cell, the minimum and maximum values from the refer-

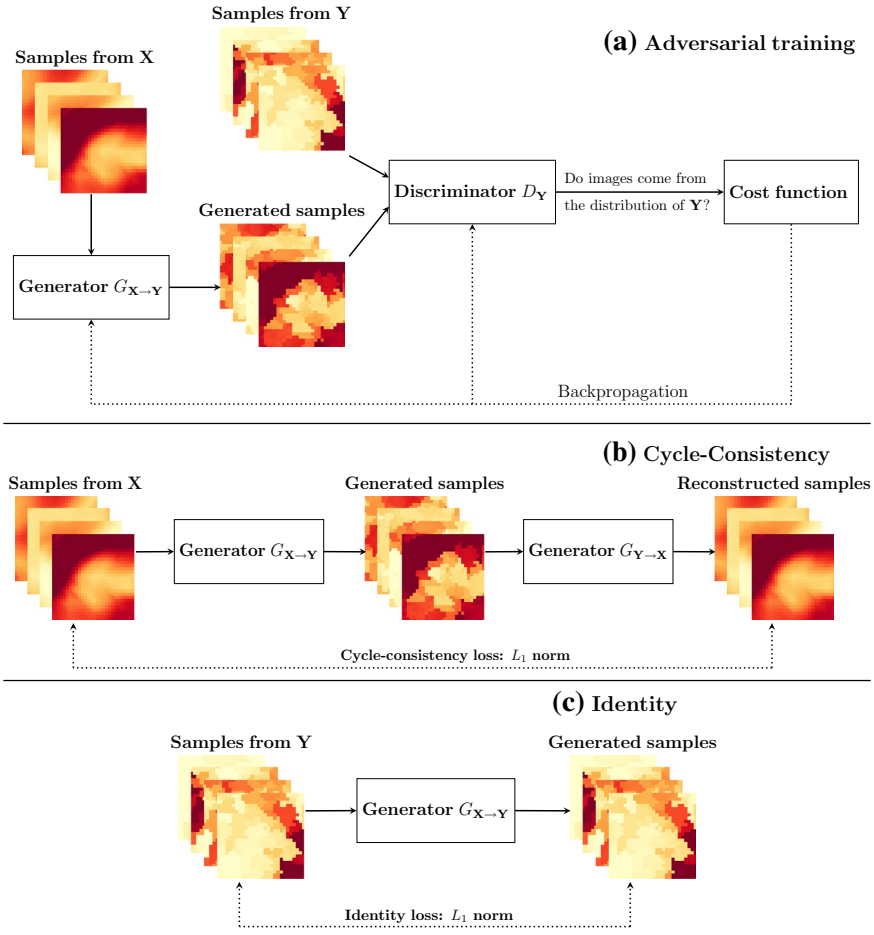

Fig. 2 **a** Illustration of the adversarial training for the mapping function $G_{\mathbf{X}\to\mathbf{Y}}$, associated with the adversarial discriminator $D_{\mathbf{Y}}$. $D_{\mathbf{Y}}$ encourages $G_{\mathbf{X}\to\mathbf{Y}}$ to generate outputs that are indistinguishable from the probability distribution of $\mathbf{Y}$. A similar adversarial training is used for $G_{\mathbf{Y}\to\mathbf{X}}$ using $D_{\mathbf{X}}$ (not presented in this figure). In CycleGAN architectures, the mappings $G_{\mathbf{X}\to\mathbf{Y}}$ and $G_{\mathbf{Y}\to\mathbf{X}}$ are enforced to be cycle-consistent, i.e., **b** if an initial image from $\mathbf{X}$ is translated using $G_{\mathbf{X}\to\mathbf{Y}}$ and back again using $G_{\mathbf{Y}\to\mathbf{X}}$, the initial image should be obtained. **c** In addition, to ensure that images from $\mathbf{X}$ that already seem to be drawn from the distribution of $\mathbf{Y}$ are not modified too much, the identity property is used by enforcing $G_{\mathbf{X}\to\mathbf{Y}}$ applied to images from $\mathbf{Y}$ to resemble to initial inputs from $\mathbf{Y}$ (and vice versa for $G_{\mathbf{Y}\to\mathbf{X}}$). In our study, samples from $\mathbf{X}$ and $\mathbf{Y}$ are replaced by QQ outputs and references, respectively

ence during the calibration are taken to compute the normalization. The resulting daily maps are then given to a CycleGAN model to learn the transfer between the two distributions of images. Generators and discriminators are trained until the spatial distribution of the corrected maps stops improving. More details about the criteria used to evaluate spatial distributions are presented thereafter.

3. Once the CycleGAN model has been trained for the calibration period, the same pointwise normalization is performed for quantile-quantile data over the projection period, i.e., using the same minimum and maximum values from the reference during the calibration period. Normalized daily maps from quantile-quantile data in the projection period are translated in the normalized reference domain using the pretrained adversarial neural network. Then, the corrected outputs obtained are rescaled to physical values by applying the inverse of the pointwise min-max normalization used.
4. Finally, by taking advantage of the Schaake Shuffle technique (Clark et al. 2004), quantile-quantile data for the projection period obtained from Step 1 are reordered such that the rank structure of the data obtained from Step 3 is reproduced. This shuffling technique, already employed in a few multivariate bias correction methods (e.g., Vrac 2018; Cannon 2018; Mehrotra and Sharma 2019), permits here to obtain bias-corrected data with marginal properties from quantile-quantile outputs and rank dependence structure from CycleGAN outputs.

A summary of the successive steps in the form of a flowchart is provided in Fig. 3. More details about the different algorithmic steps are presented in Appendix 1.

3.3.2 Network architecture

To infer the weights for the cycle-consistency mapping loss λ_{cyc} and the identity mapping loss λ_{id}, preliminary tests have been conducted by checking a couple of combinations of weights and verifying that our optimization process improved the spatial structure of the climate simulations. With respect to these results (not shown), the weights have been chosen equal to $\lambda_{cyc} = 10$ and $\lambda_{id} = 1$.

Additionally, in this paper, we only present results obtained with a simple architecture for the CycleGAN neural networks. Our work being a proof of concept, we did not tune any further the architecture or the hyperparameters of the neural networks. However, the results presented later in Sect. 5 appear sufficient to illustrate the potential of CycleGANs for MBC. Schemes for the convolutional neural networks for both generators and discriminators are presented in Fig. 4. Architecture of generators for the mapping and inverse mapping are identical and are based on deep convolutional layers (DCGAN, Radford et al. 2016). First, the daily maps, i.e. images of size 28×28 are given as inputs to the generators. Then, images flow through three 2D convolution layers with an increasing number of 3×3 filters (64–128–256). Two of them are performing convolutions that downsample input images to capture complex patterns at different scales. Then, two 2D transpose convolutional layers with a decreasing number of 4×4 filters (128–64) are used to perform inverse convolution operations and upsampling input data. Finally, one 2D convolution layer with one 1×1 filter is used to generate an output image of the same size as the initial one. Skip connections between convolution and transpose convolutional layers are used to ease the training of the CycleGAN network (He et al. 2016). All the other hyperparameters for the neural network architecture of the generators are detailed in Appendix 2.

Concerning the discriminators, they take as well as inputs images of size 28×28. Then, two 2D convolution layers with an increasing number of 3×3 filters (64–128) are used. Finally, outputs are flattened, i.e., are converted into a 1-dimensional array before being given to a fully connected layer (dense layer) that computes the sigmoïd values (i.e., probabilities) for the classification of images.

The number of parameters is equal to 1,025,281 for each generator and 80,769 for each discriminator, bringing the total number of parameters to 2,212,100 for the whole CycleGAN architecture. Please note that each convolution and transpose convolutional layer used within the neural network architectures of both generators and discriminators includes a bias vector to fit. The number of parameters added by individual convolutional layers depends on its number of filters f_2, the filter size (here 3×3) and also the number of filters f_1 from the previous convolutional layer. Adding an additional convolutional layer in a generator architecture with f_2 filters will add $(3 \times 3 \times f_1 + 1) \times f_2$ parameters. Hence, constructing a (deeper) neural network with more and more layers increases drastically the number of parameters to train. In order to keep an algorithm which is relatively fast to train while being stable, we decided not to add further layers to generators and discriminators architectures. For a concise summary of network architectures used, we refer to the Tables 3 and 4 in Appendix 2.

3.3.3 Training details

In this study, CycleGAN networks are trained using the Adam optimizer (Kingma and Ba 2017) with learning rates of 1e−4 and 5e−5 for the generators and discriminators, respectively. Please note that no grid search has been performed to determine optimal values of learning rates, and hence there is room for improvement. For the performance assessment of the CycleGAN model during training, the energy distance (Székely and Rizzo 2004; Székely and

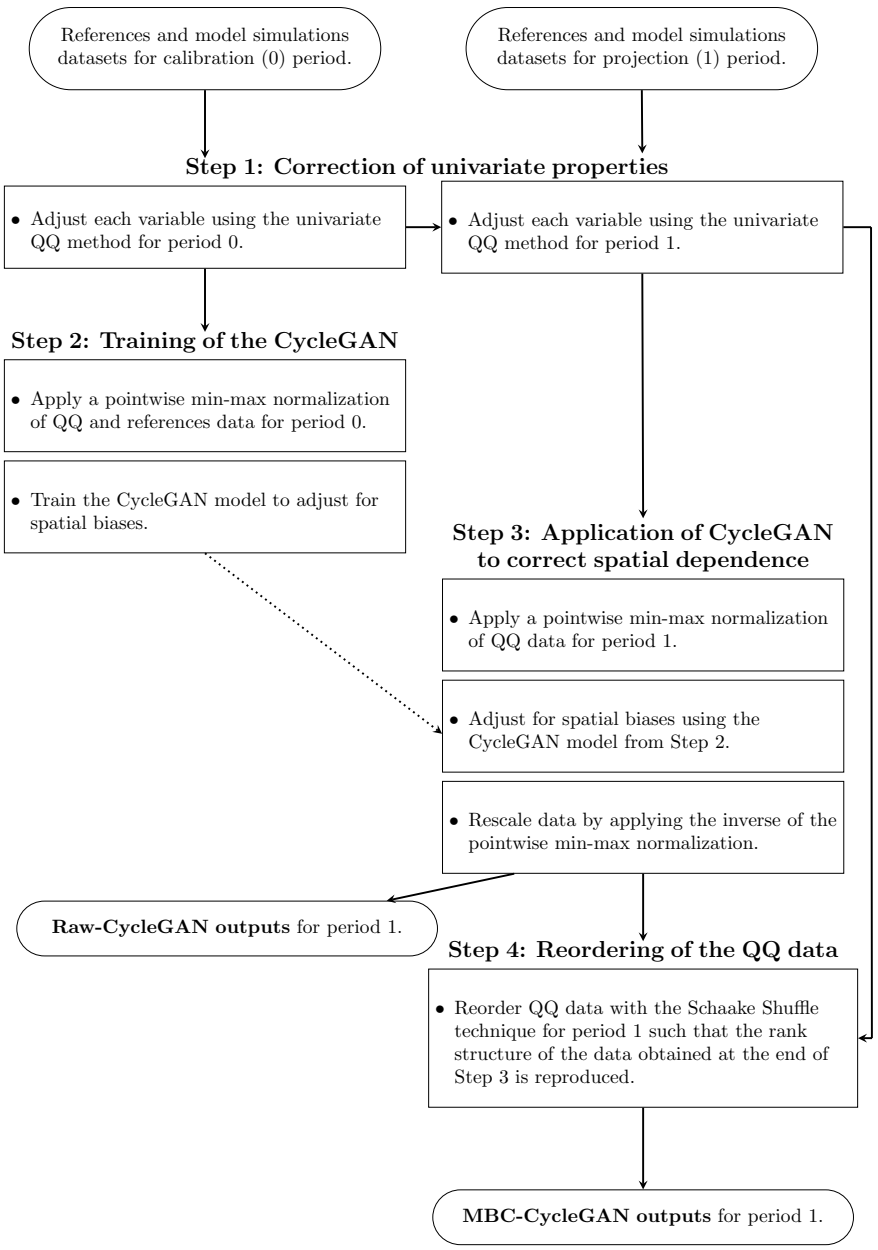

Fig. 3 Flowchart for the MBC-CycleGAN method to adjust climate simulations for the projection period

Rizzo 2013) is used. This metric, already used in the bias correction literature (e.g., Cannon 2018), permits to measure the statistical discrepancy between two multivariate distributions that are potentially in high dimension. Given two k-multivariate independent random vectors $\mathbf{P}$ and $\mathbf{Q}$ with multivariate probability distributions μ and υ respectively, the energy distance $\mathcal{E}$ between the two distributions is:

$$\mathcal{E}(\mu,\upsilon) = \sqrt{2\mathrm{E}\|\mathbf{P}-\mathbf{Q}\| - \mathrm{E}\|\mathbf{P}-\mathbf{P}'\| - \mathrm{E}\|\mathbf{Q}-\mathbf{Q}'\|},$$

with E denoting the expected value, $\mathbf{P}'$ (resp. $\mathbf{Q}'$) independent and identically distributed copy of $\mathbf{P}$ (resp. $\mathbf{Q}$) and $\|.\|$ the Euclidean distance. The corresponding energy statistic of $\mathcal{E}$ between two k- dimensional statistical samples $\mathbf{p}$ and $\mathbf{q}$ can be computed as follows:

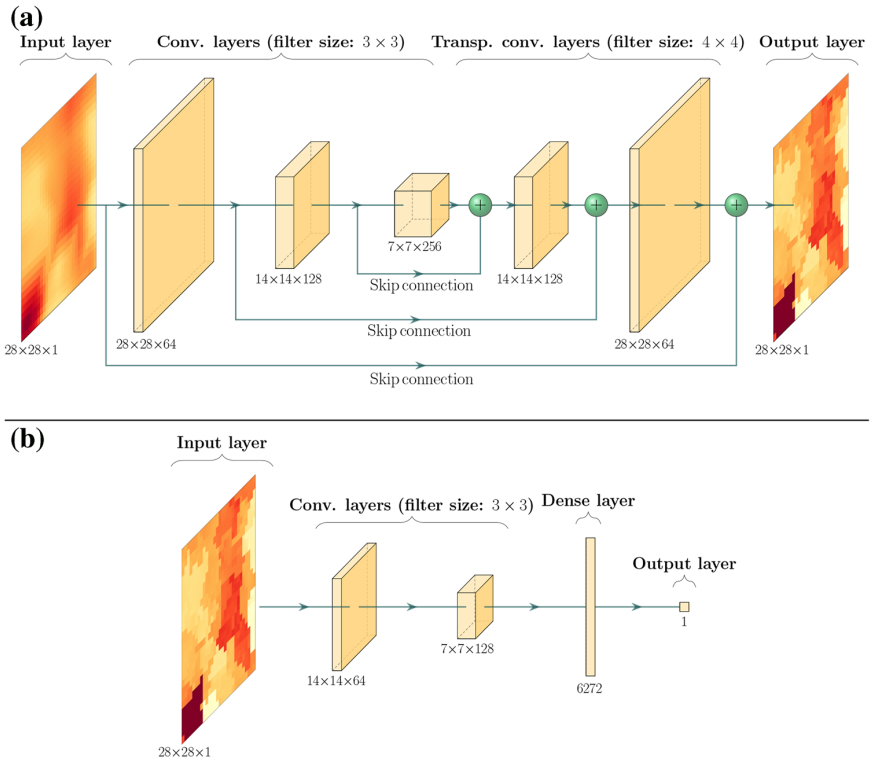

Fig. 4 Scheme of the convolutional neural networks for the **a** generators and **b** discriminators used in this study within the MBC-CycleGAN procedure. For each convolutional and transpose convolutional layers, the number of filters used is indicated by the third coordinate of their output size

$$\widehat{\mathcal{E}}(\mathbf{p},\mathbf{q}) = \left(\frac{2}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{m=1}^{n_2} \|\mathbf{p}_i - \mathbf{q}_m\| - \frac{1}{n_1^2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_1} \|\mathbf{p}_i - \mathbf{p}_j\| - \frac{1}{n_2^2} \sum_{l=1}^{n_2} \sum_{m=1}^{n_2} \|\mathbf{q}_l - \mathbf{q}_m\| \right)^{\frac{1}{2}},$$

where $\mathbf{p}_i$ denotes the realizations of $\mathbf{P}$ at the time step i across the k dimensions (and similarly for $\mathbf{q}_m$ with $\mathbf{Q}$). The energy statistic goes to zero when the two multivariate samples $\mathbf{p}$ and $\mathbf{q}$ are drawn from the same distribution.

During training, computations of energy distances are performed every 10 epochs, i.e. each time that the CycleGAN has worked 10 times through the entire training dataset. Estimated energy distances $\widehat{\mathcal{E}}$ are calculated on multivariate distributions of ranks between references and bias-corrected data. It permits to assess along the training the performance of the method to correct the whole spatial dependence structure of climate simulations. Computing energy distance using ranks instead of raw values allows the removal of the influence of univariate properties on the spatial relationships. The CycleGAN model that minimizes the energy distance on ranks during training is chosen for the correction of the projection period. Training 1000 epochs takes ~ 4 h on a single NVIDIA Tesla V100 GPU.

4 Design of experiments

For evaluation purposes, the proposed MBC-CycleGAN method is applied to adjust climate simulations outputs with SAFRAN data as references. Bias correction is performed on separate seasons in order to preserve seasonal properties. In the following, for sake of clarity, only the winter results are presented. Data are available for the 1979–2016 period (i.e, 3420 winter days), and need to be divided into a calibration period and a projection period to train and evaluate our algorithm. In accordance with common practices in machine learning, the 1979–2016 period is split as follows: 70% (2394 days) as training dataset and 30% (1026 days) as evaluation dataset. In this study, two different cross-validation methods—that differ in how calibration and projection periods are constructed—are used to evaluate our methodology.

4.1 Model output statistics (MOS) vs. Perfect prog (PP)

The first cross-validation method consists in drawing randomly the days that define the calibration and projection periods. As these periods are drawn randomly, the

potential climate change signal present in the data during the 1979–2016 period vanishes. Hence, for this cross-validation method, no changes in marginal and dependence properties are expected between the calibration and projection periods, allowing for the assessment of the method in a stationary context. We take advantage of this first stationary cross-validation technique to apply our method in both PP and MOS post-processing schemes for the adjustment of IPSL climate simulations. Implementing and evaluating both the PP and MOS approaches in such a validation context permits to determine which approach is better suited in our context of bias correction of climate simulations. For the MOS approach, MBC-CycleGAN is applied directly to IPSL data according to the 4 steps already described in Sect. 3.3. Concerning the implementation of the PP approach, the same procedure is applied but the CycleGAN model is trained in a slightly different way. Indeed, as already explained in Sect. 1, a PP approach consists in establishing the statistical relationships between large–cale predictors and local-scale predictands from observational or reanalysis data (including for the predictors) before applying them to climate model data. Hence, large-scale predictors temporally matching the SAFRAN dataset are needed to a PP approach. For this purpose, a new climate dataset is constructed for both temperature and precipitation as follows: initial local-scale SAFRAN data with 8 km × 8 km spatial resolution are upscaled using conservative interpolation on a large-scale grid of 32 km × 32 km spatial resolution. Then, the obtained large-scale data are regridded using bilinear interpolation to the initial grid of SAFRAN, allowing to train CycleGAN. It results in "biased" daily maps of temperature and precipitation (large-scale predictors) of the initial SAFRAN data (local-scale predictands), temporally matching the chronology of the SAFRAN time series. Using these new data—hereafter referred to as "low-resolution (LR) SAFRAN"—a CycleGAN model is trained for the implementation of the PP approach by learning the transfer of maps from 1d-BC large-scale predictors (QQ(LR SAFRAN)) to maps from local-scale predictands (SAFRAN). This trained model is then used to bias correct IPSL simulations over the projection period and, hence, evaluate the CycleGAN results in a PP context.

4.2 Nonstationarity investigation

To evaluate the nonstationary behavior of the proposed method, a second cross-validation method is defined, which consists in dividing the 1979–2016 period chronologically. By still defining the calibration and the projection periods based on the 70–30% split, it results in obtaining approximately the 1979–2005 and 2006–2016 portions as calibration and projection periods, respectively. Hence, the potential climate change signal between the calibration and projection periods is not removed by the cross-validation technique. Within this second cross-validation method, IPSL simulations and SAFRAN references can potentially have different marginal and spatial dependence changes between calibration and projection periods. In this respect, depending on the level of agreement in changes between simulations and references, and how MBC methods account for these changes in their correction procedure, the quality of the correction for projection periods can possibly be different. Hence, to provide a global picture of the performances of the MBC-CycleGAN method in the nonstationary context, three bias correction exercises of climate data with different statistical changes are performed with respect to SAFRAN references:

- the correction of IPSL simulations that present different marginal and spatial properties from SAFRAN, and with potentially different changes than those from SAFRAN.
- the correction of LR SAFRAN dataset (presented above), whose marginal and spatial properties as well as their changes are in line with those from SAFRAN.
- the correction of a third dataset called IPSLbis (presented below) that presents different marginal and spatial properties from SAFRAN, but for which their changes are in line with those from SAFRAN.

For the sake of clarity, a summary of the different attributes of the three datasets to correct is presented in Table 1.

LR SAFRAN dataset already presented above has, by construction, little bias with SAFRAN references: its biases are only due to the interpolation technique used to obtain data with a lower resolution. Hence, statistical changes between the calibration and projection periods for LR SAFRAN are in line with those from the SAFRAN dataset. Adjusting LR SAFRAN data for the projection period permits to assess if the MBC-CycleGAN method is able to reproduce the changes from the reference in the correction. Also, the LR SAFRAN dataset presents the particularity of being synchronous in time with references. Hence, in addition to evaluate the proposed method in terms of distributional properties, which is not considered as sufficient to identify successful bias correction techniques (Maraun 2016), this pairwise correspondence between predictors and predictands offers the possibility to directly compare corrected daily maps with those from the references using classic forecast verification statistics.

As IPSL simulations compute a different combination of variability and warming than those from the SAFRAN reanalysis, IPSL model and SAFRAN references are likely to present disagreeing changes in their statistical (marginal and dependence) properties between calibration and projection periods. To evaluate the influence of these potential disagreeing changes on the performance of correction of the

proposed method, we constructed the third dataset, referred to as "IPSLbis", for the projection period only. IPSLbis is specifically constructed so that its marginal and dependence changes between calibration and projection periods are in line with those from the reference. In order to ease the comparison of results with the first bias correction exercise, we forced IPSLbis to have the same changes as LR SAFRAN. This is reached by using a two-step procedure that takes advantage of a nonstationary quantile mapping technique for marginal changes (CDF-t, Vrac et al. 2012) and a matrix-recorrelation technique for dependence changes (Bárdossy and Pegram 2012). More details about the generation of the IPSLbis data can be found in Appendix 3 and a detailed evaluation of the evolution of statistical properties of the different dataset between the calibration and projection period is provided in Appendix 4. In particular, results presented in Appendix 4 indicate that, as expected, changes in spatial structures from SAFRAN references are (globally) in agreement with those from LR SAFRAN for both temperature and precipitation. However, concerning changes in spatial structures for IPSL simulations, conclusions are not the same depending on the physical variable. While, for temperature, simulated changes of spatial correlations are partially in line with those from LR SAFRAN, IPSL model presents discrepancy of changes for precipitation. Globally, the construction of IPSLbis with the two-step procedure described in Appendix 3 permits to impose to IPSL data spatial changes for both temperature and precipitation that are in line with those from LR SAFRAN.

4.3 Comparisons to existing MBCs: R^2D^2 and dOTC

Although evaluating the performance of correction for IPSL simulations is of primary interest, applying our method on these three datasets (IPSL, IPSLbis, LR SAFRAN) permits to assess gradually how well our method is performing depending on the biases present in the dataset to correct. Note that, as IPSL and IPSLbis data during calibration are identical, there is no need to train for a second time the CycleGAN model for IPSLbis data: the CycleGAN model trained with IPSL data can be used directly to adjust IPSLbis simulations for the projection period. In addition, two MBCs with different assumptions about nonstationarity are applied for comparison using the second cross-validation method: the "Rank Resampling For Distributions And Dependences" (R^2D^2, Vrac and Thao 2020) and the "Dynamical Optimal Transport Correction" (dOTC, Robin et al. 2019) methods.

R^2D^2, developed in the context of marginal/dependence category, relies on an analogue-based method that allows to resample ranks from a reference dataset according to some conditioning information and reconstructs dependence structure of the simulated time series. The information to condition the analogues can be multivariate by considering, for example, a set of variables to be corrected at a given time *t*. Conditioning for the ranks resampling can also be extended to ranks sequences, i.e. conditioning by not only one but several lagged time steps. Please note that, for the different implementations of R^2D^2 in this study, the multivariate conditioning used includes 4 grid points that cover uniformly the region of interest. In addition, 5 lagged time steps are used for the conditioning, as it has been found to stabilize the R^2D^2 method (not shown). Also, the QQ method is used to correct the marginal properties for R^2D^2 outputs.

Concerning the dOTC method, it was developed in the all-in-one category, i.e., adjusting the univariate distributions and dependence structures at the same time. The dOTC method takes advantage of the optimal transport theory to construct a multivariate transfer function, named a transport plan, for the adjustment of climate simulations with respect to references while minimizing an associated cost function. This particular transfer function permits to link, through conditional laws, all the multivariate elements from the biased multivariate distribution to their corrections. Corrections are then derived by drawing directly from these conditional laws to obtain the bias corrected data.

Both R^2D^2 and dOTC methods are applied according to the spatial-dimensional configuration (hereinafter referred to as "Spatial-"), where all the 784 time series for a particular physical variable are corrected jointly. While R^2D^2 assumes spatial dependence structures (i.e., the rank correlations, or copulas) to be stable in time, the dOTC method makes the hypothesis of nonstationarity of the dependence structure between the calibration and the projection periods, which allows for taking into account the changes of the model (e.g., due to climate change) in the bias correction procedure. Intercomparing the results from both Spatial-R^2D^2 and Spatial-dOTC for adjusting spatial dependence structure of climate simulations with those from MBC-CycleGAN allows to better assess how the proposed method performs in a nonstationary context.

5 Results

In this section, analyses are presented for the winter season (December, January and February) only. CycleGAN models are trained during the calibration period and selected such that energy distances on ranks are minimized. All evaluations are performed on the projection period for the corrected outputs obtained from the two cross-validation methods and results are compared to those from the reference dataset. For bias-corrected precipitation time series, thresholding of 1 mm is applied before evaluation to replace values lower than 1 mm by 0. Bias correction outputs from the first and second cross-validation methods are evaluated in terms of both marginal and spatial properties. Analyses of

temporal properties are only provided for outputs from the second cross-validation method, in which calibration and projection periods are divided chronologically and hence do not distort temporal properties, contrary to the first cross-validation method that randomly defines these periods. To assess the potential benefits of considering spatial aspects in the correction procedure, the univariate QQ method (Déqué 2007) is also included in the study as a benchmark.

5.1 MOS vs. PP

5.1.1 Training of MBC-CycleGANs

Figure 5 shows energy distances with respect to SAFRAN references for temperature computed on physical values (Fig. 5a, b) and ranks (Fig. 5c, d) for LR SAFRAN, plain IPSL simulations, 1d-QQ, and MBC-CycleGAN (MBC-CG) outputs during the training on the calibration period. In addition, results for Raw-CycleGAN (Raw-CG) are presented. Differences between Raw-CG and MBC-CG only lie in their marginal properties: while Raw-CG corresponds to the outputs obtained from the CycleGAN after denormalization at the end of Step 3, MBC-CG is the combination of the spatial structure from Raw-CG and univariate properties from QQ outputs (see the flowchart provided in Fig. 3). The results for precipitation are presented in Fig. S1 of the Supplement.

Clearly, Fig. 5a, b show large energy distances computed on physical values of temperature for LR SAFRAN and IPSL datasets, indicating some biases on spatial structures for those dataset with respect to SAFRAN references. Adjusting marginal properties with the univariate QQ method reduces values of energy distance computed on physical values, highlighting the influence of marginal properties on spatial features. Correction of the spatial dependence structure provided by MBC-CG occurs relatively quickly, with energy distances on physical variables reduced by 2 compared to QQ after approximately 1000 epochs for both PP and MOS approaches. However, for Raw-CG, marginal properties generated by the inverse pointwise min-max normalization do not seem to improve values of energy distances, which justifies the post-processing of univariate properties adopted in the MBC-CycleGAN method with the Schaake Shuffle.

Figure 5c, d show that computing energy distances on ranks for temperature removes the influence of univariate

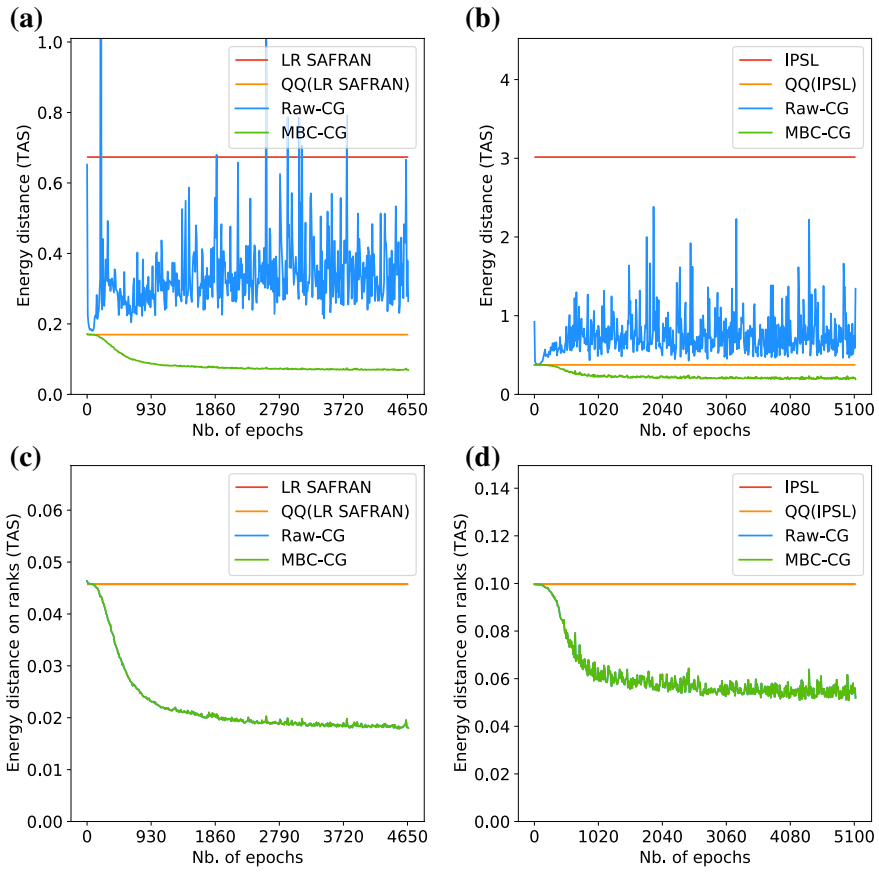

Fig. 5 Values of the energy distances with respect to SAFRAN reference for temperature computed on **a**, **b** physical values and **c**, **d** ranks during the training of MBC-CycleGAN. Results are shown for the different datasets involved in **a**, **c** the Perfect Prognosis approach and **b**, **d** the MOS approach. Please note that results of QQ and low-resolution SAFRAN (resp. IPSL) for ranks are the same. Red and orange lines are therefore superimposed in **c** (resp. **d**). This remark also applies for Raw-CycleGAN (blue line) and MBC-CycleGAN (green line)

properties on spatial features. Energy distances for both LR SAFRAN and IPSL with their respective QQ corrections are indeed the same (Fig. 5c). The same remark holds for MBC-CG and Raw-CG energy distances on ranks that have, by construction, similar spatial dependence structures. As explained in Sect. 3.3.3, the CycleGAN model that minimizes the energy distance on ranks of MBC-CycleGAN outputs is selected.

For precipitation (Fig. S1), the same conclusions hold, indicating a relative ability of the CycleGAN to adjust spatial dependence structure of precipitation fields. Nevertheless, contrary to temperature, one should remark that energy distances on ranks are different for LR SAFRAN, IPSL and their respective QQ corrections (Figs. S1c, d), which is specific to precipitation variables that can contain several null values for dry events. Indeed, ranks are computed here such that, when tied values are encountered, the minimum value of rank is attributed to each tied value. The combination of the correction with the QQ method and the thresholding for precipitation below 1 mm could modify the frequency of dry events, which could result in obtaining different rank structures, and hence, mechanically, different energy distances with respect to SAFRAN references. This mechanism is also obtained between MBC-CG and Raw-CG (Figs. S1c, d), that present different energy distances due to the difference of dry events.

5.1.2 Univariate distribution properties

Once the CycleGAN models have been selected for both the PP and MOS approaches, the corrections of IPSL simulations can be performed for the projection period. First, bias-corrected data are evaluated in terms of univariate statistics. For temperature and precipitation, differences of mean values between the bias corrected data and the SAFRAN references are computed at each grid cell. For temperature mean, absolute differences are computed, while for precipitation variables having absolute zeros, relative mean differences are more appropriate. Maps of differences with respect to the reference—for IPSL simulations and the bias-corrected data—are displayed in Fig. 6 for both temperature and precipitation. The mean absolute error (MAE) with respect to the reference dataset is also reported on each map. For more results on marginal properties, maps of standard deviation relative differences for both physical variables are also provided in Fig. S2 of the Supplement.

For both temperature and precipitation, the maps for the IPSL model (Fig. 6c, d) present large values of mean differences with respect to the SAFRAN map (Fig. 6a, b) and highlight the need to adjust univariate properties of simulations. Maps provided by 1d-QQ outputs (Fig. 6e, f) indicate that, as expected, the univariate method globally improves marginal properties at each individual site. In agreement

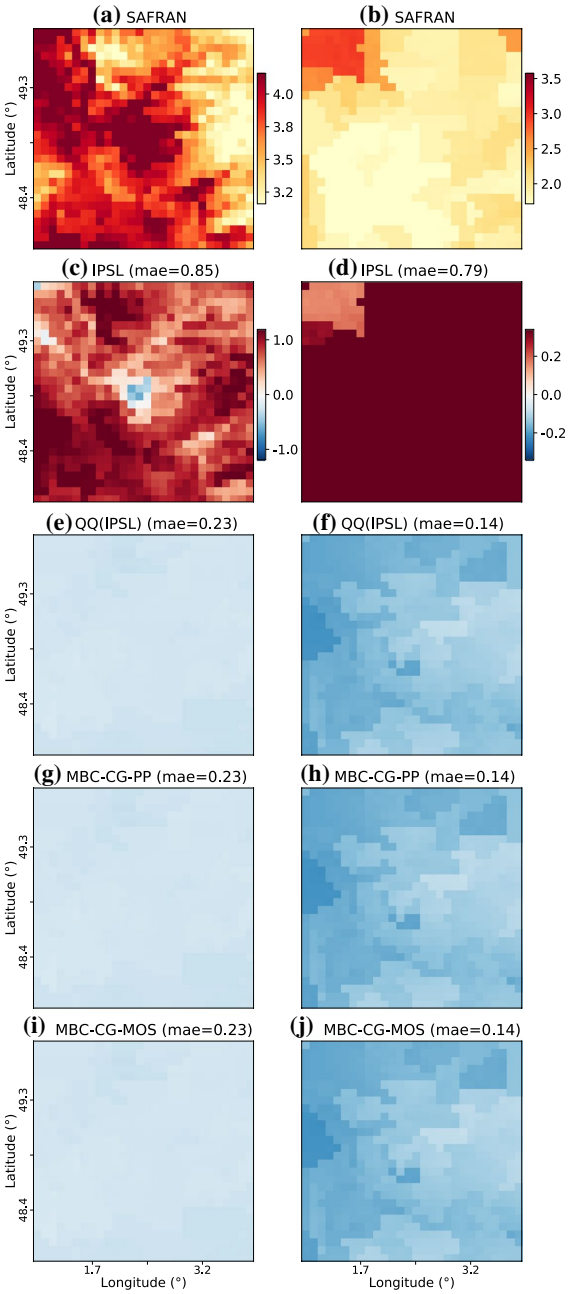

Fig. 6 Mean differences for **c**, **e**, **g**, **i** temperature and relative mean differences for **d**, **f**, **h**, **j** precipitation computed at each grid cell between SAFRAN reference and the different datasets (plain IPSL, QQ, MBC-CycleGAN-PP and MBC-CycleGAN-MOS outputs) during winter over the projection period. Note that the color scales between panels **c**, **e**, **g**, **i** and **d**, **f**, **h**, **j** are not the same to better emphasize intensities of values for the two physical variables. Maps of daily mean for SAFRAN references are also shown for **a** temperature and **b** precipitation

with the properties of the marginal/dependence MBC methods, maps for MBC-CG for PP (MBC-CG-PP, Fig. 6g, h) and MOS (MBC-CG-MOS, Fig. 6i, j) are exactly the same as those from the 1d-QQ method. Indeed, by construction, the univariate distribution properties are identical between QQ and MBC-CycleGAN outputs, regardless of the spatial correlation adjustments. Although MBC-CG-PP and MBC-CG-MOS do not use the same data for the training of the CycleGAN to adjust spatial features, same marginals are taken from the QQ outputs of IPSL data, which results in obtaining the same univariate properties between the three corrections.

5.1.3 Spatial correlations

Quality of the corrections in terms of spatial correlations is now assessed. For each grid cell, spatial dependencies are evaluated for temperature and precipitation by computing Pearson pairwise correlations between the cell of interest and each of the remaining 783 grid cells over the region of Paris for the different climate datasets. The biases of these 783 spatial Pearson correlations are then summarized by computing the Mean Squared Error (MSE) with the corresponding 783 correlations computed for the references. By computing the MSE values for each grid cell, 784 MSE values are obtained for each climate dataset and can be intercompared from one dataset to another. Figure 7 shows the boxplots of the MSE values obtained for both temperature and precipitation for the plain IPSL simulations and BC outputs. For both variables, the boxplots for the IPSL simulations indicate strong values of MSE with respect to SAFRAN references. For QQ outputs, only slight reductions of MSE of spatial correlations are observed compared to those from IPSL, indicating that QQ globally conserves the spatial structure of the IPSL model. This result could have been expected, as, for each site, the univariate QQ method does not modify (too much) rank sequences of the simulated time series. The slight improvement of spatial statistics, which is greater for precipitation (Fig. 7b) than temperature (Fig. 7a), is in fact mainly attributable to the correction of univariate properties provided by the QQ method. Concerning MBC-CycleGAN, the PP and MOS approaches display different performances in adjusting the spatial properties of simulations. Boxplots of MSE for MBC-CG-MOS indicate clear improvements of spatial correlations with respect to QQ outputs for both temperature and, to a lesser extent, precipitation. However, results for MBC-CG-PP show less pronounced improvements, suggesting a failure for the MBC-CG-PP approach to adjust spatial properties. This difference of performance for the PP approach indicates that, although CycleGAN models are able to learn the spatial relationships between large-scale predictors (LR SAFRAN) and local-scale predictands (SAFRAN) during the training of the algorithm, as previously shown in Figs. 5 and S1, these relationships do not prove to be suited for adjusting IPSL simulations. Indeed, simulated large-scale predictors seem here to present too large biases with respect to LR SAFRAN to make the CycleGAN fitted in a PP context applicable to the IPSL simulations. Hence, the perfect-prognosis approach should be discarded in our context of bias correction of climate simulations. Therefore, in the following, only the MOS approach of MBC-CG is further investigated.

5.2 MBC-CycleGAN in the nonstationary context

In the following, analyses are presented for the application of the MBC-CycleGAN method with the MOS approach in a nonstationary context using the second cross-validation method. Results for the correction of the three datasets - IPSL, IPSLbis and LR SAFRAN - with different changes in marginal and dependence properties between the calibration and projection periods are provided.

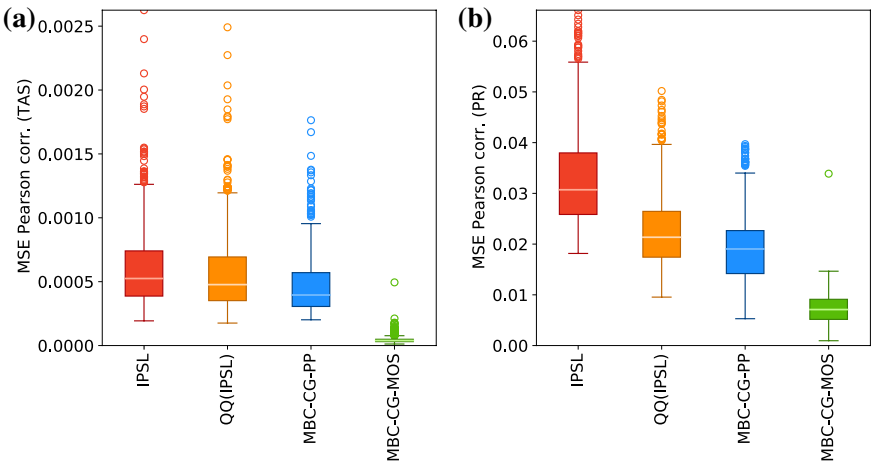

Fig. 7 Boxplots of mean squared errors of Pearson spatial correlations computed at each grid cell for **a** temperature and **b** precipitation over the projection period. Results are shown for plain IPSL, QQ, MBC-CycleGAN-PP and MBC-CycleGAN-MOS outputs

5.2.1 Univariate distribution properties

Similarly to the first cross-validation method, univariate properties are evaluated using mean differences computed at each grid cell. Figure 8 shows, for the bias-corrected outputs from the three bias correction exercises, the maps of temperature mean differences with respect to SAFRAN references. Maps for precipitation relative mean differences are presented in Fig. S6 of the Supplement. For information purposes only, standard deviation relative mean differences for temperature and precipitation are also displayed in Figs. S7 and S8, respectively.

For temperature, values of IPSL and IPSLbis mean differences (Fig. 8b, c) are high, indicating strong biases of temperature mean with respect to the SAFRAN reference dataset (Fig. 8a), although less pronounced for IPSLbis. This was somehow expected since IPSLbis data are specifically constructed to mimic the SAFRAN changes in terms of marginal (and dependence) properties. It results here in having IPSLbis temperature means closer to those from SAFRAN reference for the projection period. Map for LR SAFRAN (Fig. 8d) shows small differences with the reference. Clear improvements of the temperature mean are provided by the QQ method for each of the bias correction exercises (Fig. 8e–g). Nevertheless, quite interestingly, QQ method provides less pronounced improvements for IPSL data (Fig. 8e), suggesting a degrading effect on results of correction when changes of marginal properties between calibration and projection periods for the climate data to be corrected are not in agreement with those from the references. With regard to the performances of the MBC methods, MBC-CycleGAN presents exactly the same results as the QQ method (Fig. 8h–j), in agreement with the marginal/dependence MBC properties. For Spatial-R^2D^2 (S-R^2D^2), very slight modifications of the marginal mean values provided by QQ are observed (Fig. 8k–m), due to the use of the multivariate conditioning to adjust spatial dependence structure (Vrac and Thao 2020). Concerning Spatial-dOTC (S-dOTC), the corrected outputs for IPSLbis (Fig. 8o) and LR SAFRAN (Fig. 8p) present results similar to those obtained for QQ and MBC-CycleGAN. However, it is worth mentioning that, for the correction of IPSL, S-dOTC (Fig. 8n) slightly improves marginal properties (MAE=0.37) compared to those obtained from QQ outputs (MAE=0.42).

For precipitation relative mean differences (Fig. S6), the same conclusions hold for each (M)BC method, indicating no particular influence of the variable to correct on the results of the marginal statistics adjustment.

5.2.2 Spatial correlations

We now evaluate the ability of MBC-CycleGAN to adjust spatial dependence. First, as for the Sect. 5.1, we compute MSE of spatial Pearson correlations for both temperature and precipitation. Figure 9 displays the results with boxplots for the different datasets to correct and their adjusted outputs. Scatterplots of MSE values with respect to QQ outputs are presented in Fig. S9 to better assess the potential benefits of using MBC methods relative to univariate ones. For temperature (Fig. 9a), the positive values of MSE for IPSL suggest biases with respect to the SAFRAN references, illustrating the necessity to correct spatial properties of the model before using it in subsequent analyses. For IPSLbis, MSE values are slightly smaller, but still indicates strong differences of spatial correlations with respect to the references. The difference of results between IPSL and IPSLbis highlights that discrepancies of changes with the references can potentially have a non-negligible effect on spatial properties; in fact, reducing those discrepancies as it is done with the generation of IPSLbis leads here to reduce biases in spatial correlations. Concerning LR SAFRAN, MSE values are small, suggesting that upscaling the reference dataset deteriorates only slightly its spatial structure. By simply correcting univariate distributions, the three QQ outputs do not present a particular improvement of temperature MSE values. Clear improvements of the spatial correlation structures are provided by the MBC-CycleGAN method for the adjustment of IPSL, IPSLbis and LR SAFRAN, although some differences of performances are observed between the three corrected outputs. Temperature MSE values are indeed closer to 0 for the correction of LR SAFRAN than for the correction of IPSLbis and IPSL, for which similar results are obtained.

Concerning Spatial-R^2D^2, the corrections of IPSL and IPSLbis provide major improvements in adjusting the spatial correlations. In particular, better results are obtained for the correction of IPSLbis. However, with regard to the Spatial-R^2D^2 outputs with LR SAFRAN, the benefits provided by R^2D^2 are smaller, as not all of the spatial correlations are improved. This result can better be seen in Fig. S9e. This contrasted performance for the R^2D^2 method appears in the context of the correction of LR SAFRAN that already presents small spatial biases with respect to SAFRAN references. The correction obtained for LR SAFRAN suggests that the R^2D^2 method is too constrained by the selected conditioning to find an appropriate collection of analogues for the projection period of this specific dataset.

For Spatial-dOTC outputs, results present low MSEs values for each bias correction exercise, indicating that spatial correlations are satisfyingly corrected by this method. Nevertheless, the adjustments are slightly better for the corrected output of IPSL than for those for IPSLbis, which may be confusing here. Indeed, as dOTC is specifically designed to take into account the changes of the data to adjust in the correction procedure, better results for IPSLbis, for which changes of spatial correlations are in line with those from SAFRAN references, would have been expected. The great

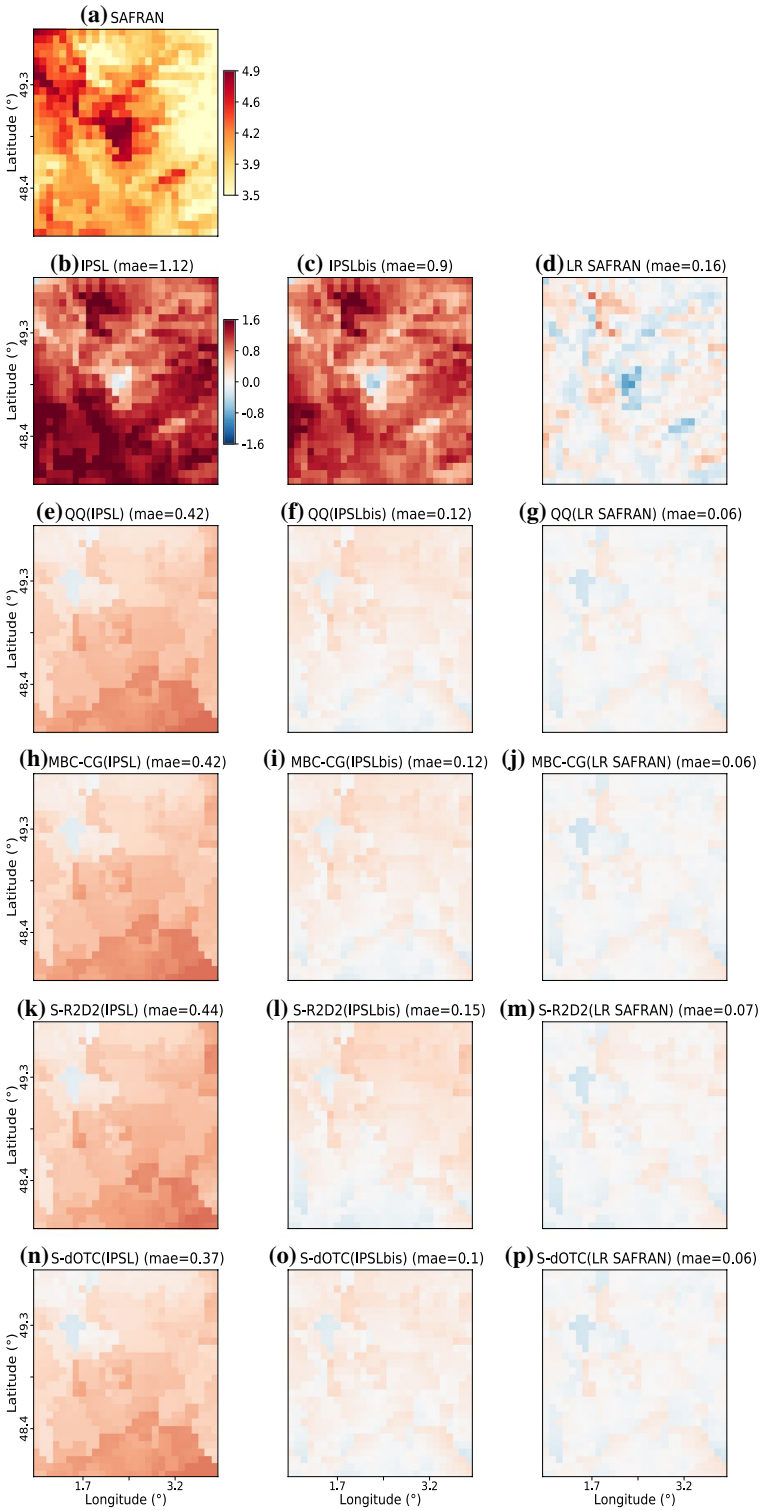

Fig. 8 Mean differences for temperature with SAFRAN reference for BC methods using as inputs **b**, **e**, **h**, **k**, **n** IPSL, **c**, **f**, **i**, **l**, **o** IPSLbis and **d**, **g**, **j**, **m**, **p** LR SAFRAN data. Results are shown during winter over the projection period for IPSL, IPSLbis, LR SAFRAN, QQ, MBC-CycleGAN, Spatial-R^2D^2 and Spatial-dOTC datasets. The map of daily mean for SAFRAN references is also shown for temperature (**a**)

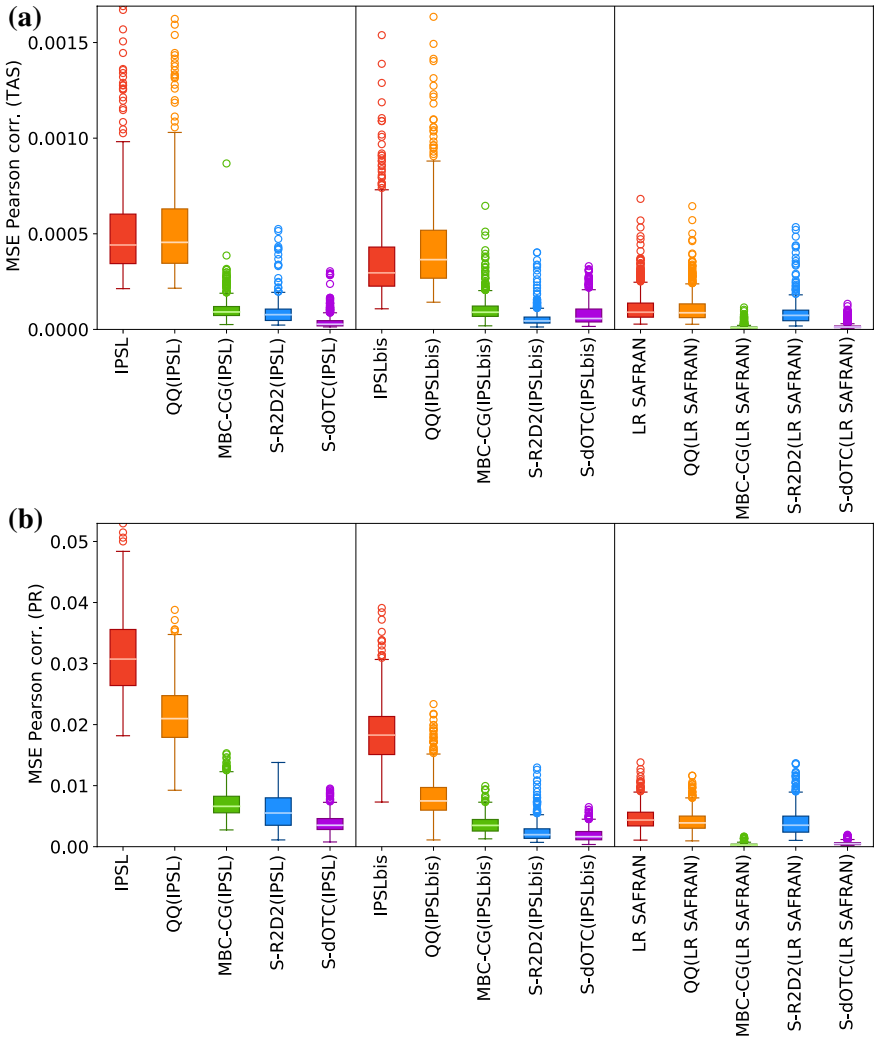

Fig. 9 Boxplots of mean squared errors of Pearson spatial correlations computed at each grid cell for **a** temperature and **b** precipitation over the projection period. Results are shown for IPSL, IPSLbis, LR SAFRAN, QQ, MBC-CycleGAN, Spatial-R^2D^2 and Spatial-dOTC datasets

performance of dOTC to correct spatial correlations for IPSL could be due to the fact that, as explained in Appendix 4, IPSL simulated changes for temperature are not in total disagreement with those from SAFRAN, and hence there is no strong discrepancy of changes affecting the corrections.

For precipitation (Fig. 9b), the same conclusions as those drawn for temperature hold. Nevertheless, quite interestingly, IPSL and IPSLbis data present even larger differences of MSE values. This shows the effects on spatial correlations of the strong discrepancies of precipitation changes between the IPSL model and the references observed in Appendix 4: reducing this discrepancy of marginal and spatial changes with IPSLbis decreases significantly the biases on spatial correlations. In contrast with temperature, these differences of spatial correlations for precipitation between IPSL and IPSLbis are significant enough to spread itself in the bias-corrected outputs: for each of the BC methods, the corrected outputs for IPSLbis present systematically lower MSE values compared to the corrections of IPSL.

To better assess spatial structure adjustments brought by MBCs, the calculation of energy distances between the bias-corrected time series and the references are performed for each physical variable according to two different multivariate distributions:

- on values of the physical variable directly over the whole region of Paris to assess differences of spatial properties (i.e., including both the marginals and their dependence);
- on ranks of the physical variable over the whole region of Paris to assess differences of spatial dependence structures (i.e., without the influence of marginal properties).

Values of energy distances are estimated using a bootstrap method. It consists for each dataset in (i) sampling (with replacement) daily fields, (ii) computing the energy distance on the bootstrapped dataset, and (iii) repeating the previous two steps 1000 times to construct the bootstrap sampling distribution. From this bootstrap sampling, distribution is deduced by the bootstrap estimator (mean of the 1000 energy distances obtained) and a 90% bootstrap sampling interval to provide uncertainty bands of the estimated distance. Results for temperature and precipitation are displayed in Fig. 10. The closer the values of the energy distances are to 0, the closer the spatial properties of the outputs are to the one of the reference data.

For temperature, the two estimators of energy distances on physical values (Fig. 10a) and ranks (Fig. 10b) for IPSL and IPSLbis data are quite high compared to those for LR SAFRAN, which is in agreement with the differences of spatial properties already observed between these datasets and the references in Fig. 9. For the three QQ outputs, while energy distances on physical values are lower (Fig. 10a), similar energy distances on ranks as those from the dataset to correct are obtained (Fig. 10b). It highlights again that, although the QQ method adjusts the univariate distributions, it is not supposed to modify rank sequence of time series, and therefore spatial dependence structures, during the correction procedure. With regard to the three MBC methods for the correction of IPSL, dOTC performs slightly better on raw values (Fig. 10a) than MBC-CycleGAN and R^2D^2, for which comparable results are obtained. For energy distances computed on ranks (Fig. 10b), dOTC and R^2D^2 produce similar results. Slightly poorer performances of MBC-CycleGAN are obtained compared to the two other MBC methods, although strongly improving the spatial dependence structures of IPSL simulations. Note that, while bootstrap sampling intervals of energy distances on temperature values are overlapping for the three MBC methods, it is less the case for energy distances on temperature ranks, thereby permitting to determine with more confidence the best method for

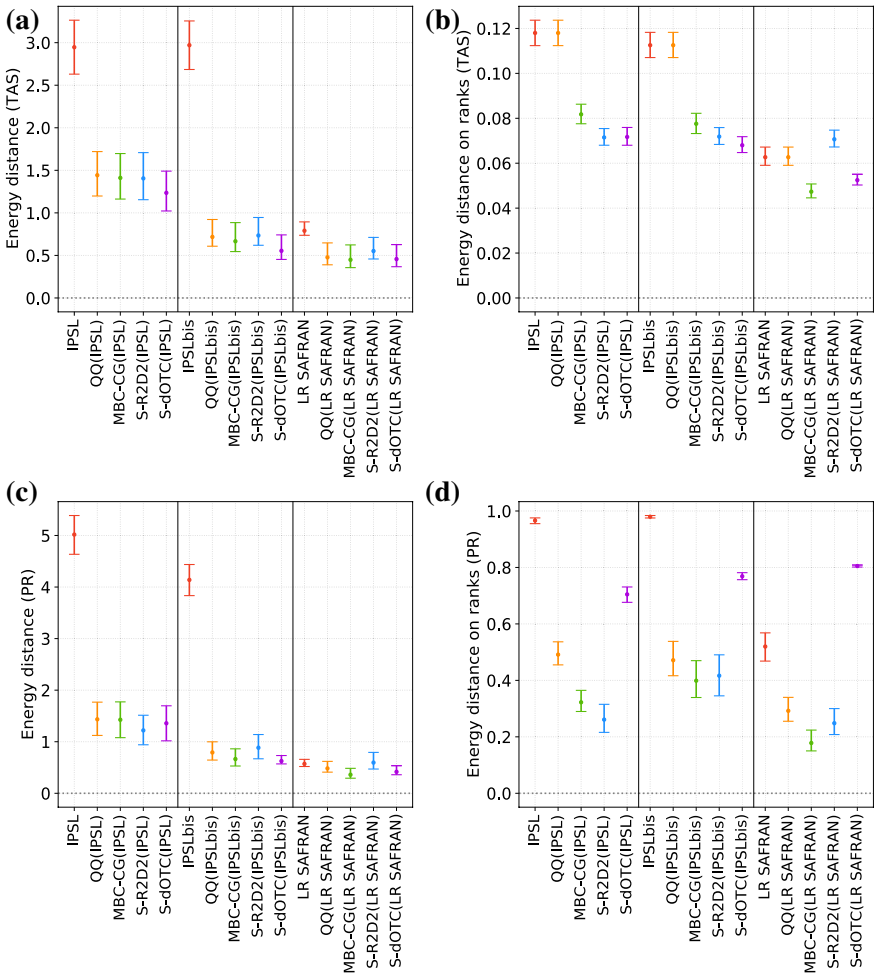

Fig. 10 Values of the estimated energy distances with respect to the reference SAFRAN for temperature (**a**, **b**) and precipitation **c**, **d** computed on physical values (**a**, **c**) and ranks (**b**, **d**) during the projection period. Results are presented for IPSL, IPSLbis, LR SAFRAN, QQ, MBC-CycleGAN, Spatial-R^2D^2 and Spatial-dOTC outputs. Estimates are evaluated using a bootstrap method (1000 replicates) that independently samples with replacement the daily fields from datasets. Note that same sequences of random days (i.e., same sampled days) are used to estimate values of energy distance for the different datasets. Error bars shows 90% bootstrap sampling intervals

the adjustment of spatial dependence properties. However, it must be mentioned that results of energy distances between the three MBCs are very close. Consequently, differences in performances between MBCs might not be significant. Concerning the correction of IPSLbis, best performances are provided by dOTC for both multivariate distributions. For multivariate distributions with raw values, MBC-CycleGAN is second best, while being third for rank dependence structure. This swap of performances between raw values and ranks for MBC-CycleGAN and R^2D^2 must be analyzed with caution as differences of estimated energy distances between the two MBC methods are again very small and thus might not be significant. This swap can however be explained by both the strong influence of marginal properties on energy distances and the slight deterioration of marginal properties provided by R^2D^2 compared to the QQ outputs, already mentioned in Sect. 5.2.1. For the corrections of LR SAFRAN, MBC-CycleGAN performs best and dOTC second best, with a more significant difference of performance for estimated energy distances evaluated on rank values (Fig. 10b).

For precipitation (Fig. 10c, d), conclusions similar to those obtained for temperature can be drawn for IPSL, IPSLbis and LR SAFRAN outputs. However, conclusions are slightly different for QQ and the MBCs. As already explained in Sect. 5.1, QQ modifies the frequency of dry events and consequently changes the rank dependence structure of precipitation, which results here in an improvement of spatial energy distances on ranks for the 1d-QQ corrections of IPSL, IPSLbis and LR SAFRAN. Concerning the performances of the three MBCs for IPSL, R^2D^2 performs best on energy distances for both raw values and ranks, while MBC-CycleGAN produces reasonable results, in particular for the adjustment of the rank dependence structure of precipitation. The dOTC method produces results that are clearly unsatisfactory concerning the rank dependence structure of precipitation. Instead of improving the rank dependence structure, dOTC correction strongly degrades it. This underperformance is in fact due to the presence of too many wet events in the corrections provided by dOTC (not shown) compared to the references, which mechanically largely affects the quality of its rank dependence structure for precipitation. For the same reason, this underperformance on precipitation rank dependence structure is also observed for the adjustments of IPSLbis and LR SAFRAN with dOTC. For IPSLbis, estimated energy distances on ranks are similar between MBC-CycleGAN and R^2D^2. Note here that similar values of energy distances do not necessarily imply that their spatial dependence structures are similar. Concerning LR SAFRAN corrections, MBC-CycleGAN again outperforms both dOTC and R^2D^2 algorithms according to estimated energy distances on raw values and ranks.

5.2.3 Temporal structure

In this section, bias-corrected data are evaluated relative to temporal properties. As a reminder, MBC-CycleGAN and dOTC methods have been specifically implemented to only adjust marginal and spatial properties of climate simulations. Similarly, the R^2D^2 algorithm is applied to adjust marginal and spatial features but, contrary to the two other methods, it also takes into account (part of) the temporal dependence properties through the multivariate conditioning chosen for its implementation, as previously explained in Sect. 4. In theory, this choice of conditioning dimensions allows R^2D^2 to partially recover temporal properties of the reference dataset (Vrac and Thao 2020). Adjusting spatial coherence necessarily modifies the rank sequences of the initial time series during the correction procedure (e.g., Vrac 2018). It is hence interesting to quantify how strong those modifications are depending on the MBC method, whether temporal properties are taken into account in the correction procedure or not. Evaluation of temporal properties is performed by computing 1-d lag Pearson autocorrelations (AR1) at each grid cell for both temperature and precipitation. The resulting maps of differences with respect to SAFRAN references for the different BC outputs are presented in Fig. 11 (resp. Fig. S10) for temperature (resp. precipitation).

For temperature, IPSL shows relatively low values of AR1 differences (Fig. 11b), indicating that temporal properties for temperature are relatively in line with those from the SAFRAN references (Fig. 11a). A similar differences map is provided by IPSLbis outputs (Fig. 11c). In fact, IPSLbis temporal properties are inherited from IPSL outputs: even in a high-dimensional context, the two-step procedure—and in particular, the matrix-recorrelation technique—used to construct IPSLbis from IPSL does not lead to a strong modification of temporal properties. This result on temporal properties of data preprocessed with this matrix-recorrelation technique is consistent with the conclusions obtained in François et al. (2020) for a MBC method (MRec) using the same matrix-recorrelation. For LR SAFRAN outputs (Fig. 11d), values of AR1 differences are very close to 0, highlighting that the upscaling step used to construct LR SAFRAN data does not strongly modify the temporal properties of the initial SAFRAN reference dataset, which was expected by construction. Difference maps for temperature from QQ outputs (Fig. 11e–g) are relatively similar to those from the three datasets to adjust, respectively. However, for the three MBC methods used to adjust spatial dependence structure, modifications of temporal properties for temperature are not equivalent. With regard to MBC-CycleGAN and dOTC outputs (Fig. 11h, i, j, n, o and p), temporal statistics are close to that from the QQ outputs. It hence suggests that both MBC-CycleGAN and dOTC algorithms, although correcting the spatial features, perform little changes of the

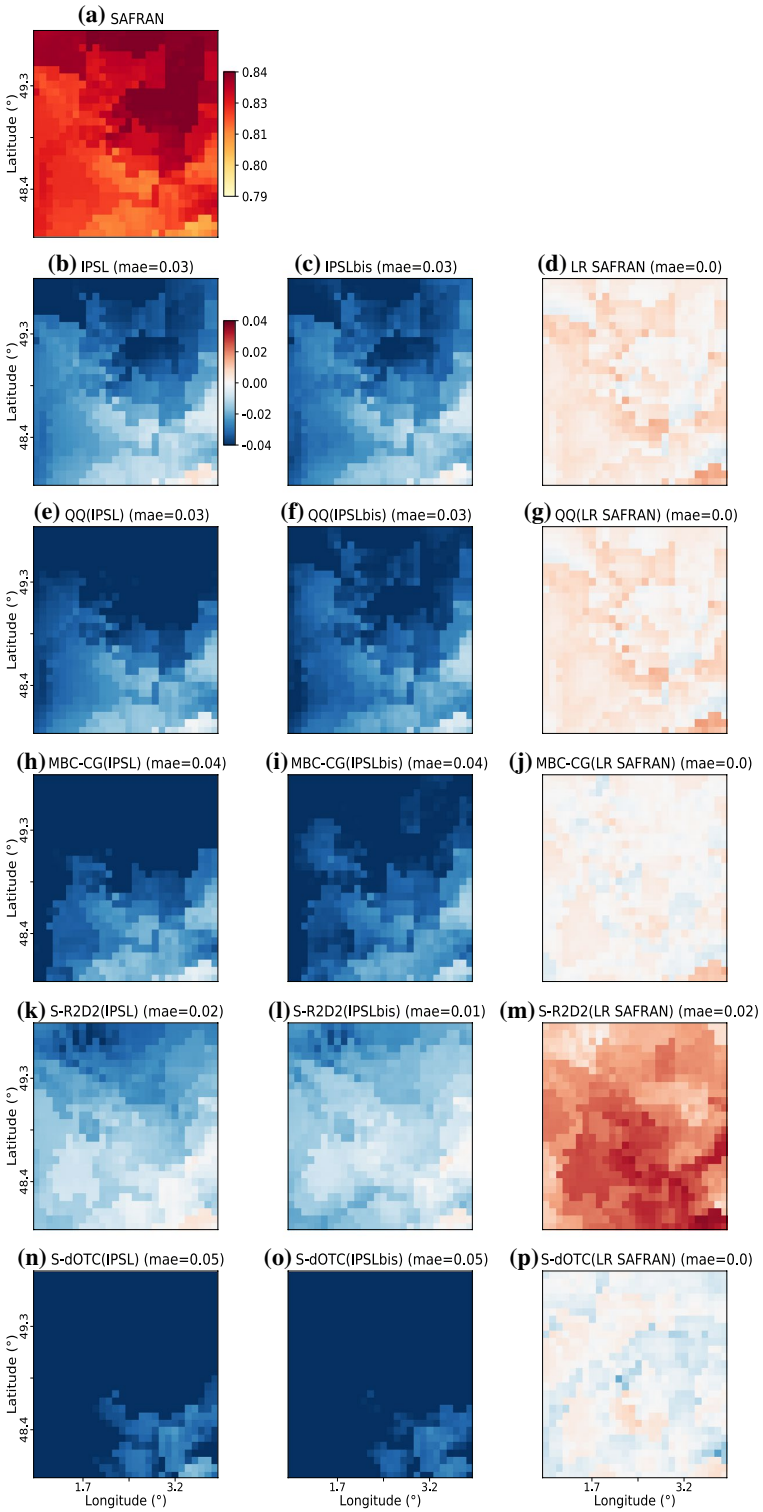

Fig. 11 Differences of order 1 Pearson autocorrelation for temperature with SAFRAN reference for BC methods using as inputs (**b**, **e**, **h**, **k**, **n**) IPSL, **c**, **f**, **i**, **l**, **o** IPSLbis and **d**, **g**, **j**, **m**, **p** LR SAFRAN data. Results are shown during winter over the projection period for IPSL, IPSLbis, LR SAFRAN, QQ, MBC-CycleGAN, Spatial-R^2D^2 and Spatial-dOTC datasets. The map of order 1 Pearson autocorrelation for SAFRAN references is also shown for temperature (**a**)

Table 1 Summary of attributes of the different climate data to correct

Climate data	Marginal prop.	Spatial prop.	Changes of marginal prop.	Changes of spatial prop.
IPSL model	From raw IPSL	From raw IPSL	Potentially not in line with SAFRAN	Potentially not in line with SAFRAN
LR SAFRAN	~ Same as SAFRAN	~ Same as SAFRAN	~ In line with SAFRAN	~ In line with SAFRAN
IPSLbis	~ Same as IPSL model	~ Same as IPSL model	~ In line with SAFRAN	~ In line with SAFRAN

temporal sequencing of the time series to correct. For MBC-CycleGAN, this is partly explained by the fact that, within the CycleGAN procedure, input maps from QQ outputs are transformed to outputs with improved spatial features, whilst not modifying too much the initial input image. It hence results in partially preserving the temporal properties of the QQ outputs used as inputs of the CycleGAN while providing improvements of the spatial representation. This particular point is thereafter discussed in greater details. Concerning R^2D^2 outputs, different results are obtained depending on the dataset to correct. For the correction of both IPSL and IPSLbis (Fig. 11k, l), R^2D^2 provides small improvements of temporal properties of temperature, which illustrates that, by including lags in the conditional dimensions, R^2D^2 is able to improve—in addition to spatial properties—temporal structure of climate datasets. However, for the correction of LR SAFRAN (Fig. 11m), a deterioration of AR1 temperature differences is obtained with respect to initial LR SAFRAN data (Fig. 11d). This result can be linked with the previously mentioned contrasted performances of the R^2D^2 method to adjust LR SAFRAN dataset in Subsect. 5.2.2.

For precipitation (Fig. S10), same conclusions hold for IPSL, IPSLbis and LR SAFRAN outputs. However, contrary to temperature, 1d-QQ corrections of IPSL and IPSLbis (Figs. S10e, f) show a pronounced improvement of temporal properties for precipitation, highlighting the potential influence of marginal properties of precipitation time series on its autocorrelation values. Moreover, the improvements of temporal properties of temperature provided by R^2D^2 for the corrections of IPSL and IPSLbis are no longer observed for precipitation (Fig. S10k, l). Instead, temporal properties with unexpected behaviors are obtained, potentially due to the difficulty of R^2D^2 to correct physical variables with events occuring at local scale, such as precipitation (Vrac and Thao 2020). It can also be due to the choice of the conditioning information made in R^2D^2. As a reminder, it is indeed the rank structure of simulated precipitation (resp. temperature) that serves as a conditioning to generate Spatial-R^2D^2 outputs for precipitation (resp. temperature). As temporal properties (including rank sequences) of precipitation time series are not well simulated by IPSL model (Fig. S10b) compared to temperature (Fig. 11b), it potentially affects the quality of the corrections—and its temporal properties—provided by Spatial-R^2D^2 for precipitation. This highlights the importance of choosing a relevant conditioning dimension for the implementation of R^2D^2 (Vrac and Thao 2020).

To illustrate the fact that MBC-CycleGAN performs little changes of the temporal sequencing of the inputs to adjust, we compare corrected daily maps from LR SAFRAN with those from the references. As the LR SAFRAN dataset is temporally matching the SAFRAN dataset by construction, classic forecast statistics such as Root Mean Square Error (RMSE) can indeed be interesting to assess the performances of MBC methods. Table 2 shows, for temperature and precipitation, the RMSE values with respect to SAFRAN references for the different BC outputs of LR SAFRAN. For temperature, the RMSE value between daily maps of the reference and the LR SAFRAN dataset is around 0.36. Slight improvement in terms of RMSE is provided by the QQ method (RMSE = 0.31). As the QQ method preserves the temporal sequencing of the times series to correct, this improvement is only due to the correction of marginal properties. The MBC-CycleGAN method presents better results (RMSE = 0.23), permitting to state with more confidence that, while adjustment of spatial dependence structure are performed, it modifies only slightly the temporal sequencing of the times series to correct. For R^2D^2 outputs, the RMSE value is quite large (RMSE=1.51), suggesting a strong modification of temporal properties. It can be linked with the underperformance of R^2D^2 already observed in Fig. 11m for the correction of LR SAFRAN. Concerning dOTC outputs, the RMSE value (= 0.42) is slightly higher than those observed for LR SAFRAN and QQ outputs. It suggests that the influence of the correction of univariate distributions and spatial dependence on temporal properties

Table 2 RMSE values between the reference SAFRAN and the different climate datasets in rows for temperature and precipitation during winter over the projection period

Physical variable	LR SAFRAN	QQ	MBC-CG	S-R^2D^2	S-dOTC
TAS	0.36	0.31	<u>0.23</u>	1.51	0.42
PR	0.75	0.73	<u>0.51</u>	3.41	1.03

As LR SAFRAN dataset is temporally matching the SAFRAN references, results are presented for LR SAFRAN and its MBC corrections only. For each physical variable, the best performing method is underlined

provided by dOTC is strong enough to affect its ability to provide appropriate forecasts at a daily scale. For precipitation, the same conclusions hold for the different BC outputs. To better illustrate the results from Table 2, two animations presenting the successive daily temperature and precipitation maps generated by MBC-CycleGAN for the correction of LR SAFRAN, as well as the corresponding daily maps from the references and the different BC methods, are provided as supplementary materials.

6 Conclusion, discussion and future work

6.1 Conclusions

Climate simulations biases are typically corrected with univariate BC methods, adjusting one physical variable and one location at a time, and thus spatial dependencies remain uncorrected. In this study, MBC-CycleGAN, an adaptation of the CycleGAN approach (Zhu et al. 2017) used to train image-to-image translation models, was presented, allowing for the adjustment of not only univariate distributions but also spatial dependence structures of climate simulations. The new suggested MBC method takes advantage of convolutional neural networks with simple architecture that are trained in competition to adjust spatial properties of simulated variables. The MBC-CycleGAN method was tested by adjusting temperature and precipitation time series from IPSL simulations with respect to the SAFRAN dataset over the region of Paris using two different cross-validation methods. The first cross-validation, that defines randomly calibration and projection periods, allows to test the new methodology in a stationary context. We took advantage of this first cross-validation method to compare two post-processing schemes (PP and MOS) approaches that differ in the statistical relationships the MBC-CycleGAN model learns to adjust spatial dependences. The MOS approach that considers biases to refer to systematic distributional differences between references and simulated climate variables was found to be more appropriate for the implementation of the MBC-CycleGAN method and was chosen to be applied for the rest of the study. The second cross-validation method, that defines chronologically calibration and projection periods, was then used to evaluate the ability of the MBC-CycleGAN method to adjust climate datasets in a non-stationary context. As IPSL simulations and SAFRAN references present different marginal and spatial changes between calibration and projection periods, two additional climate datasets (LR SAFRAN and IPSLbis) with changes that are in line with the references were specifically constructed and adjusted, allowing to better assess the quality of the corrections provided by the new method depending on the statistical biases of the data to be corrected. A wide range of metrics has been used to evaluate bias adjustment outputs with references and initial climate data and assess the corrections of univariate distributions, spatial correlations and temporal properties. In addition to the 1d-QQ method, two state-of-the-art MBC (R^2D^2 and dOTC) methods have been implemented and used as benchmarks to better evaluate the influence of nonstationary properties on the results of the MBC-CycleGAN method. The results indicate that all the (M)BC methods implemented in this study generally present similar corrections of univariate distributions. Regarding spatial properties, the benefits of using MBC methods are clear compared to the 1d-QQ method. The MBC-CycleGAN method produced reasonable adjustments of spatial correlations with respect to R^2D^2 and dOTC methods for both temperature and precipitation and the three different climate datasets to adjust. Concerning the temporal aspect, the MBC-CycleGAN method is not designed to correct this specific statistical property and tends to conserve the temporal sequencing of the time series to correct. Combined with the corrections of spatial features, this property has proved to be particularly interesting for the applications of MBC-CycleGAN when the data to correct temporally match the references (e.g., as for LR SAFRAN and SAFRAN dataset, see Sect. 5.2.2). The proposed method indeed outperformed all the others (M)BC alternatives for the correction of LR SAFRAN by generally presenting both spatial and temporal statistics closer to those from the references. Concerning nonstationary properties, it has been found that changes of both marginal and spatial properties between the calibration and projection periods of the climate data to adjust can have a non-negligible effect on the quality of corrections from the MBC-CycleGAN algorithm, and more generally from all (M)BC outputs. In a general way, better results are obtained for the corrections of simulations with changes that are in agreement with those from the references, whether the MBCs make the assumption of nonstationarity of marginal properties and dependence structures or not.

6.2 Discussion and perspectives

In this study, the development of the MBC-CycleGAN method was mainly intended as a proof of concept, in order to test if GANs can be used for multivariate bias correction of climate simulations. Although bringing results with comparable performances of correction to that of well-established MBC methods, several avenues can be considered for the improvement of the proposed algorithm.

First, in order to remain in a context of proof of concept, a simple architecture of neural networks with a small number of convolutional layers has been considered for the discriminators and generators constituting the MBC-CycleGAN method. In the same idea, a classic formulation of the CycleGAN procedure—-as initially described in Zhu

et al. (2017)—has been used with a binary-cross entropy loss function for the adversarial training (Eq. 1). Improving the training performances of GANs through more advanced architectures and optimization techniques is an active area of research (e.g., Salimans et al. 2016; Arjovsky et al. 2017; Karras et al. 2018, among others). A first natural step to potentially improve results would be to opt for a more sophisticated CycleGAN model. For example, it can be done by adding more layers in the neural network architectures of both generators and discriminators to potentially capture more complex spatial relationships for the correction of climate simulations. Also, modifying the initial adversarial loss functions (L_{GAN} in Eq. 1), as proposed in Arjovsky et al. (2017), would be interesting as it could permit to improve the stability of the learning and can prevent from mode collapse issues. However, although progress is constantly increasing concerning GANs, it is well-known that this particular class of neural networks can be more difficult to train than classical neural networks (e.g., Wu et al. 2020). The possibilities of modifications of the parameters defining a CycleGAN model are numerous, and a priori do not guarantee to improve the overall performance of the CycleGAN for the specific application of bias correction. Testing the different possibilities goes way beyond the scope of the present study and is left for future work.

Second, it has to be noted that our method, by combining the 1d-QQ method and the CycleGAN approach to adjust both marginal and spatial properties, is not designed to specifically account for any simulated changes for future periods. For marginal properties, other 1d-BC methods that are able to account for potential changes of univariate CDFs from the calibration to the projection period (e.g., CDF-t or QDM, Vrac et al. 2012; Cannon et al. 2015) can of course be employed instead of QQ, as long as they do not modify (too much) rank sequence of temperature and precipitation time series and thus do not distort the convergence of the CycleGAN procedure. Concerning changes of spatial properties, the CycleGAN approach as implemented in this study is based on the key assumption that the conditional distributions $\mathbf{X}|\mathbf{Y}$ and $\mathbf{Y}|\mathbf{X}$ are the same in the training (i.e., calibration) and test (i.e., projection) datasets. It results in our context in making a strong assumption on copula stationarity between present and future periods. Although spatial dependence structures can be considered to be stable in time as imposed by physical laws over a specific region of interest (e.g., Vrac 2018), it can not be generalized to each of the physical variables and regions. For example, more concentrated spatial rainfall events are expected with higher temperatures in the future (Guinard et al. 2015; Wasko et al. 2016). Therefore, should the changes in spatial properties in the simulations between calibration and projection periods be reproduced in the correction? By comparing our results obtained with different levels of nonstationarity in the model evolution and with two well-established MBCs based on copula stationarity (R^2D^2) and nonstationarity (dOTC) for future periods, we shed light on how the nonstationary properties of the simulations are taken into account by the different multivariate BC methods. The benefits of considering MBC methods assuming copula nonstationarity for the correction of such climate dataset are not always as clear-cut as expected compared to MBC methods assuming copula stationarity. This raises the question of whether developing MBC methods assuming copula nonstationarity is justified, i.e., whether it is worth striving for developing complicated statistical methods that consider the simulated evolution of copula in the correction procedure, and, in the end, do not produce drastically better results than MBCs assuming copula stationarity. In practice, accounting for nonstationarity of simulations in bias correction procedures still remains an open question which needs to be answered on a case-by-case basis. Developing new MBC methods that are specifically able to reproduce these simulated changes in the correction is of course an important perspective but the application of such methods would be inappropriate as long as the changes from climate simulations for future periods have not been first identified as relevant.

Third, the MBC-CycleGAN method has been developed to correct spatial correlations of climate simulations for each physical variable separately, and thus does neither consider the adjustment of inter-variable correlations nor temporal structure. A possible extension of the initial method can be the consideration of inter-variable and/or temporal correlations by providing to the CycleGAN model images with not only one but several channels of the different physical variables to correct. For example, for the adjustment of inter-variable correlations between temperature and precipitation, concatenated images of daily temperature and precipitation maps in an array of dimension $2 \times 28 \times 28$ can be provided as inputs to the adversarial neural network. Similarly, adjusting temporal correlations could be considered by adding channels with lagged versions of the physical variable. Using images with additional channels would imply to change, at least, the neural network architecture by replacing 2d-convolutional neural networks with 3d-ones to allow the CycleGAN model to consider inter-channels correlations. However, as adding additional channels can potentially make the training of the CycleGAN more complicated, it is likely that others changes relative to the architecture of neural networks and optimization techniques would be required, as those mentioned previously.

Fourth, according to the results for the correction of the references at large-scale (LR SAFRAN), MBC-CycleGAN showed greater improvements of both spatial and temporal statistics compared to the other MBC methods. These promising results suggest that MBC-CycleGAN can be used directly in downscaling applications, a practice that is not

initially recommended with univariate quantile mapping techniques (Maraun 2013; Gutmann et al. 2014). Although producing reasonable results of adjustments for temperature and precipitation spatial distributions of IPSL and IPSLbis datasets, the outperformance of MBC-CycleGAN observed for the correction of LR SAFRAN is not obtained for these climate outputs. A possible reason explaining why the performances of MBC-CycleGAN differ between these three exercises of correction concerns the importance of the distributional differences between the inputs and target dataset considered. Indeed, unsupervised image-to-image translation algorithms such as CycleGAN can present difficulties to map two random variables $\mathbf{X}$ and $\mathbf{Y}$ with probability distributions that exhibit strong differences (Gokaslan et al. 2019; Royer et al. 2020). As LR SAFRAN presents smaller bias with the references than IPSL and IPSLbis data, outstanding results are obtained for the correction of LR SAFRAN with MBC-CycleGAN, while more moderate quality results are produced for IPSL and IPSLbis. Improving the MBC-CycleGAN algorithm such that it is able to produce satisfactory results even when distributions with very strong (marginal and spatial) differences are considered is of great interest to allow its use for operational purposes.

Fifth, in this study, particular precautions have been taken to prevent overfitting during training of CycleGAN networks, such as including a regularization technique called "dropout" in both generators and discriminators architectures (see Appendix B for further details), or verifying that the performances of MBC-CycleGAN on projection periods are not deteriorated along training (not shown). These precautions permit to apply with confidence MBC-CycleGAN algorithms on projection periods. The issue of overfitting raises the question of the generalization capability of statistical models, and how they cope with new (and unseen) data. In most of the study, calibration and projection periods have been defined chronologically for the 1979–2016 period, and one can argue that small differences in terms of spatial properties are obtained between the two periods. Assessing the performances of the MBC-CycleGAN algorithm for the adjustment of climate projections with very different spatial structures remains an interesting perspective. For example, this could be done by adapting the methodology used for the generation of IPSLbis to generate alternative climate simulations for the projection period with strong spatial changes, and apply the pretrained CycleGAN neural network used for the correction of IPSL in this study.

Finally, as implemented in this study, the proposed MBC-CycleGAN algorithm produces a single correction (output) for a given input. Although essential in climate applications, uncertainty quantification of MBC-CycleGAN outputs is not estimated here. An interesting possibility of extension to model uncertainty of corrected outputs would be to introduce some stochasticity into the correction procedure by giving to the generators not only daily maps to adjust but also vectors of random noises. Then, for a given daily map, it would produce an ensemble of plausible corrections. The spread between the ensemble members would represent the uncertainty associated with the multivariate bias correction.

We hope that this study serves as a starting point for the use of GANs for multivariate bias correction of climate simulations. One of the main advantages of using MBC-CycleGAN is that adjustment is performed images by images, i.e. maps by maps. If well trained, discriminators somehow guarantee that individual generated maps produced by generators are realistic with respect to references, while daily maps with strong statistical artefacts are rejected. This is not the case for the other MBC methods such as R^2D^2 or dOTC, that provide corrected simulations with appropriate distributional statistics without being particularly constrained to generate realistic daily maps. Providing corrections with realistic maps at a daily scale can be useful for the scientific community working on climate change impacts, e.g., in hydrology, for which daily spatial features are of major concern.

Appendix A: Details on the MBC-CycleGAN method

Let consider the correction of a random variable, denoted $\mathbf{X}$ (e.g., biased climate simulations outputs) with respect to a reference random variable, denoted $\mathbf{Y}$. In our study, $\mathbf{X}$ and $\mathbf{Y}$ live in dimension $28 \times 28 = 784$ dimensions. We denote $\mathbf{X}^0$ and $\mathbf{X}^1$ the random variables to correct from climate simulations during the calibration and projection period, respectively. Similarly, $\mathbf{Y}^0$ is considered as the random variable of references for the calibration period. The goal of any BC methods is to infer future unobserved data $\mathbf{Y}^1$ from the reference variable $\mathbf{Y}^0$ during calibration, and the variables from model simulations for calibration ($\mathbf{X}^0$) and projection ($\mathbf{X}^1$) periods.

In practice, BC methods are applied to correct samples $(\mathbf{x}_1^0, \ldots, \mathbf{x}_n^0)$ and $(\mathbf{x}_1^1, \ldots, \mathbf{x}_n^1)$ from the random variables $\mathbf{X}^0$ and $\mathbf{X}^1$, with respect to a sample $(\mathbf{y}_1^0, \ldots, \mathbf{y}_n^0)$ from the random variable $\mathbf{Y}^0$. For example, 1d-bias corrections of $(\mathbf{x}_1^0, \ldots, \mathbf{x}_n^0)$ and $(\mathbf{x}_1^1, \ldots, \mathbf{x}_n^1)$ with the QQ method can be denoted $(\mathbf{qq}_1^0, \ldots, \mathbf{qq}_n^0)$ and $(\mathbf{qq}_1^1, \ldots, \mathbf{qq}_n^1)$. As explained in Sect. 3, the CycleGAN approach within the MBC-CycleGAN methodology is applied between 1d-QQ outputs and references. Hence, two generators $G_{\mathbf{QQ}\to\mathbf{Y}}$ and $G_{\mathbf{Y}\to\mathbf{QQ}}$ are considered, as well as two discriminators $D_{\mathbf{QQ}}$ and $D_{\mathbf{Y}}$. The different steps constituting the MBC-CycleGAN method are described in an algorithmic way as follows:

Algorithm 1 MBC-CycleGAN training algorithm. In all experiments presented in the paper, a batch size $m = 32$ is used.

Require: α_{disc} - the learning rate for the discriminators, α_{gen} - the learning rate for the generators, m - the batch size.

Require: $\delta_{\mathbf{QQ}}, \delta_{\mathbf{Y}}$ - the initial discriminators' parameters. $\theta_{G_{\mathbf{QQ}\to\mathbf{Y}}}, \theta_{G_{\mathbf{Y}\to\mathbf{QQ}}}$ - the initial generators' parameters.
$(\mathbf{x}_1^0, \dots, \mathbf{x}_n^0)$ a sample of the random variable $\mathbf{X}^0$.
$(\mathbf{x}_1^1, \dots, \mathbf{x}_p^1)$ a sample of the random variable $\mathbf{X}^1$.
$(\mathbf{y}_1^0, \dots, \mathbf{y}_n^0)$ a sample of the random variable $\mathbf{Y}^0$.

Ensure: $(\mathbf{z}_1^0, \dots, \mathbf{z}_n^0)$ and $(\mathbf{z}_1^1, \dots, \mathbf{z}_p^1)$ the corrections of $(\mathbf{x}_1^0, \dots, \mathbf{x}_n^0)$ and $(\mathbf{x}_1^1, \dots, \mathbf{x}_p^1)$.

1: Compute $(\mathbf{qq}_1^0, \dots, \mathbf{qq}_n^0)$ and $(\mathbf{qq}_1^1, \dots, \mathbf{qq}_p^1)$ the 1d-bias corrections of $(\mathbf{x}_1^0, \dots, \mathbf{x}_n^0)$ and $(\mathbf{x}_1^1, \dots, \mathbf{x}_p^1)$ using the quantile-mapping method (Déqué 2007).

2: Compute $(\widetilde{\mathbf{qq}_1^0}, \dots, \widetilde{\mathbf{qq}_n^0})$ and $(\widetilde{\mathbf{qq}_1^1}, \dots, \widetilde{\mathbf{qq}_p^1})$, the point-wise min-max normalizations of $(\mathbf{qq}_1^0, \dots, \mathbf{qq}_n^0)$ and $(\mathbf{qq}_1^1, \dots, \mathbf{qq}_p^1)$ with the range of $(\mathbf{y}_1^0, \dots, \mathbf{y}_n^0)$.

3: Compute $(\widetilde{\mathbf{y}_1^0}, \dots, \widetilde{\mathbf{y}_n^0})$, the point-wise min-max normalization of $(\mathbf{y}_1^0, \dots, \mathbf{y}_n^0)$.

4: **while** $\widehat{\mathcal{E}}$, the estimated energy distance on ranks between $(\mathbf{y}_1^0, \dots, \mathbf{y}_n^0)$ and $(\mathbf{z}_1^0, \dots, \mathbf{z}_n^0)$, has not converged **do**

5: Sample $\{\widetilde{\mathbf{qq}_i^0}\}_{i=1}^{m/2}$ a batch from the dataset $(\widetilde{\mathbf{qq}_1^0}, \dots, \widetilde{\mathbf{qq}_n^0})$.

6: Sample $\{\widetilde{\mathbf{y}^0}_i\}_{i=1}^{m/2}$ a batch from the dataset $(\widetilde{\mathbf{y}_1^0}, \dots, \widetilde{\mathbf{y}_n^0})$.

7: Generate "fake" samples $\{\widetilde{\mathbf{y}}_i^{0,fake}\}_{i=1}^{m/2}$:
$\forall i \in [\![1, \dots, m/2]\!], \widetilde{\mathbf{y}}_i^{0,fake} = G_{\mathbf{QQ}\to\mathbf{Y}}(\widetilde{\mathbf{qq}_i^0})$.

8: Generate "fake" samples $\{\widetilde{\mathbf{qq}}_i^{0,fake}\}_{i=1}^{m/2}$:
$\forall i \in [\![1, \dots, m/2]\!], \widetilde{\mathbf{qq}}_i^{0,fake} = G_{\mathbf{Y}\to\mathbf{QQ}}(\widetilde{\mathbf{y}_i^0})$.

9: Update $\delta_{\mathbf{QQ}}$, using Adam optimizer and the learning rate α_{disc}, by computing the adversarial loss function (Eq. 1) and its gradients with the samples $\{\widetilde{\mathbf{qq}_i^0}\}_{i=1}^{m/2}$ and $\{\widetilde{\mathbf{qq}}_i^{0,fake}\}_{i=1}^{m/2}$. The adversarial loss function must be maximized.

10: Update $\delta_{\mathbf{Y}}$, using Adam optimizer and the learning rate α_{disc}, by computing the adversarial loss function (Eq. 1) and its gradients with the samples $\{\widetilde{\mathbf{y}_i^0}\}_{i=1}^{m/2}$ and $\{\widetilde{\mathbf{y}}_i^{0,fake}\}_{i=1}^{m/2}$. The adversarial loss function must be maximized.

11: Compute the full loss function (Eq. 6) and its gradients with respect to the parameters $\theta_{G_{\mathbf{QQ}\to\mathbf{Y}}}$ and $\theta_{G_{\mathbf{Y}\to\mathbf{QQ}}}$ of the generators $G_{\mathbf{QQ}\to\mathbf{Y}}$ and $G_{\mathbf{Y}\to\mathbf{QQ}}$.

12: Update the parameters $\theta_{G_{\mathbf{QQ}\to\mathbf{Y}}}$ and $\theta_{G_{\mathbf{Y}\to\mathbf{QQ}}}$ by minimizing the full loss function, using Adam optimizer, according to its gradients and the learning rate α_{gen}.

13: Compute $(\widetilde{\mathbf{s}_1^0}, \dots, \widetilde{\mathbf{s}_n^0})$ the normalized data with a corrected spatial dependence structure for the calibration period:
$\forall i \in [\![1, \dots, n]\!], \widetilde{\mathbf{s}_i^0} = G_{\mathbf{QQ}\to\mathbf{Y}}(\widetilde{\mathbf{qq}_i^0})$.

14: Reorder each 1d-bias corrected dimension from the dataset $(\mathbf{qq}_1^0, \dots, \mathbf{qq}_n^0)$ according to its rank structure in the dataset $(\widetilde{\mathbf{s}_1^0}, \dots, \widetilde{\mathbf{s}_n^0})$ with the Schaake Shuffle method (Clark et al 2004) to obtain $(\mathbf{z}_1^0, \dots, \mathbf{z}_n^0)$, the bias correction of $(\mathbf{x}_1^0, \dots, \mathbf{x}_n^0)$.

15: Compute the estimated energy distance on ranks $\widehat{\mathcal{E}}$ evaluated between $(\mathbf{y}_1^0, \dots, \mathbf{y}_n^0)$ and $(\mathbf{z}_1^0, \dots, \mathbf{z}_n^0)$.

16: **end while**

17: Compute $(\widetilde{\mathbf{s}_1^1}, \dots, \widetilde{\mathbf{s}_p^1})$ the normalized data with a corrected spatial dependence structure for the projection period:
$\forall i \in [\![1, \dots, p]\!], \widetilde{\mathbf{s}_i^1} = G_{\mathbf{QQ}\to\mathbf{Y}}(\widetilde{\mathbf{qq}_i^1})$.

18: Reorder each 1d-bias corrected dimension from the dataset $(\mathbf{qq}_1^1, \dots, \mathbf{qq}_p^1)$ according to its rank structure in the dataset $(\widetilde{\mathbf{s}_1^1}, \dots, \widetilde{\mathbf{s}_p^1})$ with the Schaake Shuffle method to obtain $(\mathbf{z}_1^1, \dots, \mathbf{z}_p^1)$, the bias correction of $(\mathbf{x}_1^1, \dots, \mathbf{x}_p^1)$.

Appendix B: Details on the simple architecture of neural networks used in MBC-CycleGAN

The simple neural network architectures used for the discriminators and generators constituting the MBC-CycleGAN method in this study are described with more details in this appendix.

Appendix B.1: Architecture of the generators

As explained in Sect. 3.3.2, skip connections are used in the architecture of the generators to ease the training process. Skip connections permit to provide information to a given layer that comes not only from the direct previous layer, but also from other upstream convolution layers in the architecture. Skipping over layers permits to avoid vanishing gradients issues, which is a problem that can make the network hard to train. All layers except the first one have leaky rectified linear unit (leaky-ReLu) activation functions defined as: $y = \begin{cases} x & \text{if } x \geq 0, \\ \alpha x & \text{otherwise,} \end{cases}$ with $\alpha = 0.2$. Dropout regularization, that refers to ignoring neurons chosen at random during training, is used after the second and third 2D convolutional layers to prevent from overfitting (e.g., Srivastava et al. 2014). The probability used for dropout is 0.4. A summary of the simple neural network architecture used for the generators in described below in Table 3.

Appendix B.2: Architecture of the discriminators

A summary of the simple neural network architecture used for the discriminators is described below in Table 4.

Appendix C: Methodology for the generation of IPSLbis

For the generation of IPSLbis data, a two-step procedure is developed to construct, from IPSL data, climate data that present marginal and spatial changes that are in line with those from references between the calibration and projection periods. In order to stay with comparable changes as those from LR SAFRAN, LR SAFRAN changes are reproduced. We recall that, for the calibration period, IPSL and IPSLbis data are strictly identical. The two-step procedure is only used to produce alternative climate data for the projection period.

Appendix C.1: Marginal changes with CDF-t

The first step of the procedure consists in producing time series for the projection period of IPSLbis by taking into account marginal changes of LR SAFRAN with the 1d-BC named CDFt (Vrac et al. 2012). Initially, CDF-t is a version of univariate quantile mapping method designed to correct at each individual grid cell marginal properties of climate simulations outputs during the calibration and the projection period according to the data from the reference observed during calibration. CDF-t, by defining a specific transfer function, has been conceived to take into account the potential simulated changes of univariate distributions from the calibration to the projection period in order to produce the adjusted data such that the marginal changes are in line with those from the simulations. While, traditionally, this quantile-mapping approach is used to find, in a bias correction context, a mathematical transformation allowing to go from simulations to references, we here applied CDF-t to go from "large scale" references (LR SAFRAN) to simulations for future periods. By proceeding this way, the produced time series are projected distributions in the domain of IPSL simulations that have been obtained while taking into account the potential evolution of CDFs of the LR SAFRAN dataset between the calibration and projection periods. By concatenating times series from IPSL for the calibration period and those obtained from the CDF-t method for the projection period, new climate times series are obtained, presenting marginal distributions changes in line with those from references.

Appendix C.2: Spatial changes with a matrix-recorrelation technique

The second step consists in deriving a spatial dependence structure for the projection period such that spatial changes of LR SAFRAN are reproduced. To do so, we take advantage of a matrix-recorrelation technique used for the MBC method presented in Bárdossy and Pegram (2012) to impose to climate data a specific spatial dependence structure for the projection period. Our methodology is summarized in Table 5. It consists in first projecting individually each variable of both IPSL simulations and LR SAFRAN during calibration and projection periods to the univariate normal distribution with a Gaussian quantile mapping method. This "Gaussianization" step is particularly suited for variables with mixed distributions such as precipitation (composed of wet and dry events). Computing Pearson correlation matrices on such Gaussianized data instead of raw data permits to better describe its dependence structure. Thus, Pearson correlation matrices of the different Gaussianized data are computed. They are respectively denoted as $C_{I,C}$, $C_{I,P}$, $C_{I,C}^{(bis)}$, $C_{I,P}^{(bis)}$, $C_{S,C}$, $C_{S,P}$ for IPSL during calibration, IPSL during projection, IPSLbis during calibration, IPSLbis during projection, LR SAFRAN during calibration and LR SAFRAN during projection. Additionally, let $r_{I,C}$, $r_{I,P}$, $r_{I,C}^{(bis)}$, $r_{I,P}^{(bis)}$, $r_{S,C}$, $r_{S,P}$ denote one of their entry. Note that by construction, $C_{I,C}$

Table 3 The architecture of the generators used in the MBC-CycleGAN network

Layer	Layer name	Filter	Stride size	Output size
1	Input layer	n.a.	n.a.	$28 \times 28 \times 1$
2	Conv2D	$3 \times 3 \times 64$	(1, 1)	$28 \times 28 \times 64$
3	Conv2D	$3 \times 3 \times 128$	(2, 2)	$14 \times 14 \times 128$
	Leaky ReLU + Dropout	n.a.	n.a.	$14 \times 14 \times 128$
4	Conv2D	$3 \times 3 \times 256$	(2, 2)	$7 \times 7 \times 256$
	Leaky ReLU + Dropout	n.a.	n.a.	$7 \times 7 \times 256$
5	Conv2DTranspose	$4 \times 4 \times 128$	(2, 2)	$14 \times 14 \times 128$
	Skip connection (Layer 3)	n.a.	n.a.	$14 \times 14 \times 128$
	+ Leaky ReLU			
6	Conv2DTranspose	$4 \times 4 \times 64$	(2, 2)	$28 \times 28 \times 64$
	Skip connection (Layer 2)	n.a.	n.a.	$28 \times 28 \times 64$
	+ Leaky ReLU			
7	Conv2D	$1 \times 1 \times 1$	(1, 1)	$28 \times 28 \times 1$
	Skip connection (Layer 1)	n.a.	n.a.	$28 \times 28 \times 1$
	+ Leaky ReLU			

Table 4 The architecture of the discriminators used in the MBC-CycleGAN network.

Layer	Layer name	Filter	Stride size	Output size
1	Input layer	n.a.	n.a.	$28 \times 28 \times 1$
2	Conv2D	$3 \times 3 \times 64$	(2, 2)	$14 \times 14 \times 64$
	Leaky ReLU + Dropout	n.a.	n.a.	$14 \times 14 \times 64$
3	Conv2D	$3 \times 3 \times 128$	(2, 2)	$7 \times 7 \times 128$
	Leaky ReLU + Dropout	n.a.	n.a.	$7 \times 7 \times 128$
4	Flatten	n.a.	n.a.	6272
5	Dense + sigmoïd	1	n.a.	1

is the same as $C_{I,C}^{(bis)}$ and that $C_{I,P}^{(bis)}$ is unknown. Assessing the changes of LR SAFRAN spatial correlations between calibration and projection periods is now required to derive the spatial dependence structure of IPSLbis for the projection period. A simple approach to determine $r_{I,P}^{(bis)}$, the correlation of the Gaussianized data of IPSLbis for projection, would be to compute it based on the difference of correlations from Gaussianized LR SAFRAN data such as $r_{I,P}^{(bis)} = r_{I,C} + r_{S,P} - r_{S,C}$. However, computing $r_{I,P}^{(bis)}$ this way can lead to obtain correlation values that are out of range, i.e. being greater than 1 or less than -1, which is not appropriate.

From Bárdossy and Pegram (2012), given $r_{I,C}$, $r_{S,C}$, $r_{S,P}$, one can derive r_{Ib_p} using Fisher-Z transformation (Fisher 1915) as following:

$$r_{I,P}^{(bis)} = \frac{\frac{(1+r_{S,P})}{(1+r_{S,C})}(1+r_{I,C}) - \frac{(1-r_{S,P})}{(1-r_{S,C})}(1-r_{I,C})}{\frac{(1+r_{S,P})}{(1+r_{S,C})}(1+r_{I,C}) + \frac{(1-r_{S,P})}{(1-r_{S,C})}(1-r_{I,C})}. \quad (7)$$

Fisher-Z transformation permits to transform a bounded random variable to another random variable that can be assumed to be Normal, and for which additive correction can be performed (see Mehrotra and Sharma (2019) for the derivation of Eq. 7). By deriving this way all the new correlation coefficients, the potential changes in correlations in the Gaussianized LR SAFRAN data are preserved and the Pearson correlation matrix for Gaussianized IPSLbis during the projection period is obtained.

Now that the Pearson correlation matrix, $C_{I,P}^{(bis)}$, is computed, a combination of "decorrelation" and "recorrelation" steps using decompositions of correlation matrices through singular value decomposition (SVD, Beltrami 1873; Jordan 1874a, b; Stewart 1993) is applied on the Gaussianized data of IPSL during projection period, forcing its Pearson correlation matrix to be exactly the same as the Pearson correlation matrix, $C_{I,P}^{(bis)}$. The new dependence structure for IPSLbis is obtained. Finally, a reordering of time series from CDF-t outputs according to this new dependence structure is performed using the Schaake Shuffle method to obtain IPSLbis data for the projection period.

Appendix D: Spatial correlation changes analysis

We present a spatial changes analysis to provide a better picture of the properties of the climate data in terms of changes between the calibration and projection periods. As a reminder, IPSLbis data are generated using the two-step procedure described in Appendix 3 such that its marginal and dependence changes are in line with those from LR SAFRAN (and therefore SAFRAN) for the projection period. Fig. S3 displays scatterplots of differences between Spearman spatial correlations of temperature and precipitation evaluated for all pairwise combinations of sites, computed for the calibration (1979–2005) and the projection (2006–2016) period, respectively. Scatterplots compares differences of Spearman correlation with respect to those from LR SAFRAN. It permits one to visually verify if changes in the spatial dependence structure are in line to those from references at large-scale. Using rank correlation here permits to measure in isolation the spatial dependence between two sites rid of their marginal properties. Figures for the analysis of marginal changes– -in particular, mean and standard deviation changes—are also displayed in Figs. S4 and S5 for information purposes only. Results on univariate

Table 5 Summary of the different steps used to construct the spatial dependence structure of IPSLbis

Methodology step	IPSL calib. | proj.	IPSLbis proj.	LR SAFRAN calib. | proj.
0. Input data	$I_C \mid I_P$		$S_C \mid S_P$
1. Gaussianization	$G_{I,C} \mid G_{I,P}$		$G_{S,C} \mid G_{S,P}$
2. Pearson corr.	$C_{I,C} \mid C_{I,P}$	$C_{I,P}^{(bis)}$ (see Eq. 7)	$C_{S,C} \mid C_{S,P}$
3. Square root matrix		U (for $C_{I,P}^{(bis)}$)	
4. Inverse square root		T (for $C_{I,P}$)	
5. Decorr.-recorr.		$G_{I,P}^{(bis)} = G_{I,P} \times T \times U$	
6. Reorder marginals		$I_P^{(bis)}$	

properties can be briefly summarized as such: changes in marginal properties from SAFRAN references (resp. IPSL model) are in agreement (resp. disagreement) with those from LR SAFRAN for both temperature and precipitation. For IPSLbis, the application of the CDF-t method permits to obtain marginal changes for both temperature and precipitation similar to those from LR SAFRAN. Concerning spatial properties, as expected, changes in spatial correlations from SAFRAN references are (partially) in agreement with those from LR SAFRAN for both temperature (Fig. S3a) and precipitation (Fig. S3d). Concerning changes in the IPSL simulations, simulated changes of spatial correlations for temperature (Fig. S3b) are globally in line with those from LR SAFRAN, highlighting the ability of the climate model to provide appropriate temperature changes in spatial structure between the calibration and the projection periods. However, conclusions are quite different for precipitation, for which simulated changes are not in agreement at all with those from the reference at large scale (Fig. S3e). Hence, IPSL model presents discrepancy of changes for precipitation with respect to LR SAFRAN (and thus, SAFRAN references), that could potentially affect the quality of the correction depending on how MBC-CycleGAN accounts for these changes in its correction procedure. Concerning the results for IPSLbis, changes for both temperature (Fig. S3c) and precipitation (Fig. S3f) are similar to those from LR SAFRAN, confirming that the two-step methodology used to impose to IPSL specific changes of spatial correlations is appropriate here.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00382-021-05869-8.

Acknowledgements This work was granted access to the HPC resources of IDRIS under the allocation 20XX-[AD011011646] made by GENCI. MV acknowledges support from the CoCliServ project, which is part of ERA4CS, an ERA-NET initiated by JPI Climate and cofunded by the European Union.

Author contributions MV had the initial idea of the study and its structure, which was enriched by all coauthors. BF made all computations and figures, with help from ST. BF wrote the first draft of the article, with inputs, corrections and additional writing contributions from MV and ST.

Funding This research has been supported by the CoCliServ project, which is part of ERA4CS, an ERA-NET initiated by JPI Climate and cofunded by the European Union.

Availability of data and material The IPSL-CM5A-MR model data simulations as part of the CMIP5 climate model simulations can be downloaded through the Earth System Grid Federation portals. Instructions to access the data are available here: https://pcmdi.llnl.gov/mips/cmip5/data-access-getting-started.html, last access: 06 September 2020, (PCMDI, 1989). The SAFRAN reanalysis dataset is available upon request to the French National Centre for Meteorological Research (CNRM, Météo-France CNRS).

Declaration

Conflicts of interest/Competing interests The authors have no conflicts of interest to declare that are relevant to the content of this article.

Code availability The code for MBC-CycleGAN is publicly available at https://github.com/bastien-francois/MBC_CycleGAN. The R package for R^2D^2 is available at https://github.com/thaos/R2D2 (Vrac and Thao 2020). dOTC is publicly available at https://github.com/yrobink/SBCK (Robin et al. 2019).

References

Arjovsky M, Chintala S, Bottou L (2017) Wasserstein GAN. arXiv:1701.07875

Baño-Medina J, Manzanas R, Gutiérrez JM (2020) Configuration and intercomparison of deep learning neural models for statistical downscaling. Geosci Model Dev 13(4):2109–2124. https://doi.org/10.5194/gmd-13-2109-2020

Bárdossy A, Pegram G (2012) Multiscale spatial recorrelation of RCM precipitation to produce unbiased climate change scenarios over large areas and small. Water Resour Res 48:9502. https://doi.org/10.1029/2011WR011524

Bartok B, Tobin I, Vautard R, Vrac M, Jin X, Levavasseur G, Denvil S, Dubus L, Parey S, Michelangeli PA, Troccoli A, Saint-Drenan YM (2019) A climate projection dataset tailored for the European

energy sector. Clim Serv 16(100):138. https://doi.org/10.1016/j.cliser.2019.100138

Bates B, Kundzewicz Z, Wu S, Burkett V, Doell P, Gwary D, Hanson C, Heij B, Jiménez B, Kaser G, Kitoh A, Kovats S, Kumar P, Magadza C, Martino D, Mata L, Medany M, Miller K, Arnell N (2008) Climate change and water. Technical Paper of the Intergovernmental Panel on Climate Change. Tech. rep, The Intergovernmental Panel on Climate Change

Beltrami E (1873) Sulle funzioni bilineari. Giornale Mat Uso degli Stud Delle Univ 11:98–106

Berg P, Feldmann H, Panitz HJ (2012) Bias correction of high resolution regional climate model data. J Hydrol 448–449:80–92. https://doi.org/10.1016/j.jhydrol.2012.04.026

Bhatia S, Jain A, Hooi B (2020) ExGAN: adversarial generation of extreme samples. arXiv:2009.08454

Bihlo A (2020) A generative adversarial network approach to (ensemble) weather prediction. arXiv:2006.07718

Caminade C, Kovats S, Rocklov J, Tompkins AM, Morse AP, Colón-González FJ, Stenlund H, Martens P, Lloyd SJ (2014) Impact of climate change on global malaria distribution. Proc Natl Acad Sci USA 111(9):3286–3291. https://doi.org/10.1073/pnas.1302089111

Cannon AJ (2018) Multivariate quantile mapping bias correction: an N-dimensional probability density function transform for climate model simulations of multiple variables. Clim Dyn 50(1):31–49. https://doi.org/10.1007/s00382-017-3580-6

Cannon A, Sobie S, Murdock T (2015) Bias correction of gcm precipitation by quantile mapping: how well do methods preserve changes in quantiles and extremes? J Clim 28(17):6938–6959. https://doi.org/10.1175/JCLI-D-14-00754.1

Cattiaux J, Douville H, Peings Y (2013) European temperatures in CMIP5: origins of present-day biases and future uncertainties. Clim Dyn 41:2889–2907. https://doi.org/10.1007/s00382-013-1731-y

Chapman WE, Subramanian AC, Delle Monache L, Xie SP, Ralph FM (2019) Improving atmospheric river forecasts with machine learning. Geophys Res Lett 46(17–18):10627–10635. https://doi.org/10.1029/2019GL083662

Christensen JH, Boberg F, Christensen OB, Lucas-Picher P (2008) On the need for bias correction of regional climate change projections of temperature and precipitation. Geophys Res Lett 35(20):L20709. https://doi.org/10.1029/2008GL035694

Clark M, Gangopadhyay S, Hay L, Rajagopalan B, Wilby R (2004) The Schaake shuffle: a method for reconstructing space-time variability in forecasted precipitation and temperature fields. J Hydrometeor 5(1):243–262

Defrance D, Ramstein G, Charbit S, Vrac M, Famien AM, Sultan B, Swingedouw D, Dumas C, Gemenne F, Alvarez-Solas J, Vanderlinden JP (2017) Consequences of rapid ice sheet melting on the Sahelian population vulnerability. Proc Natl Acad Sci USA 114(25):6533–6538. https://doi.org/10.1073/pnas.1619358114

Dekens L, Parey S, Grandjacques M, Dacunha-Castelle D (2017) Multivariate distribution correction of climate model outputs: a generalization of quantile mapping approaches: multivariate distribution correction of climate model outputs. Environmetrics 28:e2454. https://doi.org/10.1002/env.2454

Denton E, Chintala S, Szlam A, Fergus R (2015) Deep generative image models using a laplacian pyramid of adversarial networks. arXiv:1506.05751

Déqué M (2007) Frequency of precipitation and temperature extremes over France in an anthropogenic scenario: model results and statistical correction according to observed values. Glob Planet Change 57(1):16–26. https://doi.org/10.1016/j.gloplacha.2006.11.030

Dufresne JL, Foujols MA, Denvil S, Caubel A, Marti O, Aumont O, Balkanski Y, Bekki S, Bellenger H, Benshila R, Bony S, Bopp L, Braconnot P, Brockmann P, Cadule P, Cheruy F, Codron F, Cozic A, Cugnet D, de Noblet N, Duvel JP, Ethé C, Fairhead L, Fichefet T, Flavoni S, Friedlingstein P, Grandpeix JY, Guez L, Guilyardi E, Hauglustaine D, Hourdin F, Idelkadi A, Ghattas J, Joussaume S, Kageyama M, Krinner G, Labetoulle S, Lahellec A, Lefebvre MP, Lefevre F, Levy C, Li ZX, Lloyd J, Lott F, Madec G, Mancip M, Marchand M, Masson S, Meurdesoif Y, Mignot J, Musat I, Parouty S, Polcher J, Rio C, Schulz M, Swingedouw D, Szopa S,Talandier C, Terray P, Viovy N, Vuichard N (2013) Climate change projections using the IPSL-CM5 Earth System Model: from CMIP3 to CMIP5. Clim Dyn 40(9):2123–2165. https://doi.org/10.1007/s00382-012-1636-1

Eden J, Widmann M, Grawe D, Rast S (2012) Skill, correction, and downscaling of GCM-simulated precipitation. J Clim 25:3970–3984. https://doi.org/10.1175/JCLI-D-11-00254.1

Fisher RA (1915) Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. Biometrika 10(4):507–521

François B, Vrac M, Cannon AJ, Robin Y, Allard D (2020) Multivariate bias corrections of climate simulations: which benefits for which losses? Earth Syst Dyn 2020:1–41. https://doi.org/10.5194/esd-2020-10

Gagne DJ II, Christensen HM, Subramanian AC, Monahan AH (2020) Machine learning for stochastic parameterization: generative adversarial networks in the Lorenz '96 model. J Adv Model Earth Syst 12(3):e2019MS001896. https://doi.org/10.1029/2019MS001896

Gan Z, Chen L, Wang W, Pu Y, Zhang Y, Liu H, Li C, Carin L (2017) Triangle generative adversarial networks. arXiv:1709.06548

Gauthier J (2014) Conditional generative adversarial nets for convolutional face generation. In: Class Project for Stanford CS231N: convolutional neural networks for visual recognition, Winter semester vol. 5, p 2

Gokaslan A, Ramanujan V, Ritchie D, Kim KI, Tompkin J (2019) Improving shape deformation in unsupervised image-to-image translation. arXiv:1808.04325

Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y (2014) Generative adversarial nets. Adv Neural Inf Process Syst. https://doi.org/10.1145/3422622

Gudmundsson L, Bremnes JB, Haugen JE, Engen-Skaugen T (2012) Technical note: downscaling RCM precipitation to the station scale using statistical transformations—a comparison of methods. Hydrol Earth Syst Sci 16(9):3383–3390. https://doi.org/10.5194/hess-16-3383-2012

Guinard K, Mailhot A, Caya D (2015) Projected changes in characteristics of precipitation spatial structures over North America. Int J Climatol 35:596–612. https://doi.org/10.1002/joc.4006

Guo Q, Chen J, Zhang X, Shen M, Chen H, Guo S (2019) A new two-stage multivariate quantile mapping method for bias correcting climate model outputs. Clim Dyn 53(5):3603–3623. https://doi.org/10.1007/s00382-019-04729-w

Gutmann E, Pruitt T, Clark M, Brekke L, Arnold J, Raff D, Rasmussen R (2014) An intercomparison of statistical downscaling methods used for water resource assessments in the United States. Water Resour Res 50:7167–7186. https://doi.org/10.1002/2014WR015559

Haddad Z, Rosenfeld D (1997) Optimality of empirical Z-R relations. Q J R Meteor Soc 123(541):1283–1293. https://doi.org/10.1002/qj.49712354107

He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: 2016 IEEE Conference on computer vision and pattern recognition (CVPR), pp 770–778, https://doi.org/10.1109/CVPR.2016.90

Hnilica J, Hanel M, Puš V (2017) Multisite bias correction of precipitation data from regional climate models. Int J Climatol 37:2934–2946. https://doi.org/10.1002/joc.4890

IPCC (2014) Climate change 2014: synthesis report. In: Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)]. IPCC, Geneva, Switzerland, p 151. https://www.ipcc.ch/report/ar5/syr/

Isola P, Zhu JY, Zhou T, Efros AA (2017) Image-to-image translation with conditional adversarial networks. In: 2017 IEEE Conference on computer vision and pattern recognition (CVPR), pp 5967–5976, https://doi.org/10.1109/CVPR.2017.632

Jordan C (1874a) Mémoire sur les formes bilinéaires. J Math Pures Appl 19(Deuxième Série):35–54

Jordan C (1874b) Sur la réduction des formes bilinéaires. C R Acad Sci Paris 78(Deuxième Série):614–617

Karras T, Aila T, Laine S, Lehtinen J (2018) Progressive growing of GANs for improved quality, stability, and variation. arXiv:1710.10196

Kim T, Cha M, Kim H, Lee JK, Kim J (2017) Learning to discover cross-domain relations with generative adversarial networks. arXiv:1703.05192

Kingma DP, Ba J (2017) Adam: a method for stochastic optimization. arXiv:1412.6980

Lecun Y, Bengio Y (1995) Convolutional networks for images, speech, and time-series. In: Arbib MA (ed) The handbook of brain theory and neural networks. MIT Press, Cambridge, MA, pp 255–258

Leinonen J, Berne A (2020) Unsupervised classification of snowflake images using a generative adversarial network and *K*-medoids classification. Atmos Meas Tech 13(6):2949–2964. https://doi.org/10.5194/amt-13-2949-2020

Leinonen J, Nerini D, Berne A (2020) Stochastic super-resolution for downscaling time-evolving atmospheric fields with a generative adversarial network. IEEE Trans Geosci Remote Sens. https://doi.org/10.1109/TGRS.2020.3032790

Liu Y, Racah E, Prabhat, Correa J, Khosrowshahi A, Lavers D, Kunkel K, Wehner M, Collins W (2016) Application of deep convolutional neural networks for detecting extreme weather in climate datasets. arXiv:1605.01156

Mao X, Li Q, Xie H, Lau RYK, Wang Z, Smolley SP (2017) Least squares generative adversarial networks. arXiv:1611.04076

Maraun D (2013) Bias correction, quantile mapping, and downscaling: revisiting the inflation issue. J Clim 26(6):2137–2143. https://doi.org/10.1175/JCLI-D-12-00821.1

Maraun D (2016) Bias correcting climate change simulations—a critical review. Curr Clim Chang Rep 2:211–220. https://doi.org/10.1007/s40641-016-0050-x

Maraun D, Wetterhall F, Ireson AM, Chandler RE, Kendon EJ, Widmann M, Brienen S, Rust HW, Sauter T, Themeßl M, Venema VKC, Chun KP, Goodess CM, Jones RG, Onof C, Vrac M, Thiele-Eich I (2010) Precipitation downscaling under climate change: recent developments to bridge the gap between dynamical models and the end user. Rev Geophys. https://doi.org/10.1029/2009RG000314

Marti O, Braconnot P, Dufresne J-L, Bellier J, Benshila R, Bony S, Brockmann P, Cadule P, Caubel A, Codron F, de Noblet N, Denvil S, Fairhead L, Fichefet T, Foujols M-A, Friedlingstein P, Goosse H, Grandpeix J-Y, Guilyardi E, Hourdin F, Idelkadi A, Kageyama M, Krinner G, Lévy C, Madec G, Mignot J, Musat I, Swingedouw D, Talandier C (2010) Key features of the IPSL ocean atmosphere model and its sensitivity to atmospheric resolution. Clim Dyn 34:1–26. https://doi.org/10.1007/S00382-009-0640-6

Mehrotra R, Sharma A (2016) A multivariate quantile-matching bias correction approach with auto- and cross-dependence across multiple time scales: implications for downscaling. J Clim 29(10):3519–3539. https://doi.org/10.1175/JCLI-D-15-0356.1

Mehrotra R, Sharma A (2019) A resampling approach for correcting systematic spatiotemporal biases for multiple variables in a changing climate. Water Resour Res 55(1):754–770. https://doi.org/10.1029/2018WR023270

Menick J, Kalchbrenner N (2018) Generating high fidelity images with subscale pixel networks and multidimensional upscaling. arXiv:1812.01608

Mirza M, Osindero S (2014) Conditional generative adversarial nets. arXiv:1411.1784

Mueller B, Seneviratne S (2014) Systematic land climate and evapotranspiration biases in CMIP5 simulations. Geophys Res Lett 41:128–134. https://doi.org/10.1002/2013GL058055

Muerth MJ, Gauvin St-Denis B, Ricard S, Velázquez JA, Schmid J, Minville M, Caya D, Chaumont D, Ludwig R, Turcotte R (2013) On the need for bias correction in regional climate scenarios to assess climate change impacts on river runoff. Hydrol Earth Syst Sci 17(3):1189–1204. https://doi.org/10.5194/hess-17-1189-2013

Nahar J, Johnson F, Sharma A (2018) Addressing spatial dependence bias in climate model simulations—an independent component analysis approach. Water Resour Res 54(2):827–841. https://doi.org/10.1002/2017WR021293

Nguyen H, Mehrotra R, Sharma A (2019) Correcting systematic biases across multiple atmospheric variables in the frequency domain. Clim Dyn 52:1283–1298. https://doi.org/10.1007/s00382-018-4191-6

Piani C, Haerter J (2012) Two dimensional bias correction of temperature and precipitation copulas in climate models. Geophys Res Lett 39(L20):401. https://doi.org/10.1029/2012GL053839

Racah E, Beckham C, Maharaj T, Kahou SE, Prabhat, Pal C (2017) ExtremeWeather: a large-scale climate dataset for semi-supervised detection, localization, and understanding of extreme weather events. arXiv:1612.02095

Radford A, Metz L, Chintala S (2016) Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv:1511.06434

Ramirez-Villegas J, Challinor A, Thornton P, Jarvis A (2013) Implications of regional improvement in global climate models for agricultural impact research. Environ Res Lett 8(024):018. https://doi.org/10.1088/1748-9326/8/2/024018

Randall D, Wood R, Bony S, Colman R, Fichefet T, Fyfe J, Kattsov V, Pitman A, Shukla J, Srinivasan J, Ronald S, Sumi A, Taylor K (2007) Climate models and their evaluation. Cambridge University Press, Cambridge, pp 589–662

Reichler T, Kim J (2008) how well do coupled models simulate today's climate? Bull Am Meteorol Soc 89:303–311. https://doi.org/10.1175/BAMS-89-3-303

Reichstein M, Camps-Valls G, Stevens B, Jung M, Denzler J, Carvalhais N, Prabhat M (2019) Deep learning and process understanding for data-driven Earth system science. Nature 566:195–204. https://doi.org/10.1038/s41586-019-0912-1

Robin Y, Vrac M, Naveau P, Yiou P (2019) Multivariate stochastic bias corrections with optimal transport. Hydrol Earth Syst Sci 23(2):773–786. https://doi.org/10.5194/hess-23-773-2019

Rodrigues ER, Oliveira I, Cunha RLF, Netto MAS (2018) DeepDownscale: a deep learning strategy for high-resolution weather forecast. In: 2018 IEEE 14th International Conference on e-Science (e-Science), pp 415–422, https://doi.org/10.1109/eScience.2018.00130

Roth K, Lucchi A, Nowozin S, Hofmann T (2017) Stabilizing training of generative adversarial networks through regularization. arXiv:1705.09367

Royer A, Bousmalis K, Gouws S, Bertsch F, Mosseri I, Cole F, Murphy K (2020) XGAN: unsupervised image-to-image translation for many-to-many mappings. Springer International Publishing, pp 33–49. https://doi.org/10.1007/978-3-030-30671-7_3

Räty O, Räisänen J, Bosshard T, Donnelly C (2018) Intercomparison of univariate and joint bias correction methods in changing climate from a hydrological perspective. Climate 6:33. https://doi.org/10.3390/cli6020033

Salimans T, Goodfellow I, Zaremba W, Cheung V, Radford A, Chen X (2016) Improved techniques for training GANs. arXiv:1606.03498

Scher S, Messori G (2018) Predicting weather forecast uncertainty with machine learning. Q J R Meteorol Soc 144(717):2830–2841. https://doi.org/10.1002/qj.3410

Scher S, Messori G (2019) Weather and climate forecasting with neural networks: using general circulation models (GCMs) with different complexity as a study ground. Geosci Model Dev 12(7):2797–2809. https://doi.org/10.5194/gmd-12-2797-2019

Scher S, Peßenteiner S (2020) Technical note: temporal disaggregation of spatial rainfall fields with generative adversarial networks. Hydrol Earth Syst Sci 2020:1–23. https://doi.org/10.5194/hess-2020-464

Schmidhuber J (2015) Deep learning in neural networks: an overview. Neural Netw 61:85–117. https://doi.org/10.1016/j.neunet.2014.09.003

Shi X, Chen Z, Wang H, Yeung DY, Wong W, Woo W (2015) Convolutional LSTM network: a machine learning approach for precipitation nowcasting. arXiv:1506.04214

Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R (2014) Dropout: a simple way to prevent neural networks from overfitting. J Mach Learn Res 15(56):1929–1958

Stewart GW (1993) On the early history of the singular value decomposition. SIAM Rev 35(4):551–566. https://doi.org/10.1137/1035134

Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A (2015) Going deeper with convolutions. In: 2015 IEEE Conference on computer vision and pattern recognition (CVPR), pp 1–9, https://doi.org/10.1109/CVPR.2015.7298594

Székely G, Rizzo M (2004) Testing for equal distributions in high dimension. InterStat 5:1249–1272

Székely G, Rizzo M (2013) Energy statistics: a class of statistics based on distances. J Stat Plan Inference 143:1249–1272. https://doi.org/10.1016/j.jspi.2013.03.018

Teutschbein C, Seibert J (2012) Bias correction of regional climate model simulations for hydrological climate-change impact studies: review and evaluation of different methods. J Hydrol 456:12–29. https://doi.org/10.1016/j.jhydrol.2012.05.052

Tong Y, Gao X, Han Z, Xu Y, Xu Y, Giorgi F (2020) Bias correction of temperature and precipitation over China for RCM simulations using the QM and QDM methods. Clim Dyn. https://doi.org/10.1007/s00382-020-05447-4

Tramblay Y, Ruelland D, Somot S, Bouaicha R, Servat E (2013) High-resolution Med-CORDEX regional climate model simulations for hydrological impact studies: a first evaluation of the ALADIN-Climate model in Morocco. Hydrol Earth Syst Sci 17(10):3721–3739. https://doi.org/10.5194/hess-17-3721-2013

Van Loon A, Gleeson T, Clark J, van Dijk A, Stahl K, Hannaford J, Di Baldassarre G, Teuling A, Tallaksen L, Uijlenhoet R, Hannah D, Sheffield J, Svoboda M, Verbeiren B, Wagener T, Rangecroft S, Wanders N, Van Lanen H (2016) Drought in the anthropocene. Nat Geosci 9:89–91. https://doi.org/10.1038/ngeo2646

Vandal T, Kodra E, Ganguly S, Michaelis A, Nemani R, Ganguly AR (2017) DeepSD: generating high resolution climate change projections through single image super-resolution. In: Proceedings of the 23rd ACM SIGKDD International Conference on knowledge discovery and data mining, pp 1663–1672, https://doi.org/10.1145/3097983.3098004

Vidal JP, Martin E, Franchistéguy L, Baillon M, Soubeyroux JM (2010) A 50-year high-resolution atmospheric reanalysis over France with the Safran system. Int J Climatol 30(11):1627–1644. https://doi.org/10.1002/joc.2003

Vigaud N, Vrac M, Caballero Y (2013) Probabilistic downscaling of GCM scenarios over southern India. Int J Climatol 33:1248–1263. https://doi.org/10.1002/joc.3509

Vorogushyn S, Bates PD, de Bruijn K, Castellarin A, Kreibich H, Priest S, Schröter K, Bagli S, Blöschl G, Domeneghetti A, Gouldby B, Klijn F, Lammersen R, Neal JC, Ridder N, Terink W, Viavattene C, Viglione A, Zanardo S, Merz B (2018) Evolutionary leap in large-scale flood risk assessment needed. WIREs Water 5(2):e1266. https://doi.org/10.1002/wat2.1266

Vrac M (2018) Multivariate bias adjustment of high-dimensional climate simulations: the rank resampling for distributions and dependences (R^2D^2) bias correction. Hydrol Earth Syst Sci 22(6):3175–3196. https://doi.org/10.5194/hess-22-3175-2018

Vrac M, Thao S (2020) R^2D^2 v2.0: accounting for temporal dependences in multivariate bias correction via analogue ranks resampling. Geosci Model Dev 2020:1–29. https://doi.org/10.5194/gmd-2020-132

Vrac M, Drobinski P, Merlo A, Herrmann M, Lavaysse C, Li L, Somot S (2012) Dynamical and statistical downscaling of the French Mediterranean climate: uncertainty assessment. Nat Hazards Earth Syst Sci 12(9):2769–2784. https://doi.org/10.5194/nhess-12-2769-2012

Vrac M, Noël T, Vautard R (2016) Bias correction of precipitation through singularity stochastic removal: because occurrences matter. J Geophys Res Atmos 121:5237–5258. https://doi.org/10.1002/2015JD024511

Wang J, Liu Z, Foster I, Chang W, Kettimuthu R, Kotamarthi R (2021) Fast and accurate learned multiresolution dynamical downscaling for precipitation. arXiv:2101.06813

Wasko C, Sharma A, Westra S (2016) Reduced spatial extent of extreme storms at higher temperatures. Geophys Res Lett 43(8):4026–4032. https://doi.org/10.1002/2016GL068509

Wheeler T, von Braun J (2013) Climate change impacts on global food security. Science 341(6145):508–513. https://doi.org/10.1126/science.1239402

Wilcke RAI, Mendlik T, Gobiet A (2013) Multi-variable error correction of regional climate models. Clim Change 120:871–887. https://doi.org/10.1007/s10584-013-0845-x

Wilks DS (2006) Statistical methods in the atmosphere science. Academic Press

Wu JL, Kashinath K, Albert A, Chirila D, Prabhat Xiao H (2020) Enforcing statistical constraints in generative adversarial networks for modeling chaotic dynamical systems. J Comput Phys 406(109):209. https://doi.org/10.1016/j.jcp.2019.109209

Xie Y, Franz E, Chu M, Thuerey N (2018) TempoGAN: a temporally coherent, volumetric GAN for super-resolution fluid flow. ACM Trans Graph. https://doi.org/10.1145/3197517.3201304

Xu CY (1999) From GCMs to river flow: a review of downscaling methods and hydrologic modelling approaches. Prog Phys Geogr 23:229–249. https://doi.org/10.1177/030913339902300204

Yi Z, Zhang H, Tan P, Gong M (2017) DualGAN: unsupervised dual learning for image-to-image translation. In: 2017 IEEE International Conference on computer vision (ICCV), pp 2868–2876, https://doi.org/10.1109/ICCV.2017.310

Yoo D, Kim N, Park S, Paek AS, Kweon IS (2016) Pixel-level domain transfer. arXiv:1603.07442

Zhu JY, Park T, Isola P, Efros AA (2017) unpaired image-to-image translation using cycle-consistent adversarial networks. arXiv:1703.10593

Zscheischler J, Westra S, Hurk B, Seneviratne S, Ward P, Pitman A, AghaKouchak A, Bresch D, Leonard M, Wahl T, Zhang X (2018) Future climate risk from compound events. Nat Clim Change. https://doi.org/10.1038/s41558-018-0156-3

Zscheischler J, Fischer E, Lange S (2019) The effect of univariate bias adjustment on multivariate hazard estimates. Earth Syst Dyn 10:31–43. https://doi.org/10.5194/esd-10-31-2019

3.4. Article published in *Climate Dynamics:* Adjusting spatial dependence of climate model outputs with cycle-consistent adversarial networks

3.5 Summary and conclusions

In this chapter, a new multivariate BC method based on a Machine Learning technique has been developed in a proof-of-concept context. The new method, named MBC-CycleGAN, takes advantage of convolutional neural networks to adjust spatial properties of climate simulations. The method was tested by adjusting temperature and precipitation time series from IPSL simulations with respect to the SAFRAN dataset over the region of Paris using two different cross-validation methods.

The first cross-validation method permitted to test MBC-CycleGAN in a stationary context. I took advantage of this cross-validation method to assess to different post-processing schemes (Perfect-Prog and Model Output Statistics). These two post-processing schemes differ in the way the MBC-CycleGAN model learns to adjust spatial dependencies. Applying MBC-CycleGAN with the Model Output Statistics approach was found to be more appropriate to adjust IPSL simulations.

The second cross-validation method was then used to test the MBC-CycleGAN method in a non-stationary context. In order to better assess the performances of the method to adjust IPSL simulations in a non-stationary context, two additional climate datasets (LR SAFRAN and IPSLbis) with multivariate changes in line with SAFRAN references were constructed and corrected. Statistical metrics have been applied to multivariate bias corrected outputs to evaluate the adjustment of univariate distributions, spatial correlations and temporal properties. One univariate (1d-QQ) and two multivariate BC methods (R^2D^2 and dOTC) have been implemented in order to assess the performances of the new method. The MBC-CycleGAN method produced similar adjustments of univariate distributions and reasonable adjustments of spatial correlations with respect to the others multivariate BC alternatives for both temperature and precipitation. Concerning the temporal properties, the MBC-CycleGAN method is not designed to adjust nor constrain this statistical property, but rather tends to preserve the temporal properties of the simulated time series.

Despite developed in a proof-of-concept context and with a quite simple architecture, MBC-CycleGAN presents promising performances of corrections compared to others state-of-the-art MBC methods. Several perspectives can be envisaged for the improvement of the proposed algorithm and obtain even better results. The work carried out in this chapter can serve as a starting point for the use of Machine Learning tools for multivariate bias correction.

3.6 A few comments on what did not work

Many unsuccessful attempts were made before finding the final MBC-CycleGAN algorithm described in the article. I propose here to give some details on these failures encountered during the work carried out in this chapter.

Keeping it simple in the face of the myriad of possibilities

A key challenge in the design of a neural network resides in the choice of a suitable architecture which allows to learn complex data patterns while avoiding overfitting issues. Hyperparameters which define the model architecture are numerous, and the process of searching for the optimal model architecture by exploring all the different possibilities can be quite tedious.

This is particularly true for GAN-based algorithms that use several neural networks (generators and discriminators) interacting with each other. My work being a proof of concept, I quickly opted for a simple CycleGAN neural network architecture by fixing most of the hyperparameters following best practices in the machine learning community (He et al., 2016; Srivastava et al., 2014; Zhu et al., 2017). Only the parameters controlling the "speed of learning" of generators and discriminators along training (called learning rates) were not fixed and were needed to be refined.

"Marginal/dependence" vs. "All-in-one"

As explained in the study, MBC-CycleGAN is developed in the context of the marginal/dependence category: marginal properties are adjusted separately from the spatial properties (Fig. 2.1a). However, the initial idea was not to use CycleGAN for multivariate bias correction in a marginal/dependence context but rather in a all-in-one context, i.e., adjusting univariate and spatial properties from climate simulations at the same time (Fig. 2.1b). If we refer to the algorithmic steps of MBC-CycleGAN described in Fig. 3 of the article, initial attempts therefore consisted to skip Step 1 by providing directly climate simulations data to CycleGAN without first adjusting univariate properties of climate models with the quantile-quantile method. By applying CycleGAN in the all-in-one context, I quickly obtained appropriate corrections for the adjustment of LR SAFRAN, but not for IPSL simulations. As IPSL simulations presents different marginal and spatial properties than those from SAFRAN contrary to LR SAFRAN dataset, it seemed that correcting both IPSL marginal and spatial properties at the same time was too much to ask for CycleGAN, at least with the architecture considered. It was only when adjusting univariate properties with QQ before applying CycleGAN that reasonable results were obtained, i.e., developing MBC-CycleGAN in the marginal/dependence category. By first reducing univariate properties biases, QQ permits to provide to CycleGAN source and target datasets with similar univariate distributions. The image-to-image translation problem is then simplified for CycleGAN, which can focus on adjusting spatial distributions using CNNs, what is not adjusted by the QQ method. The first results obtained using MBC-CycleGAN in the marginal/dependence context then encouraged me to pursue tries and tests in this direction, which led to the work presented in this chapter.

Machine Learning is not magical

This experience of research demonstrates the importance of domain knowledge to develop Machine Learning techniques for specific applications. Machine Learning tools, although powerful, are not "magic" and potentially need to be adapted depending on the scientific problem considered. If Machine Learning applications are not accompanied by a thorough understanding of the problem by scientists, they can be expected to fail or produce spurious results. Further research on the neural network architecture such as implementing additional layers could have produced reasonable results in the all-in-one context, but such architecture modifications would go beyond the scope of this proof-of-concept study.

3.7 Perspectives

The work developed in this chapter leads to several perspectives for the improvement of the proposed algorithm, but also to more general perspectives for bias correction using Machine

Learning tools. I will first return to some of the perspectives mentioned in the article (subsection 6.2). Then, additional perspectives will be discussed.

3.7.1 A look back at some of the perspectives of the article

From proof of concept to operational uses?

Although presenting similar performances of correction compared to state-of-the-art methods, the MBC-CycleGAN algorithm can be improved by several means. For example, a more sophisticated CycleGAN model can be adopted by increasing the number of layers in the generators and discriminators' architectures to potentially capture more complex spatial relationships that would permit to achieve better results of corrections. Also, opting for a more sophisticated adversarial loss, such as the Wasserstein loss (Arjovsky et al., 2017) to train generators and discriminators would be an interesting perspective as its effectiveness has been demonstrated to stabilise the training of GANs algorithms in several studies (e.g., Gulrajani et al., 2017; Petzka et al., 2017). However, although all these modifications of architectures may seem promising, they do not guarantee better results of corrections. Opting for these changes may even make the training of the methodology more difficult to achieve. As explained previously in subsection 3.6, MBC-CycleGAN has been developed in a proof of concept context and this work was intended to investigate if GANs can be used for multivariate bias correction of climate simulations. There is still some research to do to improve the MBC-CycleGAN method and make it highly generalisable to allow its use to adjust other climate models and larger regions.

Adjusting inter-variable and temporal properties using CycleGAN

The MBC-CycleGAN method has been designed to adjust univariate properties and spatial dependencies of climate simulations only. An interesting perspective could be to modify MBC-CycleGAN so that the CNNs within generators architecture considers inter-variable and temporal correlations. This could be done for example by considering 3-d filters (i.e., arrays of weights) instead of 2-d filters (i.e., matrices of weights) as presented in Fig. 3.2. By performing convolutional operations over more than 2 dimensions of simulated daily maps, the 3-d filters would permit to potentially capture inter-variable and/or temporal correlations and thus to allow MBC-CycleGAN to adjust these multivariate properties. However, considering these additional adjustments will necessarily modify, even slightly, performances of the method to adjust univariate and spatial properties, which should be carefully considered when evaluating the method.

Investigating uncertainty with stochasticity

As designed in this study, a single correction is provided by MBC-CycleGAN for each simulated daily maps to be adjusted. Another interesting perspective would be to introduce some stochastic components in the correction procedure in order to obtain from MBC-CycleGAN ensemble of corrections for a given simulated daily maps. This would then permit to (partially) investigate the inherent uncertainty associated with the MBC-CycleGAN method to adjust climate simulations. For example, providing vectors of random noises to the generators in addition to daily maps to adjust would be a first possibility for such investigations.

3.7.2 Additional perspectives

Towards an interpretable bias correction

Despite the relative success of my new method to produce appropriate bias corrected outputs, one of its main issue is that the transformation function approximated with neural networks is not easily interpretable. The lack of interpretability is not specific to CycleGAN but rather to complex neural networks architectures that often render difficult the analysis of the approximated function. Increasing interpretability of neural networks, often called "black-box" models, is a key challenge in Machine Learning research (Daniely et al., 2016; Zintgraf et al., 2017; Rudin, 2018; Baño Medina, 2020), as the lack of interpretability has often prevented the use of these models in scientific disciplines including climate science. For our application, although the mapping learned by CycleGAN is constrained by the cycle-consistency and identity losses, this does not tell us anything about how a specific simulated map with inappropriate spatial properties is adjusted. In our application, it resulted in obtaining mappings that somehow conserve some of the characteristics of the initial simulated maps in the corrections. This result is consistent with other successful CycleGAN applications (e.g., Zhu et al., 2017), despite no theoretical guarantees. Further constraining the solution space of mapping functions of CycleGAN is thus an important perspective that would potentially lead to a more interpretable bias correction method. An interesting idea proposed by de Bézenac et al. (2021) would be to reformulate the optimization problem solved by CycleGAN with Optimal Transport theory. It would permit to make CycleGAN converging to the mapping minimising the "transformation cost" needed to transform an image to another, and thus would increase interpretability of the algorithm.

Adjusting biases of warmer climates

As mentionned in the perspectives of the paper, it can be argued that calibration and projection periods in the non-stationary context are close in time and that the ability of MBC-CycleGAN to generalise to datasets far outside of the calibration period is not investigated. Actually, although it would be interesting to test it, there are no particular reasons why MBC-CycleGAN would generalise well far outside of the calibration set, e.g., in periods with simulated climates that are ≥ 4 °C warmer. This problem is common in many Machine Learning applications where algorithms are neither specifically designed to address this problem nor tested in this context. However, one recent study from Beucler et al. (2021) proposed a promising strategy to design "climate-invariant" neural networks aimed to present better ability to generalise to unseen climate data. By incorporating physical knowledge of climate change into neural network frameworks, they demonstrated that generalization can be achieved by modifying — or rescaling — inputs in a meaningful way so that invariant relationships can be learned. Implementing such "climate-invariant" strategy within a neural network-based algorithm for multivariate bias correction is of course an interesting perspective to explore in future research.

Machine Learning as a promising tool for MBC

Developing new MBC methods based on other Machine Learning tools is an interesting perspective. For example, implementing unsupervised variational autoencoder architectures (Kingma and Welling, 2013; Liu et al., 2017) or energy-based generative models (Zhao and Chen, 2020) to adjust statistical properties of climate simulations can be envisaged. Also, one can think about implementing other algorithms to correct not only spatial properties but also

inter-variable or temporal properties (e.g., by constraining motions between generated images, Bashkirova et al., 2018; Chen et al., 2019). In this context of developing new Machine Learning-based methods for multivariate bias correction, a benchmark dataset consisting of preprocessed climate simulations and predefined evaluation metrics would be needed. Such efforts in designing benchmark datasets have been recently carried out for weather forecasting (WeatherBench, Rasp et al., 2020; Garg et al., 2022) and climate emulation (ClimateBench, Watson-Parris et al., 2021). By defining training and validation datasets, this benchmark dataset would permit to provide a framework to researchers in which to evaluate progress in multivariate bias correction methods. This benchmark dataset would participate to boost research in Machine Learning-based MBC methods and improve their understanding and their applicability for multivariate bias correction.

Résumé

Contexte et objectifs

Les simulations climatiques présentent des biais spatiaux, c'est-à-dire que les propriétés spatiales (par exemple, la corrélation des précipitations entre différents lieux) peuvent différer des observations. Par conséquent, corriger les propriétés spatiales est nécessaire, par exemple pour les études d'impact. Dans ce chapitre, je propose d'adapter une technique de vision par ordinateur, utilisée initialement pour des problèmes de transformation d'image à image (appelée CycleGAN), pour l'ajustement des propriétés spatiales des simulations climatiques.

Méthodes

L'algorithme proposé, nommé MBC-CycleGAN, consiste à apprendre à transférer des cartes simulées (vues comme des images) issues de simulations climatiques ayant des propriétés spatiales inappropriées vers des images plus réalistes ayant des propriétés spatiales similaires à celles observées. J'ai testé MBC-CycleGAN en effectuant différents exercices de correction de biais, tous par rapport aux références SAFRAN pour les données de température et de précipitation sur la région parisienne.

Résultats

J'ai comparé les résultats de la correction des propriétés spatiales avec d'autres méthodes de correction de biais de l'état-de-l'art. Notre méthode MBC-CycleGAN produit des ajustements raisonnables des corrélations spatiales pour la température et les précipitations par rapport aux autres méthodes.

Ces résultats suggèrent que l'utilisation d'algorithmes tel que CycleGAN est très prometteuse pour la correction de biais multivariés. Bien sûr, des recherches supplémentaires sont nécessaires, comme l'extension de notre méthode pour corriger non seulement les propriétés spatiales mais aussi les propriétés inter-variables et/ou temporelles. Le travail réalisé dans ce chapitre peut servir de point de départ à l'utilisation d'outils de Machine Learning pour la correction de biais multivariés.

Chapter 4

Time of Emergence of compound events: contribution of univariate and dependence properties

The core of this chapter is an article submitted to the scientific journal *Natural Hazards and Earth System Sciences*. The article is preceded by an introduction giving some information on the work carried out. It is then completed by a section recalling the main conclusions and presenting some additional perspectives.

4.1 Introduction

In the previous two chapters, we were interested in multivariate statistical methods designed to improve the realism of the *complete* distributions of simulated variables, including univariate, inter-variable and/or spatial properties. However, in some applications, it is not the complete distribution that is of interest, but rather a subset of the multivariate distribution. This is the case when analysing multivariate hazards leading to compound events, which involves the analysis of the multivariate distribution of climatic variables simultaneously or successively reaching critical values. For example, for a study aimed at designing an effective coastal flood protection infrastructure, statistical properties of wind and precipitation extremes (including marginal and dependence properties) are of crucial importance, whereas the statistical characteristics of low wind and precipitation values would be of little interest.

Having an appropriate multivariate distribution between climate hazards is all the more important as marginal and dependence properties characterise probability of compound events. Evaluating the probability of multivariate hazards forming CEs, as well as having a sound knowledge of their changes are crucial for adaptation and mitigation strategies. An important concept for adaptation and mitigation is the Time of Emergence (ToE) of climate signals. It consists in determining the time at which a climate signal emerges from (i.e., goes out of) the natural variability (e.g., Christensen et al., 2007; Giorgi and Bi, 2009; Hawkins and Sutton, 2012; Maraun, 2013a). ToE has obvious relevance to adaptation and mitigation policy as huge impacts arise most likely when changes emerge from natural variability (e.g., Lobell and Burke, 2008). This concept has been used extensively to evaluate the emergence of different climate metrics such as mean temperatures (e.g., Hawkins and Sutton, 2012; Mahlstein et al., 2011),

precipitation (Fischer et al., 2014; Giorgi and Bi, 2009; Gaetani et al., 2020), but also extremes (e.g., Diffenbaugh and Scherer, 2011; King et al., 2015).

Several methods have been developed to evaluate ToE of climate signals. For instance, for a metric of interest, the ToE can be evaluated by computing the ratio between the estimated climate change signal (S) and the variability or noise (N) associated to natural variability (e.g., Giorgi and Bi, 2009; Hawkins and Sutton, 2012; Maraun, 2013b; Ossó et al., 2022). Fig. 4.1a illustrates schematically this method to determine the ToE for annual global mean surface temperature change anomalies (Hawkins et al., 2020). The "signal" (black line) of global temperature change is defined as the values of the smoothed time series. Standard deviation of residuals from the smooth fit defines the noise (light red). The first date for which the signal-to-noise ratio (S/N) *permanently* crosses a certain threshold is the ToE. Frame et al. (2017) proposes to describe the emergence of climate with respect to natural variability using the terms "unusual" (S/N $>$ 1), "unfamiliar" (S/N $>$ 2) and "unknown" (S/N $>$ 3). In Fig. 4.1a, a ToE for an unusual climate is detected around 1935 as the signal is greater than the noise (S/N $>$ 1) thereafter. Another method to evaluate ToE consists in using statistical tests to evaluate differences in *distribution* of the metric across sliding windows (e.g. using Kolmogorov-Smirnov tests, Mahlstein et al., 2012; Gaetani et al., 2020; Pohl et al., 2020). Fig. 4.1b illustrates this second method. A baseline period from which the natural variability is estimated is first defined. Here, the 30-year period 1871-1900 is chosen (blue area in Fig. 4.1b). The distribution of the values of global temperature within this baseline period (blue density on the right) is then compared to the distribution of the values for the different 30-year windows sliding over the period 1872-2021. Using this method, the ToE can be defined as the central year of the first period for which the *distribution* of the metric is significantly and permanently different from the baseline period distribution (Gaetani et al., 2020). For instance in Fig. 4.1b, the ToE is defined around 1946 if 1) the orange density on the right is significantly different from the blue one, 2) the distribution of the metric within the previous sliding window is not significantly different from the blue distribution and 3) if the distributions of the metric across the following sliding windows are all significantly different from the baseline period distribution.

The methodologies described above have been applied in the literature to define ToE of univariate climate metrics, such as mean temperature or precipitation changes. Thus, the various ToE studies cited above implicitly investigate the emergence of univariate properties only. However, for CEs, and as already explained, it is not only the univariate properties that define their probabilities, but also the dependence properties between climate hazards. Then, in addition to investigate if compound events probability changes significantly and from when it emerges, it is crucial to determine how much of this change is due to 1) changing univariate properties, and 2) changing dependence structure of multivariate hazards forming CEs. Quantifying the contribution of these statistical features to these probabilities changes would permit to better understand the potential future evolutions and emergences of compound events. In this chapter, I will propose a new multivariate statistical tool to evaluate the Time of Emergence of multivariate hazards probabilities. To define the ToE, I will estimate confidence intervals of baseline period's probabilities. It will permit to characterise the natural variability of our probabilities of interest. The ToE of hazard probabilities will be then the time period when the signal of CEs probability permanently goes out from the confidence interval. Based on copula, this new tool would also permit to disentangle and quantify the contribution of marginal and dependence properties in the probability changes of multivariate hazards leading to compound events. The new methodology will be applied to a 13-member multi-model ensemble (CMIP6) to analyse two different multivariate hazards with potential high-impacts: i) compound wind and precipitation

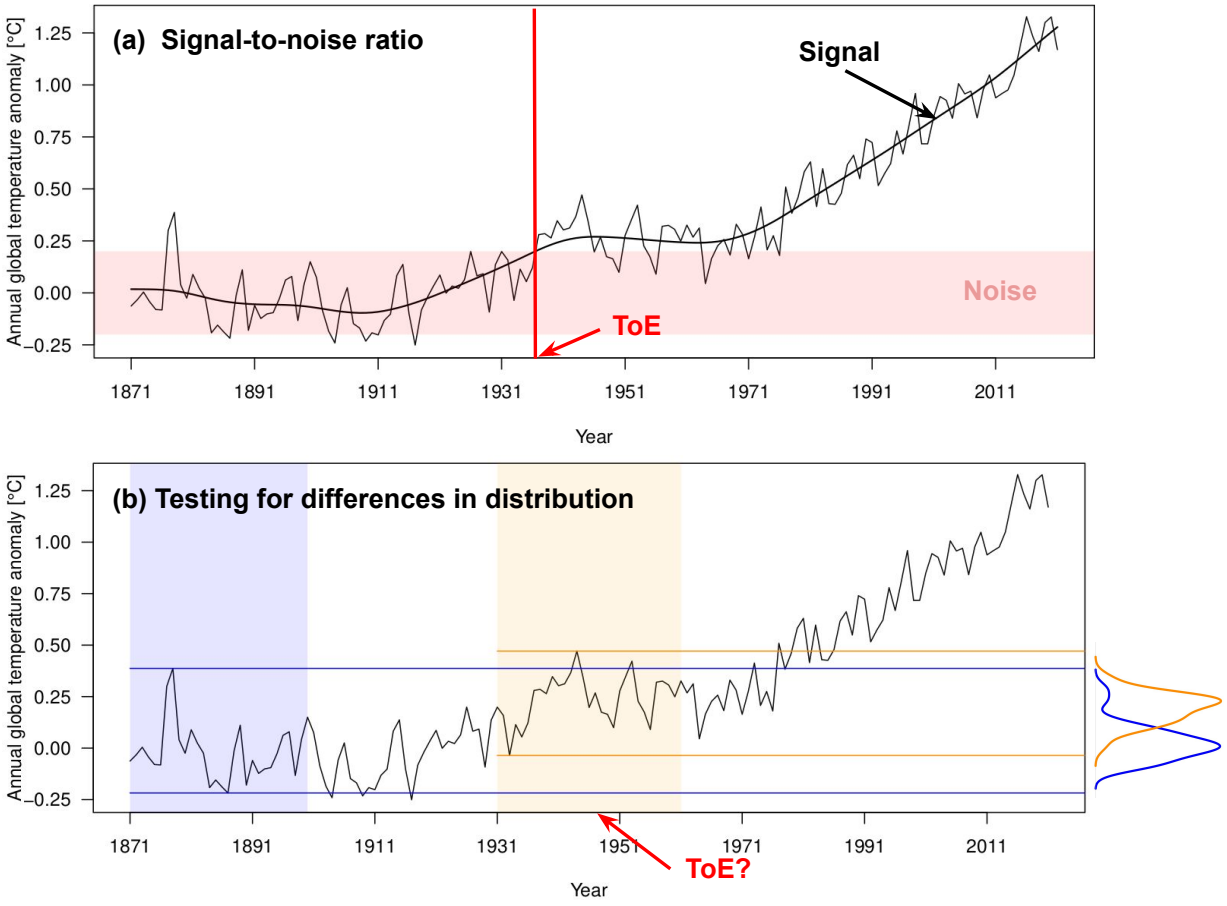

Figure 4.1: Illustration of two methodologies used in the literature to define ToE. Data are observed global temperature change from 1871 to 2021 from the Berkeley Earth temperature data set (Rohde et al., 2013). ToE can be determined using (a) Signal-to-noise ratio (graphic inspired and modified from Hawkins et al. (2020)) or by (b) testing for differences in distribution between a baseline period (blue area) and sliding windows (orange area). The reader is referred to the text for more explanations.

extremes over the region of Brittany (France), and ii) frost events occurring during the growing season preconditioned by warm temperatures over Central France. Considering 13 CMIP6 models will permit to assess inter-model variability of results for the analysis of probability changes, as well as for the contributions of the marginal and dependence properties.

Even if multivariate bias correction is not at the heart of this chapter, the new methodology will provide valuable information about simulated compound events probabilities, and how dependence contribute to these changes. This information would then be interesting to use in a context of multivariate bias correction. Indeed, as we have seen in Chapters 2 and 3, some multivariate bias correction methods permit the evolution of the simulated dependence to be taken into account in the bias correction procedure. Evaluating the simulated evolution of dependence and its contribution for changes of compound events probabilities will thus provide important information about the necessity of taking into account multi-dimensional changes in the bias correction procedure for compound events analyses.

Note that in this chapter, the analysis of CEs focuses on the multivariate statistical dis-

tribution of climate hazards. Whether or not the climate hazards lead to an impact is not investigated. The results can instead be interpreted as the climatology of potential high-impact events.

4.2 Article submitted to *Natural Hazards and Earth System Sciences*: Time of Emergence of compound events: contribution of univariate and dependence properties

Time of Emergence of compound events: contribution of univariate and dependence properties

Bastien François[1] and Mathieu Vrac[1]

[1]Laboratoire des Sciences du Climat et l'Environnement (LSCE-IPSL) CNRS/CEA/UVSQ, UMR8212, Université Paris-Saclay, Gif-sur-Yvette, France

Correspondence: B. François (bastien.francois@lsce.ipsl.fr)

Abstract.

Many climate-related disasters often result from a combination of several climate phenomena, also referred to as "compound events" (CEs). By interacting with each other, these phenomena can lead to huge environmental and societal impacts, at a scale potentially far greater than any of these climate events could have caused separately. Marginal and dependence properties of the climate phenomena forming the CEs are key statistical properties characterising their probabilities of occurrence. In this study, we propose a new methodology to assess the time of emergence of compound events probabilities, which is critical for mitigation strategies and adaptation planning. Using copula theory, we separate and quantify the contribution of marginal and dependence properties to the overall probability changes of multivariate hazards leading to compound events. It provides a better understanding of how the statistical properties of variables leading to CEs evolve and contribute to the change of their occurrences. For illustration purposes, the methodology is applied over a 13-member multi-model ensemble (CMIP6) to two case studies: compound wind and precipitation extremes over the region of Brittany (France), and frost events occurring during the growing season preconditioned by warm temperatures (growing-period frost) over Central France. For compound wind and precipitation extremes, results show that probabilities emerge before the end of the 21st century for 6 models of the considered CMIP6 ensemble. For growing-period frosts, significant changes of probability are detected for 11 models. Yet, the contribution of marginal and dependence properties to these changes of probabilities can be very different from a climate hazard to another, and from one model to another. Depending on the CE, some models give a strong importance to both marginal properties and dependence properties for probability changes. These results highlight the importance of considering both marginal and dependence properties changes, as well as their inter-model variability, for future risk assessments due to compound events.

Chapter 4. Time of Emergence of compound events: contribution of univariate and dependence properties

1 Introduction

In September 2017, heavy rainfall and storm surge associated with Hurricane Irma resulted in record-breaking floods in Jacksonville, Florida. In 2019, Australia had experienced high temperatures and prolonged dry conditions, which resulted in one of the worst bush fire seasons in its recorded history. In April 2021 and 2022, Central Europe experienced consecutive days of frost events following a warm early spring, which caused severe damages to agricultural yields. These recent climate events are some examples of so-called "compound events" (CEs), i.e., high-impact climate events that result from interactions of several climate hazards. These climate hazards are not necessarily extremes themselves, but their simultaneous or successive occurrences can generate strong impacts (Leonard et al., 2014; Zscheischler et al., 2014, 2018, 2020). Though still in its infancy, the understanding of the complex nature of compound events and the assessment of their associated risks have been the subject of numerous research studies in climate sciences (e.g., Bevacqua et al., 2017, 2021; Manning et al., 2018; Zscheischler and Seneviratne, 2017; Ridder et al., 2021, 2022; Singh et al., 2021a; Nasr et al., 2021; Raymond et al., 2022, among many others). Recently, a typology of compound events has been proposed in order to categorise them into four classes depending on how individual hazards interact to form the CEs ("preconditioned", "multivariate", "temporally compounding" and "spatially compounding" events, see Zscheischler et al., 2020). Concerning projected changes, frequency and intensity of some compound events such as co-occurring heatwaves and droughts are expected to increase for many regions of the world, even when considering climate change scenarios with limited global warming to 1.5°C above pre-industrial levels (IPCC, 2021). Determining whether probabilities of compounding climate events present significant changes between past and future periods, and to detect when these significant changes occur are of paramount importance, not only for mitigation and adaptation issues but also to inform the general public and to raise awareness of climate change. Only when the changes of probability are of sufficient magnitude relative to a baseline period can we be confident that significant changes have been detected. Detecting from which period the changes are statistically significant corresponds to the concept of "Time of Emergence" (ToE). It consists in determining the time or period in which a climate signal emerges from (i.e., goes out of) the natural variability (e.g., Christensen et al., 2007; Maraun, 2013; Hawkins et al., 2020; Ossó et al., 2022). Time of Emergence has been discussed extensively to analyse the emergence of mean temperatures (e.g., Hawkins and Sutton, 2012; Mahlstein et al., 2011), precipitation (Fischer et al., 2014; Giorgi and Bi, 2009; Gaetani et al., 2020), but also emergence of extremes (e.g. Diffenbaugh and Scherer, 2011; Fischer et al., 2014; King et al., 2015). Evaluating the ToE of compound hazards probabilities with respect to a baseline period — from which the natural variability is estimated — is valuable to analyse evolutions of compound events and attribute those to a specific cause, such as anthropogenic greenhouse gas emissions. Attribution is an important research field in climate science that aims at determining the mechanisms responsible for recent global warming and related climate changes. For example, it can be done by comparing probabilities of an event between two worlds with different forcings (the "risk-based" approach, Stott et al., 2004; Shepherd, 2016). Generally, a factual world with anthropogenic climate change and a counterfactual world in which anthropogenic emissions had never occurred are considered. Although we do not aim at performing attribution per se in the present study, the underlying philosophy is relatively similar for ToE: by considering a pre-industrial period as baseline,

compound hazards probabilities associated with natural forcings — or natural variability — could be estimated, and so the influence of future climate change on probabilities.

From a statistical point-of-view, compound events are characterised by the statistical features of the variables forming the CEs, i.e., their marginal properties (e.g., mean and variance) and dependence structures. These key statistical properties can be affected by future climate change (e.g., Wahl et al., 2015; Schär, 2015; Russo et al., 2017; Raymond et al., 2020; Jézéquel et al., 2020). In addition to potentially exacerbate impacts, these evolutions of marginal and dependence properties could also combine to change the probabilities of the CEs' hazards (e.g., Rana et al., 2017; Zscheischler and Seneviratne, 2017; Zscheischler and Lehner, 2021; Manning et al., 2019; Singh et al., 2021a). For example, rising temperatures can naturally lead to more co-occurrences of hot temperatures and droughts, despite no significant trends in droughts are detected (Diffenbaugh et al., 2015; Mazdiyasni and AghaKouchak, 2015). However, in addition to warmer temperatures, the strengthening of the dependence between hot temperatures and droughts for future periods can also contribute to an increase in their co-occurrences (as highlighted in Zscheischler and Seneviratne, 2017). Several studies concluded about the importance of considering dependencies to assess CE properties and frequencies in a robust way (e.g., Hillier et al., 2020; Singh et al., 2021a; Vrac et al., 2021). Recently, Abatzoglou et al. (2020) even showed, using reanalysis data, that changes in dependence properties have been more important than changes in univariate properties in the recent decades. Hence, to determine the ToE of hazards probabilities, quantifying the influence (or contribution) of the statistical features of the variables forming the CEs to these changes of probabilities is thus crucial to further understand the potential future evolutions of compound events (Vrac et al., 2021).

In this paper, we propose a new methodology to assess the time of emergence of compound events probabilities. We also develop a copula-based multivariate framework, which allows for an adequate description of the contribution of the marginal and dependence properties changes to the evolutions of multivariate hazard probabilities. This compound event analysis is applied to two case studies. We first analyse compound wind and precipitation extremes over the coastal region of Brittany (France). This bivariate compound event, i.e., composed of co-occurring climate hazards over the same region and time, has been analysed in several studies (e.g., Martius et al., 2016; Bevacqua et al., 2019; De Luca et al., 2020a; Reinert et al., 2021; Messmer and Simmonds, 2021) as it can have severe impacts such as important economic losses, massive damages to infrastructure and loss of human life (e.g., Fink et al., 2009; Liberato, 2014; Wahl et al., 2015; Raveh-Rubin and Wernli, 2015). We then apply our methodology to a second climate hazard: frost events occurring during the growing season preconditioned by warm temperatures (growing-period frost) over Central France. When occurring after bud burst, i.e., when the sensitive emerging leaves and flowers have started to develop, frost temperatures potentially affect growth and distribution limits of plants. It can consequently cause important economic losses to agriculture (Lamichhane, 2021). These growing-period frost events and their associated risks in past and future periods have been studied in the literature (e.g., Unterberger et al., 2018; Liu et al., 2018a; Sgubin et al., 2018; Pfleiderer et al., 2019), as well as the role of human-caused climate change on growing-period frosts probability (Vautard et al., 2021).

The rest of this paper is organised as follows: Sect. 2 describes the climate simulations used in this study, and Sect. 3 details the statistical method and experimental setup used to analyse time of emergence of compound events probabilities and contributions of the statistical features. Then, results for the analysis of the two climate compound hazards are provided in

Sect. 4 for extremes of wind and precipitation and in Sect. 5 for growing-period frost events. Conclusions, discussions and perspectives for future research are finally proposed in Sect. 6.

2 Model data

One ensemble of 13 Global Climate Models (GCMs) following the CMIP6 protocol (Eyring et al., 2016) is considered. This selection of models is listed in Table 1. To define compound wind and precipitation extremes, we use daily precipitation and wind speed maxima variables. For growing-period frosts, mean and minimum temperature variables are used. For each variable, the historical period simulations (1871-2014) have been extracted and extended until 2100 using the shared socioeconomic pathways 585 (SSP-585) scenario (Riahi et al., 2017). As the 13 selected simulations present different spatial resolutions, each climate simulation dataset has been regridded to a common spatial resolution of $0.5° \times 0.5°$ using bilinear interpolation. Considering the climate models separately will allow us to assess inter-model variability in terms of time of emergence of compound events probabilities, as well as the potentially different contributions of marginal and dependence properties to changes in probability of multivariate climate hazards. Also, by considering all climate models together using a pooling procedure, a multi-model ensemble estimate for ToE and contributions could be derived. Pooling the models together will allow us to better take into account the global uncertainty inherent in climate modelling and to reduce the influence of natural variability amongst individual ensemble members.

For compounding wind and precipitation extremes, we use spatial mean of daily wind speed maxima and spatial sum of daily precipitation time series during winter (December, January and February) over the region of Brittany, France ([-5, -2°E] × [46.5, 49°N], see Fig. 1a), which corresponds to a domain with 21 continental grid cells in our regridded climate simulations. This coastal region is regularly impacted by mid-latitude extra-tropical storms causing large damages to infrastructures (e.g., the storm Xynthia in 2010). Analysing the evolution of probability of compound wind and precipitations extremes is therefore relevant for this region. To allow for a robust statistical modelling of compounding wind and precipitation extremes, we applied our methodology to bivariate points of high values by selecting wind and precipitation data concurrently exceeding selected high thresholds. Indeed, our methodology detailed later in Sect. 3 is based on the use of parametric models and considering the complete bivariate distribution to fit marginals and copulas could be not appropriate as the representation of the extremes would be biased by the bulk of the bivariate distributions where most of the data is located (e.g., Bevacqua et al., 2019). More details on selection thresholds will be provided later in Sect. 4.

For growing-period frost events, data are extracted over Central France ([-1, 5°E] × [46, 49°N], see Fig. 1a), which corresponds to 78 continental grid cells. The region covers an important agriculture area of France, including grapevine and fruit crops with high production (Vautard et al., 2021). We focus on spatial mean of daily minimum temperature (T) in April to define frost events occurring in early spring. To account for phenology and characterise bud burst conditions by the end of March, the Growing Degree Day (GDD) model (Bonhomme, 2000) is used. The GDD model consists in computing cumulative daily mean temperatures minus a "base temperature" from a starting date. For our study, a base temperature of 5°C is used and the starting date for computing GDD values for each year is chosen to be 1 January. In this study, our aim is not to focus on the

phenology of specific plants but rather to provide a general overview of growing-period frost events. 5°C as base temperature is generally accepted for crops and grapevine (e.g., Skaugen and Tveito, 2004; Jiang et al., 2011; Ruosteenoja et al., 2016; Vautard et al., 2021). Bud burst occurs when the cumulative sum of degree-days up to 31 March is larger than some thresholds (Garcia de Cortazar-Atauri et al., 2009), which depend on species. For each year y, GDD values by the end of March are obtained via the formula:

$$GDD(y) := \sum_{i=y/01/01}^{i=y/03/31} \max(MT(i) - 5, 0),$$

with MT the daily mean temperature. GDD values are computed for each grid cell and averaged spatially over the area of Central France. We consider the threshold of 200°C.day to characterise bud burst conditions and illustrate our method. The choice of this threshold is consistent with existing studies analysing bud burst values of grapevine species (e.g., Garcia de Cortazar-Atauri et al., 2009; Vautard et al., 2021), and is useful to characterise early bud burst plants that could be impacted by frost events.

For illustration purpose, Fig. 1a displays the topographic map of France with the region of Brittany and Central France in boxes. The bivariate wind and precipitation data (Fig. 1b) and miniminal temperature and GDD data (Fig. 1c) for the CNRM-CM6 model are also displayed.

3 Statistical method

Our aim is to design a statistical method to assess the time of emergence of compound events probabilities, that is, to detect from which period changes of probability are statistically significant relative to a baseline period. Probabilities of compound events can be computed with copulas. Copulas are functions that allow to describe the dependence structure between random variables separately from their marginal distributions and greatly simplifies calculations involving multivariate distributions (Nelsen, 2006). Copulas have been widely applied in climate and geophysical science (e.g., Vrac et al., 2005; Salvadori et al., 2007; Schölzel and Friederichs, 2008; Serinaldi, 2014). In addition to allowing computations of multivariate hazards probabilities, the use of copulas in our study permits to isolate and quantify the marginal and dependence contributions of the variables forming the CEs to the overall probability changes. In the following, we first remind the concept of ToE, and then present our methodology to assess the time of emergence of compound events probabilities. Then, after some reminders about the copula theory, the methodology to assess the contribution of marginal and dependence properties to changes of probabilities is presented. For ease of presentation, the methodology is explained for compounding wind and precipitation extremes but will be applied similarly for growing-period frosts.

3.1 Time of emergence of climate hazards

The concept of Time of Emergence (ToE) has been developed to assess the significance of climate changes relative to background variability. Comparing changes of climate signal relative to the natural variability is particularly relevant as human

150 societies and ecosystems are inherently adapted to the local background level of variability, and major impacts arise most likely when changes emerge from it (e.g., Lobell and Burke, 2008). Different methodologies to assess ToE of climate signals have been used in the literature. For example, ToE can be assessed by estimating the climate change signal (S) and the variability (or noise, N) of the climate metric of interest (e.g., Hawkins and Sutton, 2012; Maraun, 2013; Hawkins et al., 2020; Ossó et al., 2022). The ToE is then estimated by determining the first period for which the S/N ratio permanently crosses a certain 155 threshold (e.g., emergence of "unusual" (S/N > 1), "unfamiliar" (S/N > 2), or "unknown" (S/N > 3) climates, Frame et al., 2017). Methodologies for ToE based on statistical tests have also been developed, which estimate the first period for which the distribution of the climate metric is significantly and permanently different from a baseline period distribution (e.g. using Kolmogorov-Smirnov tests, Mahlstein et al., 2012; Gaetani et al., 2020; Pohl et al., 2020). To define emergence of compound events probabilities, we propose to assess probabilities in a 30-year window sliding over the period 1871-2100 and compare 160 their values with respect to a baseline period's probability. In this study, we consider the reference period (1871-1900) as baseline to assess the emergence of hazard probabilities. However, there is no agreement on the choice of the baseline period for ToE studies. While most of the studies choose a pre-industrial period as baseline to attribute emergence to anthropogenic greenhouse gas forcing (e.g., 1850-1900, Hawkins et al., 2020), other studies choose a more recent baseline period (e.g., 1951-1983, Ossó et al., 2022), which can provide relevant information for adaptation planning. We further discuss the choice of the 165 reference period for emergence in Sect. 6. The ToE of hazard probabilities is then the time period when a significant change of probability occurs relatively to the probability associated with the estimated natural variability, and persists until the end of the century. To assess if probabilities are significantly different from that of the background variability, we propose to compute the 68% and 95% confidence intervals of the baseline period's probability. It permits to characterise the natural variability of our probability of interest. An emergence is detected if probability for the 30-year sliding windows permanently go out of the 170 baseline confidence intervals (i.e., out of the estimated natural variability). The ToE is then defined as the central year of the sliding window over which the probability starts to emerge. As probabilities are estimated using copula modelling (see later in subsection 3.2), 68% and 95% confidence intervals of baseline period's probabilities are computed by coupling the parameters uncertainties of both the fitted marginal distributions and the fitted copula. Considering both 68% and 95% confidence intervals allows to evaluate, with different degrees of confidence, the changes of probability of compounding events from the estimated 175 natural variability. Details on the procedure to compute confidence intervals are given in Appendix A.

3.2 A reminder on copula functions and exceedance probability

In this study, we use copula modelling to compute compound events probabilities. We first consider two random variables X (e.g., maximum wind speed) and Y (e.g., precipitation) for an arbitrary period. We denote their marginal (i.e., univariate) probability density functions (pdfs) $f_X(x)$ and $f_Y(y)$ and cumulative marginal distribution functions (CDFs) $F_X(x) = \mathbb{P}(X \leq$ 180 $x)$ and $F_Y(y) = \mathbb{P}(Y \leq y)$. Sklar's theorem (Sklar, 1959) states that, H, the joint (i.e., bivariate) CDF can be written as:

$$H_{X,Y}(x,y) = \mathbb{P}(X \leq x \cap Y \leq y) = C(F_X(x), F_Y(y)), \tag{1}$$

where C is a function called "copula", corresponding to the joint distribution function of the uniformly distributed variables $F_X(X)$ and $F_Y(Y)$. Under the assumption that the marginal distributions F_X and F_Y are continuous, Sklar's theorem states that the copula C is unique. This decomposition of the multivariate distribution into marginals distributions and copula function allows us to model the dependence among contributing variables independently of their marginals. Therefore, using copulas makes it easy to isolate the effects of marginal and dependence properties on probability of multivariate hazards.

Bivariate exceedance probability refers to the probability that both random variables exceed a certain value ("AND approach", Salvadori et al., 2016) and can be calculated relatively easily using copulas. For example, for wind and precipitation compound events, it corresponds to probabilities of wind speed and precipitation jointly exceeding established thresholds . We denote $p_{m,d}$ the bivariate exceedance probability computed with marginal (subscript m) and dependence (subscript d) properties of (X,Y). The probability $p_{m,d}(t_X,t_Y)$ that both X and Y jointly exceed some predefined thresholds t_X and t_Y is given by (Yue and Rasmussen, 2002; Shiau, 2003):

$$\begin{aligned} p_{m,d}(t_X,t_Y) &= \mathbb{P}(X \geq t_X \cap Y \geq t_Y) \\ &= 1 - F_X(t_X) - F_Y(t_Y) + C(F_X(t_X), F_Y(t_Y)). \end{aligned} \tag{2}$$

Marginal and copula distributions in Eq. (2) are estimated using parametric fitting procedures. More details on the fitting procedures for compound wind and precipitation extreme and growing-period frost events are given in Appendix B.

3.3 Change of probabilities: contribution of the marginal and dependence properties

Let us now consider the realizations $(X_{\text{ref}}, Y_{\text{ref}})$ and $(X_{\text{fut}}, Y_{\text{fut}})$ of the two random variables X and Y over the reference period (i.e., 1871-1900 in the following), and over another 30-year period (e.g. a future period such as 2071-2100). Using Eq. (2), the reference and future bivariate exceedance probability $p_{m_{\text{ref}},d_{\text{ref}}}(t_X,t_Y)$ and $p_{m_{\text{fut}},d_{\text{fut}}}(t_X,t_Y)$ for some predefined thresholds t_X and t_Y are given by:

$$p_{m_{\text{ref}},d_{\text{ref}}}(t_X,t_Y) = 1 - F_{X_{\text{ref}}}(t_X) - F_{Y_{\text{ref}}}(t_Y) + C_{\text{ref}}(F_{X_{\text{ref}}}(t_X), F_{Y_{\text{ref}}}(t_Y)), \tag{3}$$

$$p_{m_{\text{fut}},d_{\text{fut}}}(t_X,t_Y) = 1 - F_{X_{\text{fut}}}(t_X) - F_{Y_{\text{fut}}}(t_Y) + C_{\text{fut}}(F_{X_{\text{fut}}}(t_X), F_{Y_{\text{fut}}}(t_Y)). \tag{4}$$

As modeled here with Eqs. (3) and (4), $p_{m_{\text{fut}},d_{\text{fut}}}$ and $p_{m_{\text{ref}},d_{\text{ref}}}$ can differ due to:

- changes in the marginal properties of X and Y, i.e., changes between $F_{X_{\text{ref}}}$ and $F_{X_{\text{fut}}}$, as well as between $F_{Y_{\text{ref}}}$ and $F_{Y_{\text{fut}}}$,
- and changes in the dependence structure (i.e., in the copulas) between X and Y, i.e., changes between C_{ref} and C_{fut}.

Then, do exceedance probability values change significantly between reference and future periods? And if so, how much of this change is due to changing marginal properties? To changing dependence structure? In order to isolate the effects of

these potentially changing statistical properties, we propose to calculate two additional exceedance probability values. The first one is the probability $p_{m_{\text{fut}},d_{\text{ref}}}$, which assesses what the future probability would be if only the marginal properties change between the reference and future period (and thus keeping the dependence properties from the reference period). $p_{m_{\text{fut}},d_{\text{ref}}}$ is hence computed as:

$$p_{m_{\text{fut}},d_{\text{ref}}}(t_X,t_Y) = 1 - F_{X_{\text{fut}}}(t_X) - F_{Y_{\text{fut}}}(t_Y) + C_{\text{ref}}(F_{X_{\text{fut}}}(t_X), F_{Y_{\text{fut}}}(t_Y)). \tag{5}$$

Inversely, the second additional probability $p_{m_{\text{ref}},d_{\text{fut}}}$ is aimed to assess what the future probability would be if only the dependence properties change between the reference and future period (keeping the marginal properties from the reference period), and is computed as:

$$p_{m_{\text{ref}},d_{\text{fut}}}(t_X,t_Y) = 1 - F_{X_{\text{ref}}}(t_X) - F_{Y_{\text{ref}}}(t_Y) + C_{\text{fut}}(F_{X_{\text{ref}}}(t_X), F_{Y_{\text{ref}}}(t_Y)). \tag{6}$$

Illustrations of these four probabilities for artificial bivariate distributions and changes between a reference and a future period are given in Fig. 2.

To assess how much marginal and dependence contribute to exceedance probabilities change between reference and future period, we use the four probabilities derived above to decompose the overall probability change. We first define ΔP, the change of probability between the reference and future periods, as the difference between the two probabilities: $\Delta\text{P} = p_{m_{\text{fut}},d_{\text{fut}}} - p_{m_{\text{ref}},d_{\text{ref}}}$. By computing $p_{m_{\text{fut}},d_{\text{ref}}}$ and $p_{m_{\text{ref}},d_{\text{fut}}}$, one can decompose the change of probability ΔP into a sum of three terms that can yield statistical interpretations:

$$\Delta\text{P} = \Delta\text{M} + \Delta\text{D} + \Delta\text{I}. \tag{7}$$

The first term ΔM accounts for the difference of probability between the reference and future periods due to a change of marginal properties only and is hence called the "marginal" term:

$$\Delta\text{M} = p_{m_{\text{fut}},d_{\text{ref}}} - p_{m_{\text{ref}},d_{\text{ref}}}$$

Similarly, the second term ΔD assesses the difference of probability between the reference and future periods due to a change of dependence properties only and is hence called the "dependence" term:

$$\Delta\text{D} = p_{m_{\text{ref}},d_{\text{fut}}} - p_{m_{\text{ref}},d_{\text{ref}}}$$

As simultaneous changes of marginal and dependence properties between the reference and future period can affect the exceedance probability in a highly non-linear fashion (as it can be observed visually in Fig. 2), ΔP cannot be simply expressed as the sum of the differences ΔM and ΔD. Thus, a residual term ΔI, called the "interaction" term, is introduced to assess the part of the probability change that is due to the simultaneous change of marginal and dependence properties and that cannot be explained by the changes of these statistical properties separately:

$$\Delta\text{I} = p_{m_{\text{fut}},d_{\text{fut}}} - p_{m_{\text{fut}},d_{\text{ref}}} - p_{m_{\text{ref}},d_{\text{fut}}} + p_{m_{\text{ref}},d_{\text{ref}}}.$$

The decomposition of $\Delta\mathrm{P}$ into these three terms allows to isolate the effects of the changes of marginal, the effects of the changes of dependence properties and the effects of the changes of interaction on the overall change of probability value $\Delta\mathrm{P}$. By taking advantage of this decomposition, we propose to quantify the contribution (in $\%$) of the different terms $\Delta\mathrm{M}$, $\Delta\mathrm{D}$ and $\Delta\mathrm{I}$ to the change of probability $\Delta\mathrm{P}$. For example, the contribution of the changes of the marginal properties can be quantified as:

$$\mathrm{Contrib}_{\Delta\mathrm{M}} = \frac{\Delta\mathrm{M}}{\Delta\mathrm{P}} \times 100, = \frac{p_{m_{\text{fut}},d_{\text{ref}}} - p_{m_{\text{ref}},d_{\text{ref}}}}{p_{m_{\text{fut}},d_{\text{fut}}} - p_{m_{\text{ref}},d_{\text{ref}}}} \times 100. \tag{8}$$

A value of 50 % for $\mathrm{Contrib}_{\Delta\mathrm{M}}$ would indicate that the change of marginal properties is responsible for 50 % of the global change of probability $\Delta\mathrm{P}$ between the reference and future periods. The contributions of $\Delta\mathrm{D}$ (resp. $\Delta\mathrm{I}$) can be calculated the same way by simply replacing $\Delta\mathrm{M}$ in Eq. (8) by $\Delta\mathrm{D}$ (resp. $\Delta\mathrm{I}$). The sum of the three contributions adds up to 100 %, by construction. Please note that, for illustration, changes of probability $\Delta\mathrm{P}$, $\Delta\mathrm{M}$ and $\Delta\mathrm{D}$ are here considered as differences of probabilities. One could also consider analysing other metrics such as relative differences ("r. diff") by dividing each of the terms in Eq. (7) by $p_{m_{\text{ref}},d_{\text{ref}}}$:

$$\begin{aligned}
\Delta\mathrm{P}^{\text{r. diff}} &= \frac{p_{m_{\text{fut}},d_{\text{fut}}} - p_{m_{\text{ref}},d_{\text{ref}}}}{p_{m_{\text{ref}},d_{\text{ref}}}}, \\
\Delta\mathrm{M}^{\text{r. diff}} &= \frac{p_{m_{\text{fut}},d_{\text{ref}}} - p_{m_{\text{ref}},d_{\text{ref}}}}{p_{m_{\text{ref}},d_{\text{ref}}}}, \\
\Delta\mathrm{D}^{\text{r. diff}} &= \frac{p_{m_{\text{ref}},d_{\text{fut}}} - p_{m_{\text{ref}},d_{\text{ref}}}}{p_{m_{\text{ref}},d_{\text{ref}}}}, \\
\Delta\mathrm{I}^{\text{r. diff}} &= \frac{p_{m_{\text{fut}},d_{\text{fut}}} - p_{m_{\text{fut}},d_{\text{ref}}} - p_{m_{\text{ref}},d_{\text{fut}}} + p_{m_{\text{ref}},d_{\text{ref}}}}{p_{m_{\text{ref}},d_{\text{ref}}}}.
\end{aligned}$$

In addition, bivariate fraction of attributable risk ("FAR", e.g., Stott et al., 2016; Chiang et al., 2021; Zscheischler and Lehner, 2021) can also be computed by dividing each of the term by $p_{m_{\text{fut}},d_{\text{fut}}}$:

$$\begin{aligned}
\Delta\mathrm{P}^{\text{FAR}} &= \frac{p_{m_{\text{fut}},d_{\text{fut}}} - p_{m_{\text{ref}},d_{\text{ref}}}}{p_{m_{\text{fut}},d_{\text{fut}}}}, \\
\Delta\mathrm{M}^{\text{FAR}} &= \frac{p_{m_{\text{fut}},d_{\text{ref}}} - p_{m_{\text{ref}},d_{\text{ref}}}}{p_{m_{\text{fut}},d_{\text{fut}}}}, \\
\Delta\mathrm{D}^{\text{FAR}} &= \frac{p_{m_{\text{ref}},d_{\text{fut}}} - p_{m_{\text{ref}},d_{\text{ref}}}}{p_{m_{\text{fut}},d_{\text{fut}}}}, \\
\Delta\mathrm{I}^{\text{FAR}} &= \frac{p_{m_{\text{fut}},d_{\text{fut}}} - p_{m_{\text{fut}},d_{\text{ref}}} - p_{m_{\text{ref}},d_{\text{fut}}} + p_{m_{\text{ref}},d_{\text{ref}}}}{p_{m_{\text{fut}},d_{\text{fut}}}}.
\end{aligned}$$

However, by construction, results for contributions, either for relative differences or bivariate FAR, would be identical to those obtained for differences.

3.4 Application to the multi-model ensemble

The methodology described above to assess time of emergence of compound events probabilities and marginal and dependence contributions to these changes is applied to the 13 CMIP6 models by considering successively all 30-year sliding windows spanning the period 1871-2100. Moreover, the methodology is applied to the ensemble in two different versions:

- the "Indiv-Ensemble" version, for which the methodology is applied to each climate model individually. In particular for contributions and ToE, multi-model median estimates are derived to summarise the information given by all the models.
- the "Full-Ensemble" version, which consists of pooling the contributing variables of the 13 climate models together and applying the methodology to these pooled data to derive a pooled estimate of time of emergence, as well as marginal and dependence contributions.

Depending on the versions, the objectives are not the same: whereas the Indiv-Ensemble version permits to analyse the modelling of hazards separately and assess the uncertainty in ToE arising from the inter-model differences, the Full-Ensemble version permits to derive unique ToE estimates and contribution values accounting for the global uncertainty in climate modelling. This Full-Ensemble version assumes that the variables of interest are drawn from the same distribution.

Concerning the Full-Ensemble version, a post-processing step of the different models is required for the analysis of compound wind and precipitation extremes only. Indeed, as already explained in Section 2, wind and precipitation data concurrently exceeding high selection thresholds are selected for each climate model in order to focus on compounding extremes. However, climate models can present very different values of wind and precipitation data: for example, a model may not be capable of simulating wind and precipitation events as intense as other models. Hence, each model potentially has different selection thresholds over which values of wind and precipitation are selected. Because of this, selected compound wind and precipitation data from the different climate models cannot be directly pooled, and data need first to be transformed to apply our methodology and analyse pooled extreme events. The transformation step is reached by using a univariate quantile mapping technique (CDF-t, Vrac et al., 2012) that makes the univariate distributions of the wind and precipitation extremes similar to those from a model of reference without modifying their dependence structure. In the following, we choose the CNRM-CM6 model as reference. As values of wind and precipitation extremes of the different models will be modified on purpose by the CDF-t method, note that exceedance thresholds in terms of probabilities (instead of physical values) will be considered. This way, it will enable an interpretation of the results from the Full-Ensemble version. More details about the application of the CDF-t method to transform compound wind and precipitation data for the Full-Ensemble version can be found in Appendix C.

To analyse growing-period frost events with the Full-Ensemble version, no transformation step is needed before pooling. Indeed, contrary to wind and precipitation extreme, the definition of growing-period frost events does not depend on climate models and can be based on well-established thresholds. A summary of the successive steps of our methodology for the Indiv- and Full-Ensemble versions is provided in the form of a flowchart in Fig. 3.

4 Results for compounding wind and precipitation extremes

In this section, results are presented for compound wind and precipitation extremes during winter in Brittany. Please note that, for this section as well as for the rest of the study, the period 1871-1900 is considered as the baseline period for natural variability to evaluate time of emergence and contributions. To focus on wind and precipitation extremes, we applied our methodology to points of high values. For each model, we selected points where, concurrently, wind and precipitation values exceed the individual 90th percentiles (denoted x_{sel} and y_{sel}, respectively) of the 1871-1900 reference period. In the following, we denote $S^{i}_{90,90}$ the ensemble of the selected points of high values for a model i. For illustration purpose, the ensemble $S^{\text{CNRM-CM6}}_{90,90}$ for the CNRM-CM6 model is shown in orange in Fig 1b. We first illustrate our method with a single climate model (CNRM-CM6). Then, results obtained for the Indiv- and Full-Ensemble versions are presented.

4.1 Results for an individual model and a single exceeding threshold: CNRM-CM6

To illustrate our methodology, we first explain the results obtained for compound wind and precipitation extremes and a single bivariate exceeding threshold before extending the results to several bivariate thresholds. We evaluate the probabilities of exceeding the 80th percentiles of the bivariate points belonging to $S^{\text{CNRM-CM6}}_{90,90}$. The 80th percentiles for wind and precipitation correspond to $x_{80\text{lsel}} \approx 17.8$ m/s and $y_{80\text{lsel}} \approx 338$ mm/d, respectively.

Before computing any probability, Fig. 4 gives a first overview of the fitted bivariate distributions of compound wind and precipitation extremes in our study. It displays the evolutions of the bivariate distributions over a selection of sliding windows due to changing marginal and dependence properties ("Marg.-dep.", Fig. 4a), changing marginal properties only ("Marg.", Fig. 4b) and changing dependence only ("Dep.", Fig. 4c). Plotting these bivariate distributions already indicates the changes in probability of wind and precipitation extremes, and the potential influences of marginal and dependence properties on these changes. Indeed, at first sight in Fig. 4a, the area of bivariate distributions where wind speed and precipitation jointly exceed $x_{80\text{lsel}}$ and $y_{80\text{lsel}}$ appears to increase for future periods, suggesting that such bivariate events are more likely to occur according to CNRM-CM6 projections. But is this change of probability significant? And is this change due to marginal properties changes? Dependence properties changes? Or both? By keeping the dependence properties of the reference period and considering changing marginal properties only (Fig. 4b), an increase of exceedance probability seems to be observed, although less pronounced. Similar observations can be made by keeping the marginal properties of the reference period and considering changing dependence properties only (Fig. 4c). If both marginal and dependence changes seem to have an importance in the increase of probability, it is important to quantify how much these statistical properties contribute to the change of the overall probability, as well as their respective influence on the time of emergence of probabilities of compounding wind and precipitation extremes.

Time series of exceedance probabilities over all sliding windows for the bivariate threshold ($x_{80\text{lsel}}$,$y_{80\text{lsel}}$) are presented in Fig. 5 by considering changes of marginal and dependence properties together (Fig. 5a) and separately (Figs. 5b and c). 68% and 95% confidence intervals resulting from marginal and copula uncertainties are also displayed for each probability. All three time series present an increase with time, which is consistent with the visual analysis made in Fig. 4. Probability increase is less

pronounced when future marginal (Fig. 5b) and future dependence properties (Fig. 5c) are considered separately. It illustrates that the effects of these changing statistical properties combine on exceedance probabilities. Yet, all three probability signals permanently go out of the reference natural variability confidence intervals, suggesting that an emergence of probability occurs: for probabilities computed with future marginal and dependence properties (Fig. 5a), the time of emergence is detected in 2009 (1994-2023) and 2072 (2057-2086) for 68 % and 95% confidence levels, respectively. Concerning probabilities influenced by future marginal changes and future dependence changes separately (Figs. 5b and c), probability signals emerge later at the 68 % confidence level, in 2073 (2058-2087) and 2063 (2048-2077), respectively. If contributions of the statistical properties to time of emergence in itself are not computed here, one can get an idea of the importance of the statistical properties on ToE: at the 68% confidence level, ignoring the dependence change would induce a ToE $2073 - 2009 = 64$ years later. Similarly, ignoring marginal changes would induce a ToE $2063 - 2009 = 54$ years later. It thus indicates that both marginal and dependence properties have a non-negligible effect on time of emergence.

Evolution of the bivariate FAR $\Delta \mathrm{P}^{\mathrm{FAR}}$ with respect to the reference period over sliding windows, as well as its decomposition in terms of "marginal" ($\Delta \mathrm{M}^{\mathrm{FAR}}$), "dependence" ($\Delta \mathrm{D}^{\mathrm{FAR}}$) and "interaction" ($\Delta \mathrm{I}^{\mathrm{FAR}}$) terms are displayed in Fig. 5d. As explained in Sect. 3, for each sliding window, the sum of $\Delta \mathrm{M}^{\mathrm{FAR}}$, $\Delta \mathrm{D}^{\mathrm{FAR}}$ and $\Delta \mathrm{I}^{\mathrm{FAR}}$ is by construction equal to $\Delta \mathrm{P}^{\mathrm{FAR}}$. The decomposition highlights that the influences of the marginal and of the dependence properties on bivariate FAR can vary with time. Also, the combination of individual effects of marginal and dependence changes on the overall probability changes is again illustrated: for example, by 2100, considering both future marginal and dependence changes leads to a value of FAR $\Delta \mathrm{P}^{\mathrm{FAR}}$ twice as high as those of $\Delta \mathrm{M}^{\mathrm{FAR}}$ and $\Delta \mathrm{D}^{\mathrm{FAR}}$, respectively. Concerning the interaction term, its associated bivariate FAR is negligible, highlighting that most of the changes can be explained by the changing marginal and dependence properties separately. Results for relative differences are displayed in Fig. 5e, and same conclusions can be drawn. Fig. 5f shows the evolution of the contributions from the marginal, dependence and interaction terms to probability values over sliding windows. By computing the median of contributions over all sliding windows, we can see that both changes in the marginal and in the dependence properties contribute greatly to probability changes ($\approx 50\%$) in the CNRM-CM6 simulations, with a slightly more important contribution from dependence properties (dashed lines in figure 5f). One could remark a symmetry between the contribution values of the marginal and the dependence terms over sliding windows. This can be explained by the way contribution values are computed. Indeed, as the sum of the three contributions adds up to 100 %, by construction, and that the contribution from the interaction term is close to 0, contribution values of the marginal and the dependence terms covary symmetrically around 50%.

4.2 Results for CNRM-CM6 and several exceeding thresholds

Until now, results for ToE and contributions have been presented for the probability of events exceeding the 80th percentiles of selected points belonging to $S_{90,90}^{\text{CNRM-CM6}}$. In order to have a broader analysis of exceedance probabilities of compound wind and precipitation extremes, we repeat the methodology for all pairs of exceedance thresholds between the 5th and 95th percentiles (with steps of 5 percentiles) of selected points belonging to $S_{90,90}^{\text{CNRM-CM6}}$. Fig. 6 displays the results obtained for the CNRM-CM6 time of emergence at the 68% confidence level, by considering marginal and dependence changes (Fig. 6a), marginal changes

only (Fig. 6b) and dependence changes only (Fig. 6c). Moreover, for each bivariate exceedance threshold, median contributions (over all sliding windows) of marginal (Fig. 6d), dependence (Fig. 6e) and interaction terms (Fig. 6f) are displayed. Results for ToE obtained at 95 % confidence level are displayed in Fig. S1 and differences of ToE are displayed in Fig. S2 of the Supplement. When varying exceedance thresholds, different ToE results are obtained, depending on whether marginal and dependence changes are considered (Figs. 6a, b and c). ToE are found for most of the exceedance thresholds when considering both marginal and dependence changes (Fig. 6a) or marginal changes only (Fig. 6b). It is however not the case for dependence changes only (Fig. 6c), for which only specific pairs of exceedance thresholds can find times of emergence. Interestingly, these pairs correspond to very high compound wind and precipitation extremes. It indicates that dependence change plays an important role for the probability of such high extreme events. The importance of dependence properties can also be assessed visually by comparing Figs. 6a and b. Indeed, for approximately the same pairs of exceedance thresholds as those already identified in Fig. 6c, earlier times of emergence are obtained when considering both marginal and dependence changes (Fig. 6a), than when considering only marginal changes (Fig. 6b). Concerning the median contributions over all sliding windows of the marginal (Fig. 6d), dependence (Fig. 6e) and interactions terms (Fig. 6f), results vary according to the exceedance thresholds considered. While, for a large proportion of the exceedance thresholds, marginal properties changes contribute strongly to probability changes (Fig. 6d), dependence properties changes contribute dominantly to probability changes of very high wind and precipitation extremes (Fig. 6e). Regarding the "interaction" term, its contributions are close to 0, indicating little influence on the probability changes.

4.3 Results for Indiv- and Full-Ensemble version and a single exceeding threshold

We now present the results obtained for time of emergence and contributions for the Indiv- and Full-Ensemble versions for a single exceeding threshold. The methodology, previously illustrated on the CNRM-CM6 simulations, is now applied to each of the 13 models. Concerning the Indiv-Ensemble version, only one model (INMCM-5.0) had more than 5% of goodness-of-fit tests over all sliding windows rejecting the hypothesis that the copula is a good fit, and hence was excluded from the analysis (see Appendix B for further details).

We first present the results obtained for probabilities of exceeding the 80th percentiles of selected points of high values of wind and precipitation for the 1871-1900 reference period. Fig. 7 presents time series of exceedance probabilities obtained for the Indiv- and Full-Ensemble versions. Probability time series obtained for the 12 models when considering changes of marginal and dependence (Fig. 7a), marginal (Fig. 7b) and dependence properties (Fig. 7c) are displayed, as well as ToE at the 68 % confidence level for the individual models and their multi-model median estimate. Similarly, probability time series are shown for the Full-Ensemble version in Figs. 7d, e and f. Results for time of emergence at the 95 % confidence level are presented in Fig. S3 of the Supplement. When considering future changes of both marginal and dependence properties (Fig. 7a), half of the models (6/12) detects a time of emergence at the 68% confidence level. When found, a relatively important variability of ToE across climate models is obtained (varying between 2009 (1994-2023) and 2083 (2068-2097), Fig. 7a). These different results — i.e. either a ToE is detected or not, and the important variability of the year of emergence when found — indicate discrepancies of statistical properties of compound wind and precipitation extremes between climate models. For marginal

changes (Fig. 7b), 7 models out of 12 detect a time of emergence, within a smaller range of values. It suggests a slightly better agreement of marginal changes for future periods between models when time of emergence is defined. Moreover, models that show emergence when considering marginal changes only are not necessary those that show emergence when considering both future marginal and dependence changes. Indeed, 2 out of the 7 models emerging with marginal changes are not those from the 6 emerging when marginal and dependence changes are taken into account (not shown). Hence, marginal changes alone are not always sufficient to make the probability signal emerge. Concerning dependence changes (Fig. 7c), 2 models out of 12 detect a time of emergence, indicating that dependence property changes for these two models influence greatly exceedance probabilities by 2100. However, it also suggests that, for most of the models, the influence of the dependence properties changes on exceedance probabilities are too small to make the probability signals go out from the reference confidence interval by 2100. These results on the stationarity of dependence structures complement those of Vrac et al. (2022), where the ability of CMIP6 models to capture and represent significant changes in inter-variable dependencies is questioned.

Concerning the results for the Full-Ensemble version, emergence at the 68% confidence level is detected when considering marginal and dependence changes (Fig. 7d), marginal changes only (Fig. 7e) and dependence changes only (Fig. 7f) of pooled data. Emergence for the Full-Ensemble version can be partly explained by the pooling step which mechanically reduces uncertainties in marginal and copula fitting. Then, confidence intervals, including that of the reference period, are smaller than those obtained for individual models, which leads to emergence of probability signals with small probability changes (as for probability changes induced by dependence changes only in Fig. 7f). Thus, ToE are here detected for the Full-Ensemble version despite the pooling procedure that could reduce the signal by combining models simulating different evolutions of probabilities. Results for time of emergence presented in Fig. 7 for both Indiv- and Full-Ensemble versions are summarised in Fig. S4 of the Supplement.

Now, contributions of marginal, dependence and interaction terms in probability changes are quantified for the Indiv- and Full-Ensemble versions. For the Indiv-Ensemble versions, contributions are computed for each model separately and summarised by computing the median contribution of the models. Fig. 8 displays the median contributions over all sliding windows for the 12 climate models separately, as well as for the Indiv- and Full-Ensemble versions. Time series of bivariate FAR, relative differences and contributions along sliding windows for the Indiv- and Full-Ensemble versions are also displayed in Fig. S5 of the Supplement. Fig. 8 shows that, depending on the model, different results are obtained for the contributions to probability changes. Indeed, while some models present balanced contributions, i.e. marginal and dependence terms contributing to $\approx 50\%$ each to probability changes (e.g., CMCC-ESM2, CNRM-CM6-1 and CNRM-CM6-1-HR), other models show very unbalanced contributions, with one statistical property mainly driving the probability changes. For example, the dependence term contributes dominantly ($\geq 65\%$) to probability changes for the models CanESM5, FGOALS-g3 and INM-CM-4-8, while the marginal term contributes the most for EC-Earth3, GFDL-CM4, IPSL-CM61-LR, MIROC6, MPI-ESM1-2-LR and MRI-ESM2-0. Results for Indiv- and Full-Ensemble versions are also reported, both indicating a contribution to probability changes of ≈ 60 % from changes in marginal properties and ≈ 40 % from changes in dependence properties. Concerning the interaction term, as obtained previously in Sect. 4.1, its contribution is close to zero for each model individually, and for Indiv- and Full-Ensemble versions.

4.4 Results for Indiv- and Full-Ensemble versions and several exceeding thresholds

As previously done in Sect. 4.2, we now compute times of emergence for all combinations of exceedance thresholds between the 5th and 95th percentiles, for both Indiv- and Full-Ensemble versions, in Fig. 9. Note that, here, exceedance thresholds are now expressed in terms of percentiles to enable a comparison of results. Fig. 9a shows multi-model medians of ToE values induced by both marginal and dependence changes, i.e., results obtained for the Indiv-Ensemble version. A median value of time of emergence is obtained for any considered bivariate threshold, indicating that, for each exceedance threshold, at least one model presents an emergence. However, median ToE values show a variability depending on the bivariate exceedance thresholds. Note that, for the Indiv-Ensemble version, the number of models presenting a time of emergence can also vary from one bivariate threshold to another. For each exceedance threshold, the number of models emerging at the 68 % confidence level, as well as interquartile values, are shown in Fig. S6 of the Supplement. In particular, Fig. S6a indicates that all of the 12 models present a time of emergence for the probability of events exceeding very high precipitation and relatively low wind speed values (upper-left corner of the subplot). It suggests that all models agree on a change of the probability of occurrence of such events. This large consensus between models is not reached for events exceeding relatively low precipitation and very high wind speed values. Therefore, while all models simulate a significant increase of extreme precipitation events, it is not necessarily the case for extreme wind speed events. Results obtained for time of emergence induced by marginal properties only (Figs. 9b and S6b) are quite similar, although still indicating small differences with those obtained by considering marginal and dependence changes. Indeed, small differences of time of emergence can be observed, in particular for the upper-right area corresponding to very high wind speed and precipitation extremes. As observed in Sect. 4.1, this area corresponds to the area where dependence properties changes make emerging exceedance probability from the reference period (Fig. 9c), suggesting their importance for the probability changes of such events. This result however should not be overstated, as only ≈ 2 models show dependence changes large enough to lead to the emergence of probability (Fig. S6c).

The results of the Full-Ensemble approach are quite different from those of the Indiv-Ensemble one. For example, Fig. 9d indicates that the time of emergence for exceedance probabilities of low wind speed and high precipitation values is ≈ 2000 (while later for Indiv-Ensemble version, i.e. ≈ 2040). The results when considering marginal and dependence changes (Fig. 9d) and marginal changes only (Fig. 9e) are quite similar, indicating that changes in marginal properties mainly drive emergence of probabilities for each of the exceedance thresholds. A clear gradient of ToE values across exceedance thresholds is present: the more extreme the precipitation and the less extreme the wind speed, the sooner the time of emergence of exceedance probability. Conversely, the less extreme the precipitation and the more extreme the wind speed, the later the ToE. In fact, pooling data somehow strengthens the results for time of emergence when models agree on probability changes. Indeed, as seen previously, individual models agree in simulating a significant increase in probability of events exceeding low wind speed and high precipitation values. For ToE induced by dependence properties changes only (Fig. 9f), quite interestingly, probabilities emerge for exceedance thresholds more or less corresponding to the ones identified for Indiv-Ensemble in Fig. 9c. Although dependence properties seem to be stable over time for the majority of the models as observed in Fig. 7c, the resulting dependence structure of pooled data and its changes over sliding windows lead to obtain ToE values of exceeding probabilities.

One should also keep in mind that the reduced uncertainty for probability estimations resulting from the pooling process plays an important role in ToE detection for the Full-Ensemble version. For illustration purposes, evolutions of the bivariate distributions for the Full-Ensemble version are shown in Fig. S7 of the Supplement. Also, results for time of emergence at 95% and the number of models emerging for each exceedance threshold are displayed in Fig. S8 and Fig. S9 of the Supplement.

Median contribution of marginal, dependence and interactions terms are displayed in Fig. 10 for both Indiv- and Full-Ensemble versions. The results obtained previously concerning the importance of the marginal properties on probability changes are here confirmed: for all exceedance thresholds, marginal properties changes contribute to more than 50 % of probability changes for both Indiv- and Full-Ensemble versions (Figs. 10a and d). Concerning dependence changes' contribution (Figs. 10b and e), the median values obtained are less than 50%, but specific pairs of exceedance thresholds highlight again the varying importance of dependence properties on exceedance probability changes: for both Indiv- and Full-versions, median contribution of dependence properties are high for the probability changes of events exceeding high wind speed and high precipitation values. The area of exceedance thresholds for which dependence properties contribute greatly to probability changes is however greater for the Full-version (Fig. 10e) than for the Indiv-Ensemble version (Fig. 10b). Again, these results have to be directly linked with those obtained for the emergence of probabilities of such events due to dependence changes in Figs. 9c and f. Concerning the interaction term (Figs. 10c and e), contribution values are equal to 0 for both Indiv- and Full-Ensemble versions, highlighting again the negligible role of this term in probability changes.

5 Results for growing-period frosts

We now apply our methodology to analyse a second type of compound events: growing-period frosts. Contrary to compound wind and precipitation extremes, for which we were interested in exceedance probabilities (i.e. both contributing variables exceeding thresholds), we are interested here in probability of growing-period frosts, i.e. the probability of having a GDD value exceeding a threshold of 200 ($GDD \geq 200$) by the end of March — and hence characterising bud burst conditions — and having a frost in April, i.e. having $T \leq 0$. Hence, we applied our methodology described in Sect. 3 on bivariate points of GDD and minimal temperature data (one pair by year) by adapting Eq. (2) to compute the probabilities of interest. For example, for the probability of growing-period frosts in the reference period, it is computed as follows (Yue and Rasmussen, 2002):

$$\begin{aligned} p_{m,d}(0,150) &= \mathbb{P}(T \leq 0 \cap GDD \geq 200) \\ &= F_T(0) - C(F_T(0), F_{GDD}(200)). \end{aligned} \tag{9}$$

Although the main results are presented for a threshold of 200 °C.day, additional results for thresholds of 150 °C.day and 250 °C.day are displayed in the Supplement to assess risks of growing-period frosts for earlier and later bud burst plants.

5.1 Indiv- and Full-Ensemble results

We now present the results for the growing-period frosts. For the Indiv-Ensemble version, as previously, only one model (CMCC-ESM2) is excluded from the ensemble as it presents more than 5% of goodness-of-fit tests rejecting the hypothesis that fitted copulas are a good fit (see Appendix B for further details). Before computing any probability, Fig 11 displays the changes along sliding windows of the fitted bivariate distributions for the Full-Ensemble version, i.e., after pooling GDD and minimal temperature data of the different models. Clearly, a change of bivariate distributions for future periods can be visually assessed when marginal properties changes are considered (Figs. 11a and b). In particular, it presents an increase of both minimal temperature and GDD values, which could be expected in a context of global warming. The upper-left areas corresponding to probabilities of growing-period frost events ($\{G \geq 200 \cap T \leq 0\}$) are approximately similar for the first sliding windows, but their sizes increase for future periods, suggesting a greater probability of growing-period frosts induced by marginal properties changes. However, when dependence properties changes are only considered without marginal changes (Fig. 11c), bivariate distributions are quite similar and the upper-left area is almost identical in size, suggesting that the effect of dependence properties changes on growing-period frost probability is small.

Fig. 12 presents the time series of probabilities obtained for the Indiv-and Full-Ensemble versions for growing-period frost events. Results for 150°C.d and 250°C.d GDD thresholds are presented in the Supplement in Figs. S10 and S11, respectively. By considering climate models separately, a time of emergence at 68 % confidence level is detected for 11 out of 12 models when marginal properties changes are taken into account (Figs. 12a and b). Although a large majority of models agrees by simulating a significant change of growing-period frost probability with respect to the reference period, times of emergence are quite scattered, indicating differences in simulations of growing-period frosts. By considering dependence changes only (Fig. 12c), none of the 12 models within the Indiv-Ensemble presents a time of emergence, indicating that the influence of dependence changes alone is not strong enough to modify growing-period frost probabilities. For the Full-Ensemble version, changes of marginal and dependence properties (Fig. 12d) and changes of marginal properties only (Fig. 12e) lead to increase growing-period frosts probability such that time of emergence is detected at 1905 and 1906, respectively. Probability time series are quite similar, suggesting again that changes of dependence properties do not influence strongly probability of growing-period frosts. It is confirmed in Fig. 12f, for which no significant change of probability induced by dependence changes only are observed between the reference and future periods. Times of emergence obtained for growing-period frosts are summarised in Fig. S12 of the Supplement.

Fig. 13 displays the median contribution of the marginal, dependence and interaction terms to probability changes for each climate model individually and for the Indiv- and Full-Ensemble versions. For the climate models individually, as well as for the Indiv- and Full-Ensemble versions, the results are quite clear: marginal properties are the statistical properties contributing the most to probability changes of growing-period frosts. Fig. S13 shows contribution across sliding windows and hereby confirms that contributions of the dependence and interaction to change of probability are rather limited along the whole time period.

6 Conclusion, discussion and future work

6.1 Conclusions

In this study, we have presented a new methodology to assess time of emergence of compound hazards probabilities. Using a copula-based multivariate framework, we also propose to quantify the contributions of marginal and dependence properties to probability changes of hazards leading to compound events. The methodology has been applied to analyse two different climate hazards with potentially high-impacts, using a 13-member multi-model ensemble (CMIP6): compounding wind and precipitation extremes in Brittany and growing-period frosts over Central France. For each hazard, the methodology has been applied in two different versions: the Indiv-Ensemble version, for which the methodology is applied to individual climate models to derive time of emergence of probabilities and contributions of statistical properties of each model separately, and the Full-Ensemble version, for which the methodology is applied to bias-corrected and pooled data from the different models. Depending on the version, the objectives are not exactly the same: whereas the Indiv-Ensemble version enables us to estimate the uncertainty in ToE values and contributions to multivariate hazards probability changes arising from inter-model differences, the Full-version allows us to get unique ToE and contribution values accounting for the whole ensemble, that is, by taking into account the global uncertainty inherent in climate modelling.

Results for compounding wind and precipitation extremes over Brittany show that occurrence probabilities of such events are likely to increase and potentially emerge before the end of the 21st century. However, the reason of these increased probabilities can be different depending on climate models: while, for some models, probability changes are mainly driven by marginal changes only, other models give a strong importance to both marginal properties and dependence properties. It results in having a mixed importance ($\sim 65\%$ and 35%) of both marginal and dependence properties that contribute to probability changes within the Full-Ensemble version. These results highlight the importance of carefully taking into consideration the dependence structure when studying the evolution of probabilities of compound wind and precipitation extremes.

Concerning growing-period frosts over Central France, a large majority of models agrees on the emergence of probabilities of such events. They also agree on the dominant contribution of marginal properties changes, while the contribution of dependence properties are mostly negligible.

By analysing two different case studies, our results highlight that the importance of marginal and dependence properties to probability changes can differ from a compound hazard to another, and from one climate model to another. It thus stresses the importance of considering both marginal and dependence properties carefully, as well as their inter-model variability, to analyse the future evolution of multivariate hazards leading to compound events.

6.2 Discussion and perspectives

In this study, emergence of probabilities of multivariate hazards has been investigated with respect to the baseline period 1871-1900. This period can be considered as representative of the beginning of the industrial era (e.g., Hawkins et al., 2020) and can hence be of interest to assess if anthropogenic climate change has contributed to an emergence of probability of multivariate hazards. However, other baseline periods could have been chosen, such as more recent ones which would provide useful

results for adaptation planning (e.g., Ossó et al., 2022). Of course, depending on the chosen baseline period, the estimated natural variability that serves as reference for assessing changes would be different, and thus would affect the ToE results. As an illustration, Fig. S14 shows results from a quick sensitivity experiment for the time of emergence of probabilities of compounding wind and precipitation depending on the choice of the baseline period for the CNRM-CM6 model. It illustrates that results of emergence can vary strongly depending on the chosen baseline period. In addition to modifying the potential time of emergence, the choice of the baseline period can also influence the results of contributions from the statistical properties changes (not shown), as these statistical changes are also assessed with respect to the baseline period.

Moreover, in this study, time of emergence of probability signals is defined as the year or time period for which the probability signal *permanently* excesses a certain threshold (e.g., Hawkins and Sutton, 2012; Maraun, 2013; Hawkins et al., 2020). As the Earth's climate system is highly nonlinear and non-monotonic, detecting the emergence of a signal in this way can be limited depending on the climate signal under study. Analysing "periods of emergence" (PoE) instead of time of emergence may be more relevant to rather describe specific periods where probability signals emerge significantly — but temporarily — from reference natural variability. This notion of PoE would better highlight not only the non-linearities of the CE changes but also the differences of evolution of probability between climate models, as it was observed for growing-period frosts in Sec. 5. Indeed, in Fig. 12, while some climate models reach their highest growing-period frosts probability for the late 21st century, other climate models present a decrease of probability to 0 for the end of the century after having reached maximum growing-period frosts probability earlier. In other words, probabilities for future periods may differ, not permanently, but only temporarily from the estimated probability associated with natural variability. This could justify the development, the investigation and the use of the notion of temporary periods of emergence.

In addition, changes in marginal properties of the different variables and their contributions to probability changes have been assessed together, i.e., without separating the changes and contributions from wind and precipitation, nor those from GDD and minimum temperature. Thus, it does not allow us quantifying by how much individual variables' changes drive probability changes. Some studies already concluded about the importance of individual variables in the change of occurrence of multivariate hazards (e.g., Manning et al., 2018; Brunner et al., 2021; Calafat et al., 2022). Our methodology can however be easily adapted to quantify such information by keeping fixed marginal properties of only one contributing variable and assess probability changes. By doing this for the different variables in turn, the contribution of marginal changes to probability changes would be decomposed according to individual variables changes.

This study shows that both univariate and multivariate properties can be essential in determining CE properties. However, despite substantial improvements in climate modelling, climate simulations often remain biased compared to observations or reanalyses in terms of both univariate and multivariate properties (e.g., Cannon, 2018; Vrac, 2018; François et al., 2020). This could have major consequences on the ability of climate models to simulate compound events accurately (Zscheischler et al., 2019; Villalobos-Herrera et al., 2021; Vrac et al., 2021; Ridder et al., 2021), and then on the resulting analyses involved in decision-making processes. A few multivariate bias correction methods, i.e. statistical methods that are able to adjust both univariate and multivariate properties of simulations with respect to reference dataset, have been recently developed (e.g., Cannon, 2018; Guo et al., 2019; Mehrotra and Sharma, 2019; Robin et al., 2019; Vrac and Thao, 2020; François et al., 2021).

However, such MBC methods are designed to adjust the whole statistical distribution of climate simulations, and their abilities to increase the realism of specific parts of the statistical distribution (such as multivariate extremes) have never been tested, while it can be crucial for specific CEs. This is therefore an important perspective and the methodology developed in the present study could be a way to evaluate the consequences of MBC methods, e.g., in terms of ToE and contributions of marginal and dependence properties.

It has to be noted that uncertainty in probabilities of multivariate hazards has been assessed by considering uncertainty in both statistical fitting procedures and model-to-model differences. However, uncertainty arisen from internal climate variability, i.e., from the inherent chaotic nature of the climate system, has not been investigated. Assessing and analysing these uncertainties is however key to better characterise them and thus provide useful information for policy-makers (Raymond et al., 2022; Bevacqua et al., 2022). Future extensions of the framework presented herein could thus focus on using multimodel large-ensemble simulations to assess more robustly probabilities of hazards, contributions of statistical properties changes to their emergence, and their associated uncertainties resulting from both internal variability and structural model differences.

It is also important noting that the role of physical drivers of multivariate hazards has not been investigated in this study. Indeed, recent studies highlight the importance of large-scale climate modes (e.g., De Luca et al., 2020b; Singh et al., 2021b) and atmospheric circulation regimes (e.g., Faranda et al., 2020; Jézéquel et al., 2020; Vrac et al., 2021) on compound and extreme events. Understanding the influences of physical drivers and their changes on the statistical features and probabilities of multivariate hazards is a key research which has important implications for predicting their occurrence and characterising their impacts.

As mentioned in Sect. 1, the present methodology has been developed and applied in a ToE framework that is different from attribution. We have not considered factual and counterfactual worlds with different forcings to assess the effects of climate change on multivariate hazards probabilities. Adapting and applying our methodology in an attribution setting is thus an interesting perspective that would complement the existing multivariate event attribution framework recently developed (e.g., Kiriliouk and Naveau, 2020; Zscheischler and Lehner, 2021). In addition to attributing changes of compound events, our methodology would permit to quantify the underlying contributions of the changes in marginal and dependence properties, hence better characterising the statistical features of climate change.

Code availability. Custom codes developed for the analyses are publicly available at https://github.com/bastien-francois/ToE_CE.

Data availability. CMIP6 climate model data can be downloaded through the Earth System Grid Federation portals. Instructions to access the data are available here: https://pcmdi.llnl.gov/CMIP6/Guide/dataUsers.html, last access: 23 January 2022.

Sample availability. TEXT

Video supplement. TEXT

Appendix A: Procedure for confidence intervals estimation

Confidence intervals of bivariate exceedance probabilities are estimated by combining the confidence intervals from the fitted parameters for both marginal distributions and copulas. For both marginal distributions and copula, the fitted parameters and their 68 % (resp. 95 %) confidence intervals are estimated using MLE (as described in Appendix B) and profile likelihood (e.g., Venzon and Moolgavkar, 1988; Hofert et al., 2012). Estimating the 68 % (resp. 95 %) confidence intervals for bivariate exceedance probabilities consists in: (i) resampling uniformly and independently the fitted parameters of the two marginal distributions within their 68 % (resp. 95 %) profile likelihood confidence intervals, (ii) computing the bivariate exceedance probability using the resampled parameters for marginal distributions and the copula parameter estimated using MLE, (iii) repeating the two previous step 100 times to construct a sampling distribution for the bivariate exceedance probability, (iv) searching which combinations of the resampled parameters lead to the 16 % and 84 % (resp. 2.5 % and 97.5 %) percentiles of the re-estimated bivariate exceedance probabilities, (v) using the copula parameter uncertainty, estimating the 68 % (resp. 95 %) confidence intervals of the 16 % and 84 % (resp. 2.5 % and 97.5 %) percentiles of the bivariate exceedance probabilities. The lower and upper bounds of these two confidence intervals define the final confidence interval combining both marginal and copula parameters uncertainty.

Appendix B: Marginal and copula fitting

For the fitting of the marginal distributions, we considered the Akaike information criterion (AIC) to select the best families among Gaussian, generalized extreme value and generalized Pareto distributions. The marginal distributions of wind speed and precipitation beyond the selection thresholds were modeled by generalized Pareto distributions. For growing-period frost events, the marginal distributions of the GDD indices were modeled using Gaussian distributions. We modeled the negative of the minimal temperatures using GEV distributions and transformed back.

For fitting of the copulas, marginal distributions are transformed into uniform distribution using normalized ranks (e.g., Salvadori et al., 2011; Serinaldi, 2015; Bevacqua et al., 2019). This procedure is common for copula analysis as it allows to perform appropriate goodness of fit tests (Genest et al., 2009). In this study, four Archimedean copulas (Clayton, Frank, Gumbel and Joe) are considered. These copulas have been widely used in hydrology and climate studies (e.g., Zscheischler and Seneviratne, 2017; Liu et al., 2018b; Tavakol et al., 2020) and allow the dependence structure to be modelled with a single parameter that determines the strength of the dependence. Moreover, the four Archimedean copulas differ in how they model dependence structures. For instance, the Gumbel and Joe copulas have upper tail dependence, which means that they are able to model correlated extremes. The Clayton copula has lower tail dependence and the Frank copula has no tail dependence. A complete overview of copula families, their related functions and the range of their parameters is offered by Sadegh et al. (2017). For each climate model and each sliding window, the best copula family is determined using the Akaike Information

Criterion. Copulas were fitted through maximum likelihood estimators (MLE) using the copula (Hofert et al., 2020) and VineCopula: (Schepsmeier et al., 2016) R-packages. Goodness of fit are tested based on the White's information matrix equality (White, 1982; Huang and Prokhorov, 2014) implemented in the R package VineCopula (Schepsmeier et al., 2016). To evaluate exceedance probabilities, we select the copula family that has been the most selected along all the sliding windows and for which less than 5% of the goodness of fit tests conclude to the rejection that data fits well the considered copula distribution. For the Indiv-Ensemble version, climate models for which more than 5% of the goodness of fit tests conclude to a rejection are excluded.

Appendix C: Transformation of wind and precipitation data using CDF-t

As selection thresholds for wind and precipitation extremes are not the same for all the climate models, we need to transform selected wind and precipitation data. For each model, bivariate points of high values are selected using the individual 90th percentiles of wind and precipitation variables. Then, the selected bivariate data from the different models are adjusted with respect to a model taken as reference, using a univariate bias correction technique called the "Cumulative Distribution Function – transform" method (CDF-t, Michelangeli et al., 2009; Vrac et al., 2012). The CDF-t method allows to correct the univariate distribution of a modeled climate variable via a quantile-quantile method that takes into account potential changes of the univariate distribution in the correction procedure. By choosing a model as reference (CNRM-CM6), we use here the CDF-t method to transform marginal properties of selected wind and precipitation values of each climate dataset with respect to CNRM-CM6. This way, marginal distributions of wind and precipitation extremes are similar between the different climate models and are thus more consistent with each other. We consider the 1871-1900 sliding window as reference period for the calibration of the bias correction. Once data have been transformed for each climate model, bivariate wind and precipitation extreme values from the different models can be pooled and the Full-Ensemble methodology can be applied.

Author contributions. MV had the initial idea of the study. MV and BF designed the experiments and protocols. BF made all computations and figures. BF and MV made the analyses and interpretations. BF wrote the first complete draft of the article, with inputs, corrections and additional writing contributions from MV.

Competing interests. The authors declare that they have no competing interests.

Acknowledgements. We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modeling groups (listed in Table 1 of this paper) for producing and making available their models outputs. For CMIP, the U.S. Department of Energy's Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals. The authors

acknowledge support from the EUR IPSL Climate Graduate School project managed by the ANR under the "Investissements d'avenir" programme with the reference ANR-11-IDEX-0004-17-EURE-0006, the European Union's Horizon 2020 research and innovation programme via the "XAIDA" project (Grant agreement No. 101003469), as well as from the "COESION" project funded by the French National program LEFE (Les Enveloppes Fluides et l'Environnement).

References

Abatzoglou, J. T., Dobrowski, S. Z., and Parks, S. A.: Multivariate climate departures have outpaced univariate changes across global lands, Sci. Rep., 10, 2020.

Bevacqua, E., Maraun, D., Hobæk Haff, I., Widmann, M., and Vrac, M.: Multivariate statistical modelling of compound events via pair-copula constructions: analysis of floods in Ravenna (Italy), Hydrol. Earth Syst. Sci., 21, 2701–2723, https://doi.org/10.5194/hess-21-2701-2017, 2017.

Bevacqua, E., Maraun, D., Vousdoukas, M. I., Voukouvalas, E., Vrac, M., Mentaschi, L., and Widmann, M.: Higher probability of compound flooding from precipitation and storm surge in Europe under anthropogenic climate change, Sci. Adv., 5, https://doi.org/10.1126/sciadv.aaw5531, 2019.

Bevacqua, E., De Michele, C., Manning, C., Couasnon, A., Ribeiro, A. F. S., Ramos, A. M., Vignotto, E., Bastos, A., Blesic, S., Durante, F., and et al.: Bottom-up identification of key elements of compound events, Earth Space Sci., p. 29, https://doi.org/10.1002/essoar.10507809.1, 2021.

Bevacqua, E., Zappa, G., Lehner, F., and Zscheischler, J.: Precipitation trends determine future occurrences of compound hot–dry events, Nat. Clim. Chang., https://doi.org/10.1038/s41558-022-01309-5, 2022.

Bonhomme, R.: Bases and limits to using 'degree.day' units, Eur. J. Agron., 13, 1–10, https://doi.org/10.1016/S1161-0301(00)00058-7, 2000.

Boucher, O., Denvil, S., Levavasseur, G., Cozic, A., Caubel, A., Foujols, M.-A., Meurdesoif, Y., Cadule, P., Devilliers, M., Ghattas, J., Lebas, N., Lurton, T., Mellul, L., Musat, I., Mignot, J., and Cheruy, F.: IPSL IPSL-CM6A-LR model output prepared for CMIP6 CMIP, https://doi.org/10.22033/ESGF/CMIP6.1534, 2018.

Brunner, M. I., Swain, D. L., Gilleland, E., and Wood, A. W.: Increasing importance of temperature as a contributor to the spatial extent of streamflow drought, Environ. Res. Lett., 16, 024 038, https://doi.org/10.1088/1748-9326/abd2f0, 2021.

Calafat, F. M., Wahl, T., Tadesse, M. G., and Sparrow, S. N.: Trends in Europe storm surge extremes match the rate of sea-level rise, Nature, 603, 841–845, https://doi.org/10.1038/s41586-022-04426-5, 2022.

Cannon, A. J.: Multivariate quantile mapping bias correction: an N-dimensional probability density function transform for climate model simulations of multiple variables, Clim. Dynam., 50, 31–49, https://doi.org/10.1007/s00382-017-3580-6, 2018.

Cherchi, A., Fogli, P. G., Lovato, T., Peano, D., Iovino, D., Gualdi, S., Masina, S., Scoccimarro, E., Materia, S., Bellucci, A., and Navarra, A.: Global Mean Climate and Main Patterns of Variability in the CMCC-CM2 Coupled Model, J. Adv. Model. Earth Syst., 11, 185–209, https://doi.org/10.1029/2018MS001369, 2019.

Chiang, F., Greve, P., Mazdiyasni, O., Wada, Y., and AghaKouchak, A.: A Multivariate Conditional Probability Ratio Framework for the Detection and Attribution of Compound Climate Extremes, Geophys. Res. Lett., 48, e2021GL094 361, https://doi.org/10.1029/2021GL094361, 2021.

Christensen, J., Hewitson, B., Busuioc, A., Chen, A., Gao, X., Held, I., Jones, R., Kolli, R., Kwon, W.-T., Laprise, R., Rueda, V., Mearns, L., Menéndez, C., Räisänen, J., Rinke, A., Sarr, A., and Whetton, P.: Regional climate projections. Climate change 2007: The physical science basis, Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, pp. 847–940, 2007.

De Luca, P., Messori, G., Pons, F. M. E., and Faranda, D.: Dynamical systems theory sheds new light on compound climate extremes in Europe and Eastern North America, Q. J. Roy. Meteor. Soc., 146, 1636–1650, https://doi.org/10.1002/qj.3757, 2020a.

De Luca, P., Messori, G., Wilby, R. L., Mazzoleni, M., and Di Baldassarre, G.: Concurrent wet and dry hydrological extremes at the global scale, Earth Syst. Dynam., 11, 251–266, https://doi.org/10.5194/esd-11-251-2020, 2020b.

Diffenbaugh, N. and Scherer, M.: Observational and model evidence of global emergence of permanent, unprecedented heat in the 20th and 21st centuries, Clim. Change, 107, 615–624, https://doi.org/10.1007/s10584-011-0112-y, 2011.

Diffenbaugh, N. S., Swain, D. L., and Touma, D.: Anthropogenic warming has increased drought risk in California, Proc. Natl. Acad. Sci. U.S.A., 112, 3931–3936, https://doi.org/10.1073/pnas.1422385112, 2015.

EC-Earth: EC-Earth-Consortium EC-Earth3 model output prepared for CMIP6 ScenarioMIP ssp585, https://doi.org/10.22033/ESGF/CMIP6.4912, 2019.

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., and Taylor, K. E.: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, Geosci. Model Dev., 9, 1937–1958, https://doi.org/10.5194/gmd-9-1937-2016, 2016.

Faranda, D., Vrac, M., Yiou, P., Jézéquel, A., and Thao, S.: Changes in Future Synoptic Circulation Patterns: Consequences for Extreme Event Attribution, Geophys. Res. Lett., 47, e2020GL088 002, https://doi.org/10.1029/2020GL088002, 2020.

Fink, A. H., Brücher, T., Ermert, V., Krüger, A., and Pinto, J. G.: The European storm Kyrill in January 2007: synoptic evolution, meteorological impacts and some considerations with respect to climate change, Nat. Hazards Earth Syst. Sci., 9, 405–423, https://doi.org/10.5194/nhess-9-405-2009, 2009.

Fischer, E. M., Sedláček, J., Hawkins, E., and Knutti, R.: Models agree on forced response pattern of precipitation and temperature extremes, Geophys. Res. Lett., 41, 8554–8562, https://doi.org/10.1002/2014GL062018, 2014.

Frame, D., Joshi, M., Hawkins, E., Harrington, L., and Róiste, M.: Population-based emergence of unfamiliar climates, Nat. Clim. Chang., 7, 407–411, https://doi.org/10.1038/nclimate3297, 2017.

François, B., Vrac, M., Cannon, A. J., Robin, Y., and Allard, D.: Multivariate bias corrections of climate simulations: Which benefits for which losses?, Earth Syst. Dyn., 11, 537–562, https://doi.org/10.5194/esd-2020-10, 2020.

François, B., Thao, S., and Vrac, M.: Adjusting spatial dependence of climate model outputs with cycle-consistent adversarial networks, Clim. Dynam., 57, 3323–3353, https://doi.org/10.1007/s00382-021-05869-8, 2021.

Gaetani, M., Janicot, S., Vrac, M., Famien, A. M., and Sultan, B.: Robust assessment of the time of emergence of precipitation change in West Africa, Sci. Rep., 10, https://doi.org/10.1038/s41598-020-63782-2, 2020.

Garcia de Cortazar-Atauri, I., Brisson, N., and Gaudillere, J.: Performance of several models for predicting budburst date of grapevine (Vitis vinifera L.), Int. J. Biometeorol., 53, 317–326, https://doi.org/10.1007/s00484-009-0217-4, 2009.

Genest, C., Remillard, B., and Beaudoin, D.: Goodness-of-fit tests for copulas: A review and a power study, Insur. Math. Econ., 44, 199–213, https://doi.org/10.1016/j.insmatheco.2007.10.005, 2009.

Giorgi, F. and Bi, X.: Time of emergence (TOE) of GHG-forced precipitation change hot-spots, Geophys. Res. Lett., 36, https://doi.org/10.1029/2009GL037593, 2009.

Guo, H., John, J. G., Blanton, C., McHugh, C., Nikonov, S., Radhakrishnan, A., Rand, K., Zadeh, N. T., Balaji, V., Durachta, J., Dupuis, C., Menzel, R., Robinson, T., Underwood, S., Vahlenkamp, H., Dunne, K. A., Gauthier, P. P., Ginoux, P., Griffies, S. M., Hallberg, R., Harrison, M., Hurlin, W., Lin, P., Malyshev, S., Naik, V., Paulot, F., Paynter, D. J., Ploshay, J., Schwarzkopf, D. M., Seman, C. J., Shao, A., Silvers, L., Wyman, B., Yan, X., Zeng, Y., Adcroft, A., Dunne, J. P., Held, I. M., Krasting, J. P., Horowitz, L. W., Milly, C., Shevliakova, E., Winton, M., Zhao, M., and Zhang, R.: NOAA-GFDL GFDL-CM4 model output prepared for CMIP6 ScenarioMIP ssp585, https://doi.org/10.22033/ESGF/CMIP6.9268, 2018.

Guo, Q., Chen, J., Zhang, X., Shen, M., Chen, H., and Guo, S.: A new two-stage multivariate quantile mapping method for bias correcting climate model outputs, Clim. Dynam., 53, 3603–3623, https://doi.org/10.1007/s00382-019-04729-w, 2019.

Hawkins, E. and Sutton, R.: Time of emergence of climate signals, Geophys. Res. Lett., 39, https://doi.org/10.1029/2011GL050087, 2012.

Hawkins, E., Frame, D., Harrington, L., Joshi, M., King, A., Rojas, M., and Sutton, R.: Observed Emergence of the Climate Change Signal: From the Familiar to the Unknown, Geophys. Res. Lett., 47, e2019GL086 259, https://doi.org/10.1029/2019GL086259, 2020.

Hillier, J., Matthews, T., Wilby, R., and Murphy, C.: Multi-hazard dependencies can increase or decrease risk, Nat. Clim. Chang., 10, 1–4, https://doi.org/10.1038/s41558-020-0832-y, 2020.

Hofert, M., Mächler, M., and McNeil, A. J.: Likelihood inference for Archimedean copulas in high dimensions under known margins, J. Multivar. Anal., 110, 133–150, https://doi.org/10.1016/j.jmva.2012.02.019, 2012.

Hofert, M., Kojadinovic, I., Maechler, M., and Yan, J.: copula: Multivariate Dependence with Copulas, https://CRAN.R-project.org/package=copula, R package version 1.0-1, 2020.

Huang, W. and Prokhorov, A.: A Goodness-of-fit Test for Copulas, Econom. Rev., 33, 751–771, https://doi.org/10.1080/07474938.2012.690692, 2014.

IPCC, 2021: *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* [Masson-Delmotte, V., P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (eds.)]., Cambridge University Press. In Press.

Jézéquel, A., Bevacqua, E., d'Andrea, F., Thao, S., Vautard, R., Vrac, M., and Yiou, P.: Conditional and residual trends of singular hot days in Europe, Environ. Res. Lett., 15, 064 018, https://doi.org/10.1088/1748-9326/ab76dd, 2020.

Jiang, F., Hu, R.-j., Zhang, Y.-w., Li, X., and Tong, L.: Variations and trends of onset, cessation and length of climatic growing season over Xinjiang, NW China, Theor. Appl. Climatol., 106, 449–458, https://doi.org/10.1007/s00704-011-0445-5, 2011.

King, A. D., Donat, M. G., Fischer, E. M., Hawkins, E., Alexander, L. V., Karoly, D. J., Dittus, A. J., Lewis, S. C., and Perkins, S. E.: The timing of anthropogenic emergence in simulated climate extremes, Environ. Res. Lett., 10, 094 015, https://doi.org/10.1088/1748-9326/10/9/094015, 2015.

Kiriliouk, A. and Naveau, P.: Climate extreme event attribution using multivariate peaks-over-thresholds modeling and counterfactual theory, Ann. Appl. Stat., 14, 1342–1358, https://doi.org/10.1214/20-AOAS1355, 2020.

Lamichhane, J.-R.: Rising risks of late-spring frosts in a changing climate, Nat. Clim. Chang., 11, 554–555, https://doi.org/10.1038/s41558-021-01090-x, 2021.

Leonard, M., Westra, S., Phatak, A., Lambert, M., Hurk, B., Mcinnes, K., Risbey, J., Schuster, S., Jakob, D., and Stafford Smith, M.: A compound event framework for understanding extreme impacts, Wiley Interdiscip. Rev. Clim. Change, 2014.

Li, L.: CAS FGOALS-g3 model output prepared for CMIP6 ScenarioMIP ssp585, https://doi.org/10.22033/ESGF/CMIP6.3503, 2019.

Liberato, M. L.: The 19 January 2013 windstorm over the North Atlantic: large-scale dynamics and impacts on Iberia, Weather. Clim. Extremes, 5-6, 16–28, https://doi.org/10.1016/j.wace.2014.06.002, 2014.

Liu, Q., Piao, S., Janssens, I., Fu, Y., Peng, S., Lian, X., Ciais, P., Myneni, R., Penuelas, J., and Wang, T.: Extension of the growing season increases vegetation exposure to frost, Nat. Commun., 9, https://doi.org/10.1038/s41467-017-02690-y, 2018a.

Liu, Y., Cheng, Y., Zhang, X., Li, X., and Cao, S.: Combined Exceedance Probability Assessment of Water Quality Indicators Based on Multivariate Joint Probability Distribution in Urban Rivers, Water, 10, https://doi.org/10.3390/w10080971, 2018b.

Lobell, D. B. and Burke, M. B.: Why are agricultural impacts of climate change so uncertain? The importance of temperature relative to precipitation, Environ. Res. Lett., 3, 034 007, https://doi.org/10.1088/1748-9326/3/3/034007, 2008.

Mahlstein, I., Knutti, R., Solomon, S., and Portmann, R. W.: Early onset of significant local warming in low latitude countries, Environ. Res. Lett., 6, 034 009, https://doi.org/10.1088/1748-9326/6/3/034009, 2011.

Mahlstein, I., Hegerl, G., and Solomon, S.: Emerging local warming signals in observational data, Geophys. Res. Lett., 39, https://doi.org/10.1029/2012GL053952, 2012.

Manning, C., Widmann, M., Bevacqua, E., Loon, A. F. V., Maraun, D., and Vrac, M.: Soil Moisture Drought in Europe: A Compound Event of Precipitation and Potential Evapotranspiration on Multiple Time Scales, J. Hydrometeorol., 19, 1255–1271, https://doi.org/10.1175/JHM-D-18-0017.1, 2018.

Manning, C., Widmann, M., Bevacqua, E., Loon, A. F. V., Maraun, D., and Vrac, M.: Increased probability of compound long-duration dry and hot events in Europe during summer (1950–2013), Environ. Res. Lett., 14, 094 006, https://doi.org/10.1088/1748-9326/ab23bf, 2019.

Maraun, D.: When will trends in European mean and heavy daily precipitation emerge?, Environ. Res. Lett., 8, 014 004, https://doi.org/10.1088/1748-9326/8/1/014004, 2013.

Martius, O., Pfahl, S., and Chevalier, C.: A global quantification of compound precipitation and wind extremes, Geophys. Res. Lett., 43, 7709–7717, https://doi.org/10.1002/2016GL070017, 2016.

Mazdiyasni, O. and AghaKouchak, A.: Substantial increase in concurrent droughts and heatwaves in the United States, Proc. Natl. Acad. Sci. U.S.A., 112, 11 484–11 489, https://doi.org/10.1073/pnas.1422945112, 2015.

Mehrotra, R. and Sharma, A.: A Resampling Approach for Correcting Systematic Spatiotemporal Biases for Multiple Variables in a Changing Climate, Water Resour. Res., 55, 754–770, https://doi.org/10.1029/2018WR023270, 2019.

Messmer, M. and Simmonds, I.: Global analysis of cyclone-induced compound precipitation and wind extreme events, Weather. Clim. Extremes, 32, 100 324, https://doi.org/10.1016/j.wace.2021.100324, 2021.

Michelangeli, P.-A., Vrac, M., and Loukos, H.: Probabilistic downscaling approaches: Application to wind cumulative distribution functions, Geophys. Res. Lett., 36, L11 708, https://doi.org/10.1029/2009GL038401, 2009.

Nasr, A. A., Wahl, T., Rashid, M. M., Camus, P., and Haigh, I. D.: Assessing the dependence structure between oceanographic, fluvial, and pluvial flooding drivers along the United States coastline, Hydrol. Earth Syst. Sci., 25, 6203–6222, https://doi.org/10.5194/hess-25-6203-2021, 2021.

Nelsen, R. B.: An Introduction to Copulas , 2nd ed. Springer, 2006.

Ossó, A., Allan, R., Hawkins, E., Shaffrey, L., and Maraun, D.: Emerging new climate extremes over Europe, Clim. Dyn., 58, https://doi.org/10.1007/s00382-021-05917-3, 2022.

Pfleiderer, P., Menke, I., and Schleussner, C.-F.: Increasing risks of apple tree frost damage under climate change, Clim. Change, 157, https://doi.org/10.1007/s10584-019-02570-y, 2019.

Pohl, E., Grenier, C., Vrac, M., and Kageyama, M.: Emerging climate signals in the Lena River catchment: a non-parametric statistical approach, Hydrol. Earth Syst. Sci., 24, 2817–2839, https://doi.org/10.5194/hess-24-2817-2020, 2020.

Rana, A., Hamid, M., and Qin, Y.: Understanding the Joint Behavior of Temperature and Precipitation for Climate Change Impact Studies, Theor. Appl. Climatol., 129, https://doi.org/10.1007/s00704-016-1774-1, 2017.

Raveh-Rubin, S. and Wernli, H.: Large-scale wind and precipitation extremes in the Mediterranean: a climatological analysis for 1979–2012, Q. J. Roy. Meteor. Soc., 141, 2404–2417, https://doi.org/10.1002/qj.2531, 2015.

Raymond, C., Matthews, T., and Horton, R. M.: The emergence of heat and humidity too severe for human tolerance, Sci. Adv., 6, https://doi.org/10.1126/sciadv.aaw1838, 2020.

Raymond, C., Suarez-Gutierrez, L., Kornhuber, K., Pascolini-Campbell, M., Sillmann, J., and Waliser, D. E.: Increasing spatiotemporal proximity of heat and precipitation extremes in a warming world quantified by a large model ensemble, Environ. Res. Lett., 17, 035 005, https://doi.org/10.1088/1748-9326/ac5712, 2022.

Reinert, M., Pineau-Guillou, L., Raillard, N., and Chapron, B.: Seasonal Shift in Storm Surges at Brest Revealed by Extreme Value Analysis, J. Geophys. Res. Oceans, 126, https://doi.org/10.1029/2021JC017794, 2021.

Riahi, K., van Vuuren, D. P., Kriegler, E., Edmonds, J., O'Neill, B. C., Fujimori, S., Bauer, N., Calvin, K., Dellink, R., Fricko, O., Lutz, W., Popp, A., Cuaresma, J. C., KC, S., Leimbach, M., Jiang, L., Kram, T., Rao, S., Emmerling, J., Ebi, K., Hasegawa, T., Havlik, P., Humpenöder, F., Da Silva, L. A., Smith, S., Stehfest, E., Bosetti, V., Eom, J., Gernaat, D., Masui, T., Rogelj, J., Strefler, J., Drouet, L., Krey, V., Luderer, G., Harmsen, M., Takahashi, K., Baumstark, L., Doelman, J. C., Kainuma, M., Klimont, Z., Marangoni, G., Lotze-Campen, H., Obersteiner, M., Tabeau, A., and Tavoni, M.: The Shared Socioeconomic Pathways and their energy, land use, and greenhouse gas emissions implications: An overview, Global Environmental Change, 42, 153–168, https://doi.org/10.1016/j.gloenvcha.2016.05.009, 2017.

Ridder, N., Pitman, A., and Ukkola, A.: Do CMIP6 Climate Models simulate Global or Regional Compound Events skilfully?, Geophys. Res. Lett., 48, https://doi.org/10.1029/2020GL091152, 2021.

Ridder, N., Ukkola, A., Pitman, A., and Perkins-Kirkpatrick, S.: Increased occurrence of high impact compound events under climate change, NPJ Clim. Atmos. Sci., 5, https://doi.org/10.1038/s41612-021-00224-4, 2022.

Robin, Y., Vrac, M., Naveau, P., and Yiou, P.: Multivariate stochastic bias corrections with optimal transport, Hydrol. Earth Syst. Sci., 23, 773–786, https://doi.org/10.5194/hess-23-773-2019, 2019.

Ruosteenoja, K., Räisänen, J., Venäläinen, A., and Kämäräinen, M.: Projections for the duration and degree days of the thermal growing season in Europe derived from CMIP5 model output, Int. J. Climatol., 36, 3039–3055, https://doi.org/10.1002/joc.4535, 2016.

Russo, S., Sillmann, J., and Sterl, A.: Humid heat waves at different warming levels, Sci. Rep., 7, https://doi.org/10.1038/s41598-017-07536-7, 2017.

Sadegh, M., Ragno, E., and AghaKouchak, A.: Multivariate Copula Analysis Toolbox (MvCAT): Describing dependence and underlying uncertainty using a Bayesian framework, Water Resour. Res., 53, 5166–5183, https://doi.org/10.1002/2016WR020242, 2017.

Salvadori, G., de michele, C., Kottegoda, N., and Rosso, R.: Extremes in Nature: An Approach Using Copulas, Springer, Dordrecht, the Netherlands, https://doi.org/10.1007/1-4020-4415-1, 2007.

Salvadori, G., De Michele, C., and Durante, F.: On the return period and design in a multivariate framework, Hydrol. Earth Syst. Sci., 15, 3293–3305, https://doi.org/10.5194/hess-15-3293-2011, 2011.

Salvadori, G., Durante, F., De Michele, C., Bernardi, M., and Petrella, L.: A multivariate copula-based framework for dealing with hazard scenarios and failure probabilities, Water Resour. Res., 52, 3701–3721, https://doi.org/10.1002/2015WR017225, 2016.

Schepsmeier, U., Stoeber, J., Brechmann, E. C., Graeler, B., Nagler, T., and Erhardt, T.: VineCopula: Statistical inference of vine copulas, https://github.com/tnagler/VineCopula, r package version 2.0.5, 2016.

Schölzel, C. and Friederichs, P.: Multivariate non-normally distributed random variables in climate research - introduction to the copula approach, Nonlinear Process. Geophys., 15, 761–772, https://doi.org/10.5194/npg-15-761-2008, 2008.

Schär, C.: Climate extremes: The worst heat waves to come, Nat. Clim. Chang., 6, https://doi.org/10.1038/nclimate2864, 2015.

Serinaldi, F.: Dismissing return periods!, Stoch. Environ. Res. Risk Assess., 29, https://doi.org/10.1007/s00477-014-0916-1, 2014.

Serinaldi, F.: Can we tell more than we can know? The limits of bivariate drought analyses in the United States, Stoch. Environ. Res. Risk Assess., 30, 1691–1704, 2015.

Sgubin, G., Swingedouw, D., Dayon, G., Garcia de Cortazar-Atauri, I., Ollat, N., Page, C., and van Leeuwen, C.: The risk of tardive frost damage in French vineyards in a changing climate, Agric. For. Meteorol., 250–251, 226–242, https://doi.org/10.1016/j.agrformet.2017.12.253, 2018.

Shepherd, T. G.: A Common Framework for Approaches to Extreme Event Attribution, Curr. Clim. Change Rep., 2, 28–38, https://doi.org/10.1007/s40641-016-0033-y, 2016.

Shiau, J.: Return Period of Bivariate Distributed Hydrological Events, Stoch. Environ. Res. Risk Assess., 17, 42–57, https://doi.org/10.1007/s00477-003-0125-9, 2003.

Shiogama, H., Abe, M., and Tatebe, H.: MIROC MIROC6 model output prepared for CMIP6 ScenarioMIP, https://doi.org/10.22033/ESGF/CMIP6.898, 2019.

Singh, H., Najafi, M., and Cannon, A.: Characterizing non-stationary compound extreme events in a changing climate based on large-ensemble climate simulations, Clim. Dynam., 56, 1–17, https://doi.org/10.1007/s00382-020-05538-2, 2021a.

Singh, J., Ashfaq, M., Skinner, C. B., Anderson, W. B., and Singh, D.: Amplified risk of spatially compounding droughts during co-occurrences of modes of natural ocean variability, NPJ Clim. Atmos. Sci., 4, 7, https://doi.org/10.1038/s41612-021-00161-2, 2021b.

Skaugen, T. E. and Tveito, O. E.: Growing-season and degree-day scenario in Norway for 2021-2050, Clim. Res., 26, 221–232, 2004.

Sklar, A.: Fonctions de Répartition à n Dimensions et Leurs Marges, Publications de l'Institut Statistique de l'Université de Paris, 8, 229–231, 1959.

Stott, P. A., Stone, D. A., and Allen, M. R.: Human contribution to the European heatwave of 2003, Nature, 432, 610–614, https://doi.org/10.1038/nature03089, 2004.

Stott, P. A., Christidis, N., Otto, F. E. L., Sun, Y., Vanderlinden, J.-P., van Oldenborgh, G. J., Vautard, R., von Storch, H., Walton, P., Yiou, P., and Zwiers, F. W.: Attribution of extreme weather and climate-related events, Wiley Interdiscip. Rev. Clim. Change, 7, 23–41, https://doi.org/10.1002/wcc.380, 2016.

Swart, N. C., Cole, J. N., Kharin, V. V., Lazare, M., Scinocca, J. F., Gillett, N. P., Anstey, J., Arora, V., Christian, J. R., Jiao, Y., Lee, W. G., Majaess, F., Saenko, O. A., Seiler, C., Seinen, C., Shao, A., Solheim, L., von Salzen, K., Yang, D., Winter, B., and Sigmond, M.: CCCma CanESM5 model output prepared for CMIP6 ScenarioMIP, https://doi.org/10.22033/ESGF/CMIP6.1317, 2019.

Tavakol, A., Rahmani, V., and Jr, J.: Probability of compound climate extremes in a changing climate: A copula-based study of hot, dry, and windy events in the central United States, Environ. Res. Lett., 15, https://doi.org/10.1088/1748-9326/abb1ef, 2020.

Unterberger, C., Brunner, L., Nabernegg, S., Steininger, K. W., Steiner, A. K., Stabentheiner, E., Monschein, S., and Truhetz, H.: Spring frost risk for regional apple production under a warmer climate, PLOS ONE, 13, 1–18, https://doi.org/10.1371/journal.pone.0200201, 2018.

Vautard, R., van Oldenborgh, G. J., Bonnet, R., Li, S., Philip, S., Soubeyroux, J. M., Dubuisson, B., Viovy, N., Reichstein, M., and Otto, F. E. L.: Human influence on growing period frosts like the early april 2021 in Central France, 2021.

Venzon, D. J. and Moolgavkar, S. H.: A Method for Computing Profile-Likelihood-Based Confidence Intervals, J. R. Stat. Soc. Ser. C Appl. Stat., 37, 87–94, 1988.

Villalobos-Herrera, R., Bevacqua, E., Ribeiro, A. F. S., Auld, G., Crocetti, L., Mircheva, B., Ha, M., Zscheischler, J., and De Michele, C.: Towards a compound-event-oriented climate model evaluation: a decomposition of the underlying biases in multivariate fire and heat stress hazards, Nat. Hazards Earth Syst. Sci., 21, 1867–1885, https://doi.org/10.5194/nhess-21-1867-2021, 2021.

Voldoire, A.: CNRM-CERFACS CNRM-CM6-1 model output prepared for CMIP6 CMIP, https://doi.org/10.22033/ESGF/CMIP6.1375, 2018.

Voldoire, A.: CNRM-CERFACS CNRM-CM6-1-HR model output prepared for CMIP6 ScenarioMIP ssp585, https://doi.org/10.22033/ESGF/CMIP6.4225, 2019.

Volodin, E., Mortikov, E., Gritsun, A., Lykossov, V., Galin, V., Diansky, N., Gusev, A., Kostrykin, S., Iakovlev, N., Shestakova, A., and Emelina, S.: INM INM-CM4-8 model output prepared for CMIP6 ScenarioMIP, https://doi.org/10.22033/ESGF/CMIP6.12321, 2019a.

Volodin, E., Mortikov, E., Gritsun, A., Lykossov, V., Galin, V., Diansky, N., Gusev, A., Kostrykin, S., Iakovlev, N., Shestakova, A., and Emelina, S.: INM INM-CM5-0 model output prepared for CMIP6 ScenarioMIP ssp585, https://doi.org/10.22033/ESGF/CMIP6.12338, 2019b.

Vrac, M.: Multivariate bias adjustment of high-dimensional climate simulations: the Rank Resampling for Distributions and Dependences (R^2D^2) bias correction, Hydrol. Earth Syst. Sci., 22, 3175–3196, https://doi.org/10.5194/hess-22-3175-2018, 2018.

Vrac, M. and Thao, S.: R^2D^2 v2.0: Accounting for temporal dependences in multivariate bias correction via analogue ranks resampling, Geosci. Model Dev., 2020, 1–29, https://doi.org/10.5194/gmd-2020-132, 2020.

Vrac, M., Chédin, A., and Diday, E.: Clustering a Global Field of Atmospheric Profiles by Mixture Decomposition of Copulas, J. Atmos. Ocean Technol., 22, 1445–1459, https://doi.org/10.1175/JTECH1795.1, 2005.

Vrac, M., Drobinski, P., Merlo, A., Herrmann, M., Lavaysse, C., Li, L., and Somot, S.: Dynamical and statistical downscaling of the French Mediterranean climate: uncertainty assessment, Nat. Hazards Earth Syst. Sci., 12, 2769–2784, https://doi.org/10.5194/nhess-12-2769-2012, 2012.

Vrac, M., Thao, S., and Yiou, P.: Changes in temperature-precipitation correlations over Europe: Are climate models reliable?, https://doi.org/10.21203/rs.3.rs-1008080/v1, 2021.

Vrac, M., Thao, S., and Yiou, P.: Should multivariate bias corrections of climate simulations account for changes of rank correlation over time?, https://doi.org/10.1002/essoar.10510318.1, 2022.

Wahl, T., Jain, S., Bender, J., Meyers, S., and Luther, M.: Increasing risk of compound flooding from storm surge and rainfall for major US cities, Nat. Clim. Chang., 5, 1093–1097, https://doi.org/10.1038/nclimate2736, 2015.

White, H.: Maximum Likelihood Estimation of Misspecified Models, Econometrica, 50, 1–25, 1982.

Wieners, K.-H., Giorgetta, M., Jungclaus, J., Reick, C., Esch, M., Bittner, M., Gayler, V., Haak, H., de Vrese, P., Raddatz, T., Mauritsen, T., von Storch, J.-S., Behrens, J., Brovkin, V., Claussen, M., Crueger, T., Fast, I., Fiedler, S., Hagemann, S., Hohenegger, C., Jahns, T., Kloster, S., Kinne, S., Lasslop, G., Kornblueh, L., Marotzke, J., Matei, D., Meraner, K., Mikolajewicz, U., Modali, K., Müller, W., Nabel, J., Notz, D., Peters-von Gehlen, K., Pincus, R., Pohlmann, H., Pongratz, J., Rast, S., Schmidt, H., Schnur, R., Schulzweida, U., Six, K., Stevens, B., Voigt, A., and Roeckner, E.: MPI-M MPI-ESM1.2-LR model output prepared for CMIP6 ScenarioMIP ssp585, https://doi.org/10.22033/ESGF/CMIP6.6705, 2019.

Yue, S. and Rasmussen, P.: Bivariate frequency analysis: Discussion of some useful concepts in hydrological application, Hydrol. Process., 16, 2881–2898, https://doi.org/10.1002/hyp.1185, 2002.

Yukimoto, S., Koshiro, T., Kawai, H., Oshima, N., Yoshida, K., Urakawa, S., Tsujino, H., Deushi, M., Tanaka, T., Hosaka, M., Yoshimura, H., Shindo, E., Mizuta, R., Ishii, M., Obata, A., and Adachi, Y.: MRI MRI-ESM2.0 model output prepared for CMIP6 CMIP, https://doi.org/10.22033/ESGF/CMIP6.621, 2019.

Zscheischler, J. and Lehner, F.: Attributing compound events to anthropogenic climate change, Bull. Amer. Meteor. Soc., pp. 1–45, https://doi.org/10.1175/BAMS-D-21-0116.1, 2021.

Zscheischler, J. and Seneviratne, S.: Dependence of drivers affects risks associated with compound events, Sci. Adv, 3, e1700 263, https://doi.org/10.1126/sciadv.1700263, 2017.

Zscheischler, J., Michalak, A. M., Schwalm, C., Mahecha, M. D., Huntzinger, D. N., Reichstein, M., Berthier, G., Ciais, P., Cook, R. B., El-Masri, B., Huang, M., Ito, A., Jain, A., King, A., Lei, H., Lu, C., Mao, J., Peng, S., Poulter, B., Ricciuto, D., Shi, X., Tao, B., Tian, H., Viovy, N., Wang, W., Wei, Y., Yang, J., and Zeng, N.: Impact of large-scale climate extremes on biospheric carbon fluxes: An intercomparison based on MsTMIP data, Glob. Biogeochem. Cycles, 28, 585–600, https://doi.org/10.1002/2014GB004826, 2014.

Zscheischler, J., Westra, S., Hurk, B., Seneviratne, S., Ward, P., Pitman, A., AghaKouchak, A., Bresch, D., Leonard, M., Wahl, T., and Zhang, X.: Future climate risk from compound events, Nat. Clim. Chang., pp. 469–477, https://doi.org/10.1038/s41558-018-0156-3, 2018.

Zscheischler, J., Fischer, E., and Lange, S.: The effect of univariate bias adjustment on multivariate hazard estimates, Earth Syst. Dynam., 10, 31–43, https://doi.org/10.5194/esd-10-31-2019, 2019.

Zscheischler, J., Martius, O., Westra, S., Bevacqua, E., Raymond, C., Horton, R., Hurk, B., AghaKouchak, A., Jézéquel, A., Mahecha, M., Maraun, D., Ramos, A., Ridder, N., Thiery, W., and Vignotto, E.: A typology of compound weather and climate events, Nat. Rev. Earth Environ., 1, 1–5, 2020.

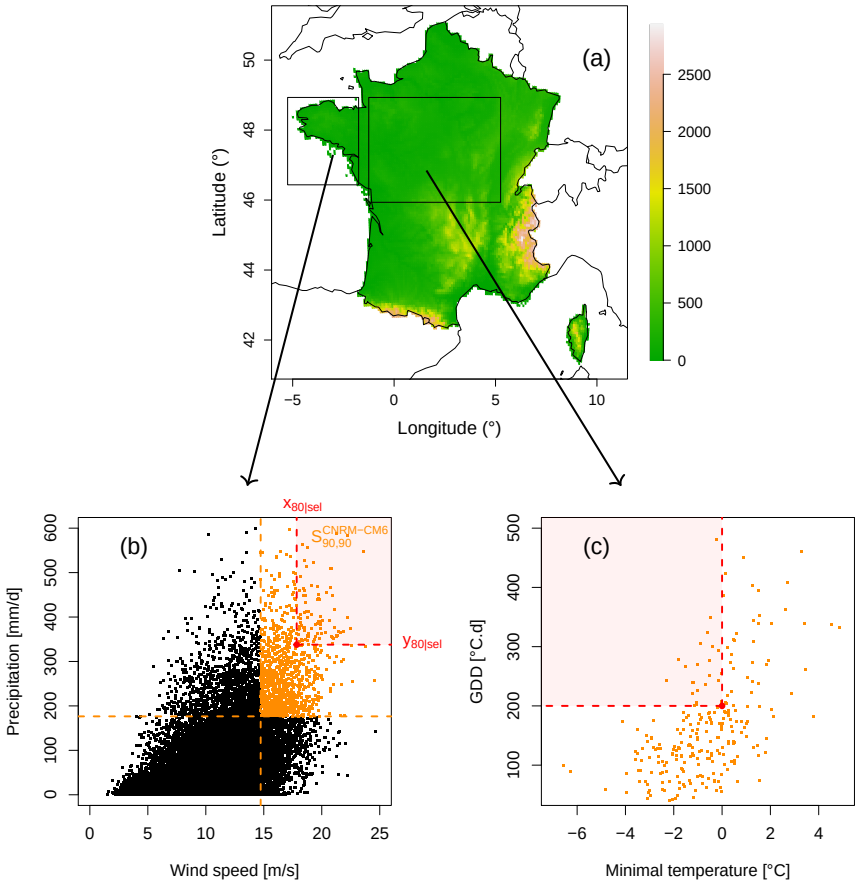

Figure 1. (a) Map of France with the regions of interest in boxes. Scatterplots of CNRM-CM6 (b) DJF compounding wind and precipitation in Brittany and (c) miniminal temperature in April and GDD values by the end of March over Central France for the 1871-2100 period. Parametric fitting for marginal and dependence over the 30-years sliding windows spanning the 1871-2100 period are performed to bivariate points in orange. For compounding wind and precipitation, these points correspond to high values of wind and precipitation data belonging to $S_{90,90}^{\text{CNRM-CM6}}$, i.e. simultaneously exceeding the individual 90th percentiles of the 1871-1900 reference period. Bivariate exceedance probabilities are then computed for varying exceedance thresholds between the 5th and 95th percentile of wind speed and precipitation already belonging to $S_{90,90}^{\text{CNRM-CM6}}$ (for more details, see Sect. 4). The red area contains bivariate points exceeding the 80th percentiles of points already belonging to $S_{90,90}^{\text{CNRM-CM6}}$. For growing-period frosts, exceedance thresholds of interest for minimal temperature and GDD index are fixed to values of 0°C and 200°C.day, respectively.

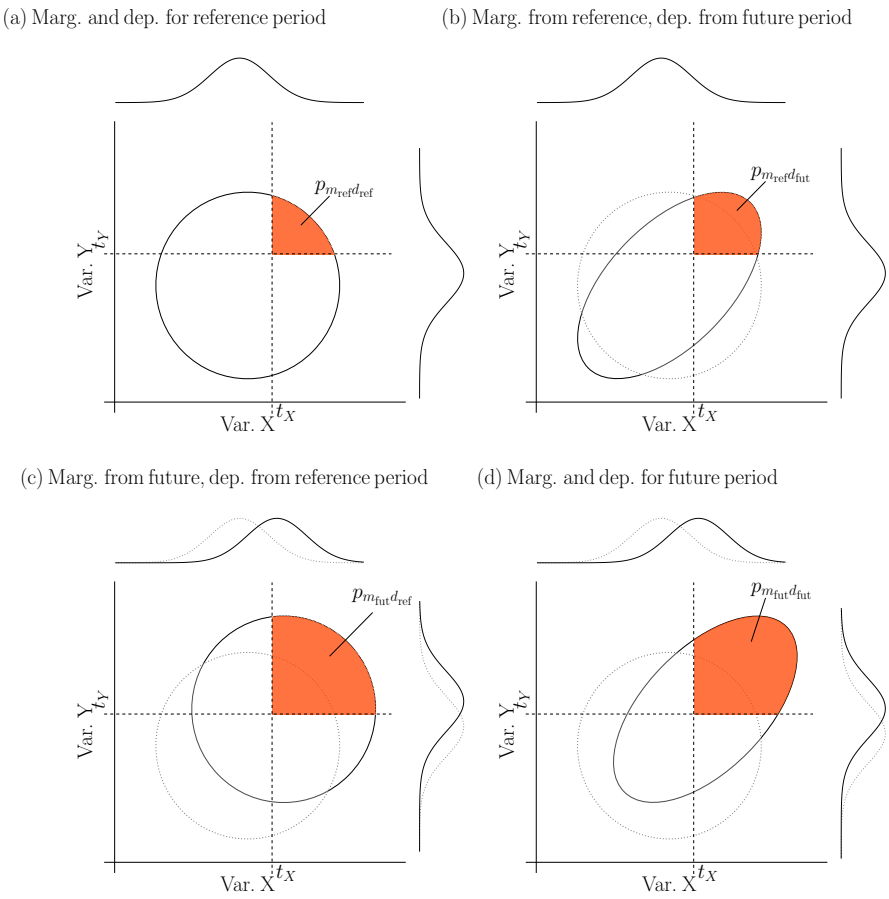

Figure 2. Illustration of the influence of marginal and dependence properties on bivariate exceedance probabilities for an artifical distribution of two contributing variables X and Y during (a) the reference period and (d) a future period with a shift in means and an increase in dependence between the variables. The distribution of the two contributing variables (b) with marginal properties from the reference period and dependence structure from the future period, and (c) with marginal properties from the future period and dependence structure from the reference period. Orange areas show bivariate exceedance probabilities for the thresholds (t_X, t_Y) of the two contributing variables.

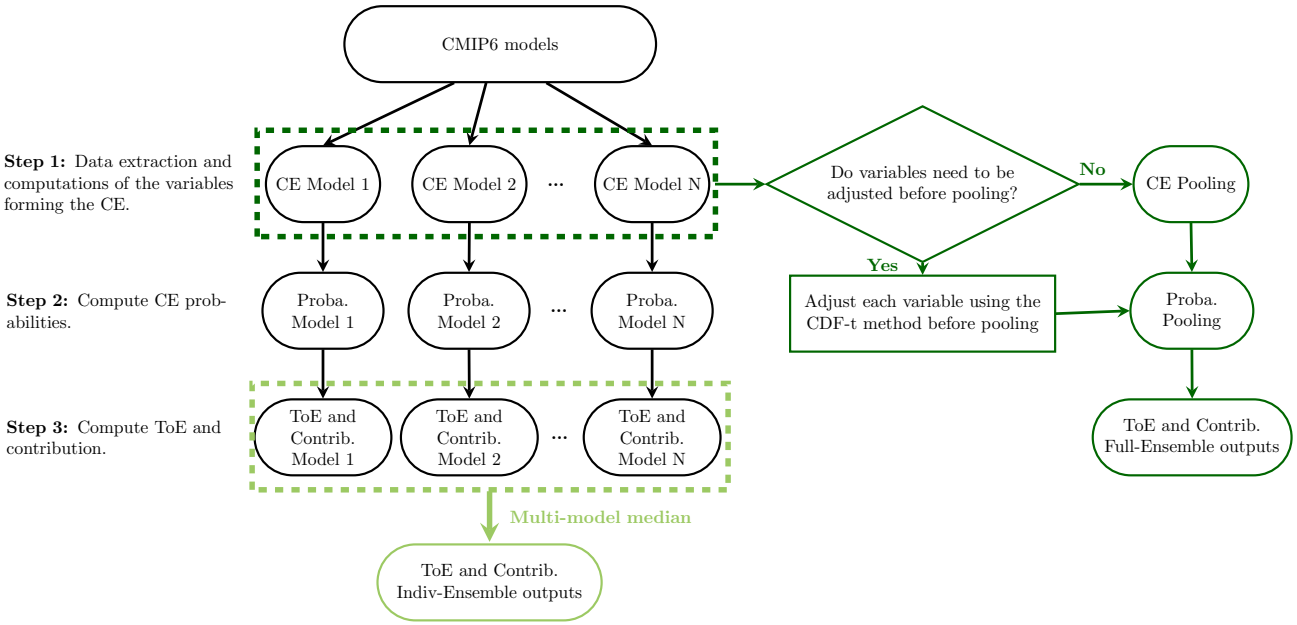

Figure 3. Flowchart for the computations of time of emergence and contributions for Indiv- and Full-Ensemble versions.

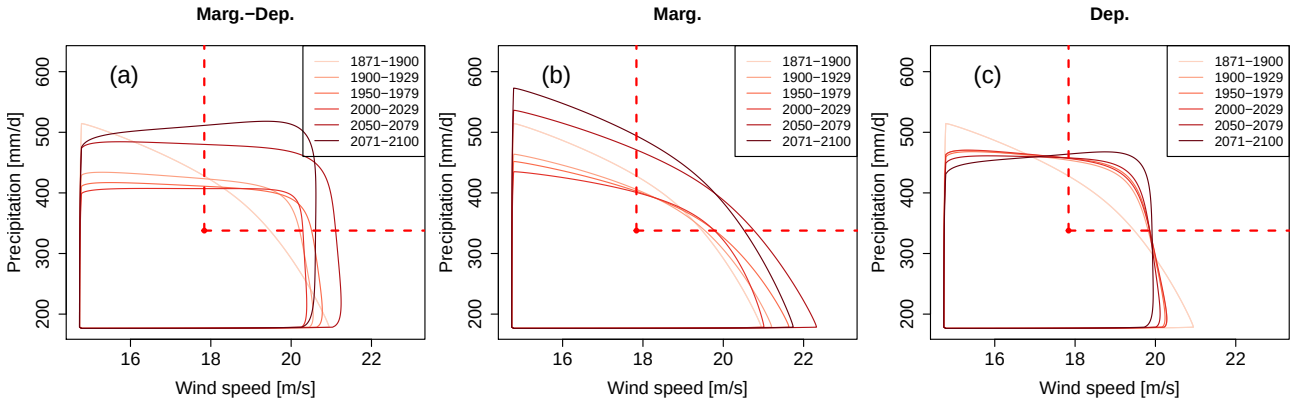

Figure 4. Change of winter (December-to-February) bivariate wind and precipitation extremes distributions in Brittany based on CNRM-CM6 simulations due to (a) future marginal and dependence changes ("Marg.-dep."), (b) future marginal changes while keeping dependence of the reference period ("Marg.") and (c) dependence changes while keeping marginal of the reference period ("Dep."). For the bivariate distributions, contour lines encompassing 90 % of all data points are shown. A selection of six 30-years sliding windows is presented using a color gradient from light (1871-1900) to dark (2071-2100). The red dashed lines characterises the bivariate exceeding thresholds defined here as the 80th quantile of each variable.

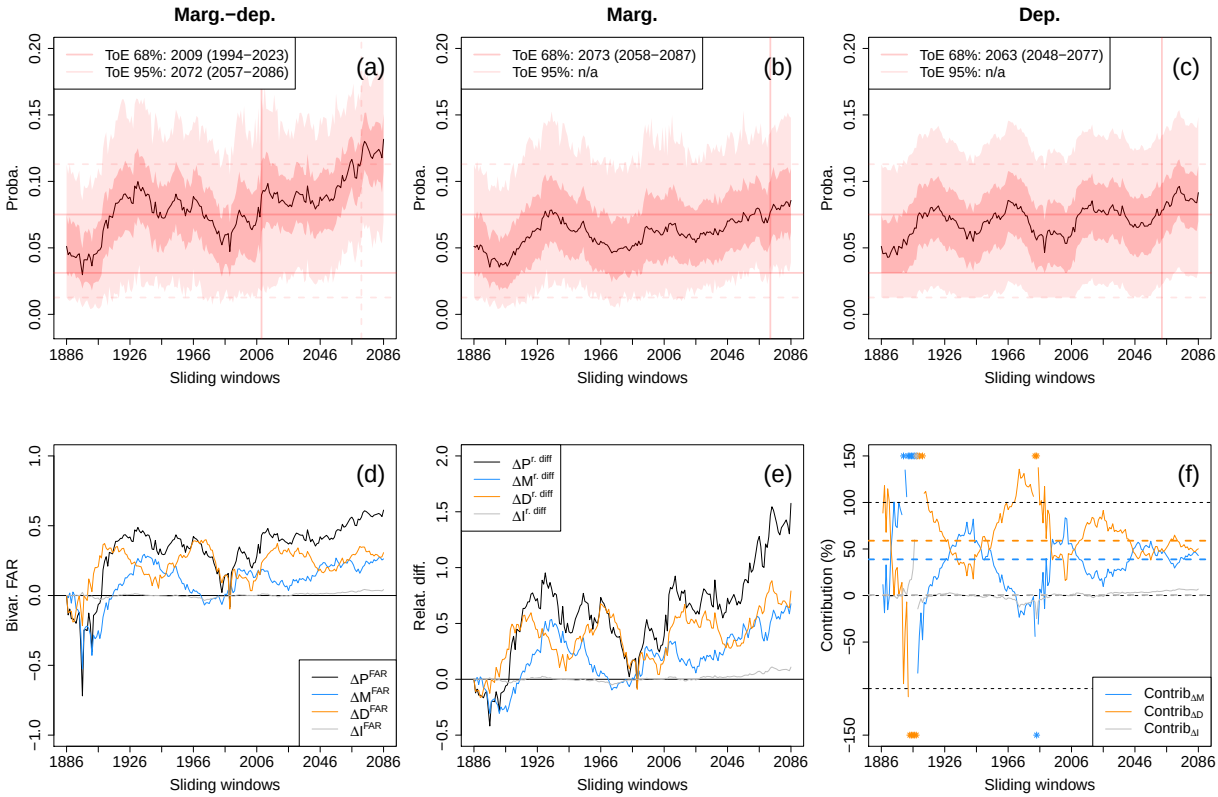

Figure 5. (a-c) Probability changes and time of emergence of compound wind and precipitation extremes ($\mathbb{P}(X > x_{80\text{lsel}} \cap Y > y_{80\text{lsel}} \mid (X,Y) \in S_{90,90}^{\text{CNRM-CM6}}$) based on CNRM-CM6 simulations due to changes of (a) both marginal and dependence properties, (b) marginal properties only, and (c) dependence properties only. The shaded bands indicate 68% and 95% confidence intervals of the probabilities. Evolutions of (d) the bivariate fraction of attributable risk (FAR), (e) relative difference of probabilities with respect to the reference period (1871-1900) and (f) contribution of the marginal, dependence and interaction terms to probability values. Median contributions computed over all sliding windows are displayed with dashed lines. Asterisks indicate values lying outside the plotted range. Not-applicable (n/a) is indicated when no time of emergence is detected.

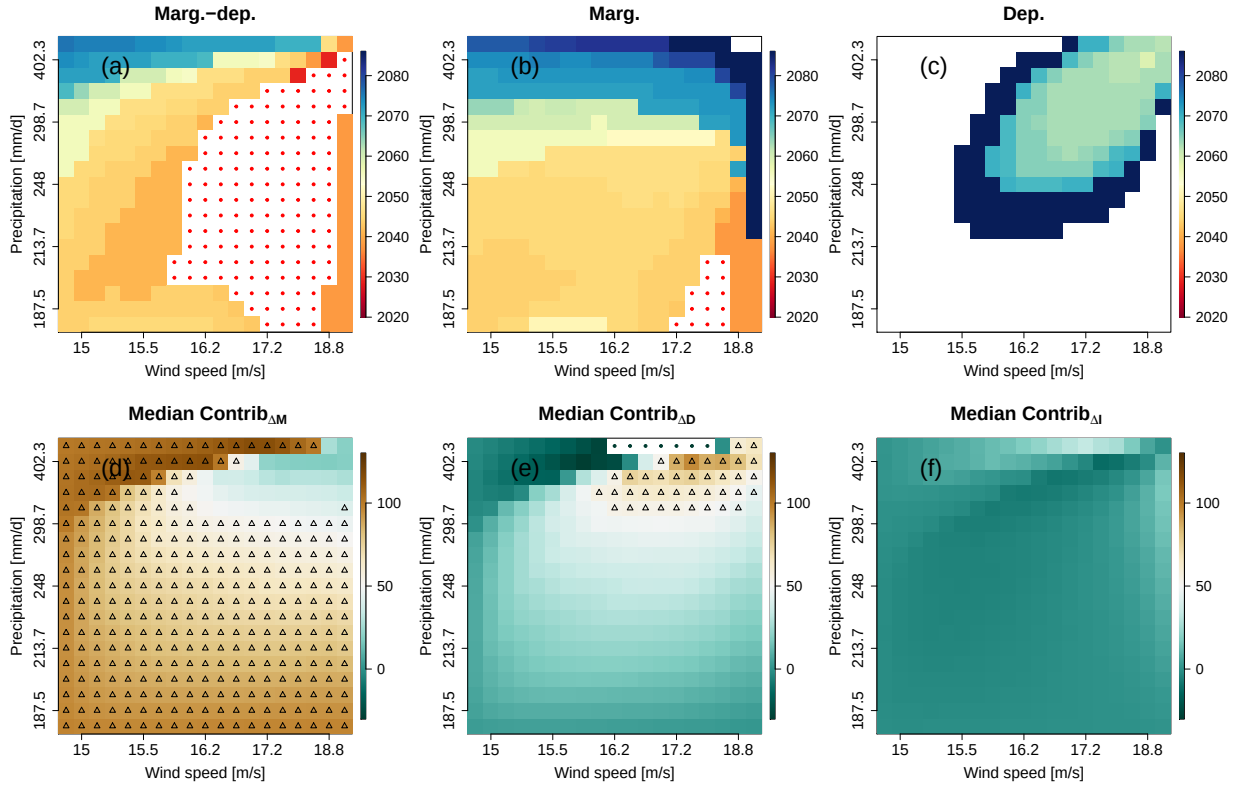

Figure 6. CNRM-CM6 (a-c) time of emergence (at 68% confidence level) for compound wind and precipitation extremes due to changes of (a) both marginal and dependence properties, (b) marginal properties only, and (c) dependence properties only. White cells indicate that no time of emergence is detected, while white cells with red points indicate ToE values before 2020. (d-f) Matrices of median contributions of the (d) marginal, (e) dependence and (f) interaction terms. Results are presented for varying exceedance thresholds between the 5th and 95th percentile of compound wind and precipitation extremes data. Upper triangles show where contribution ≥ 50 %.

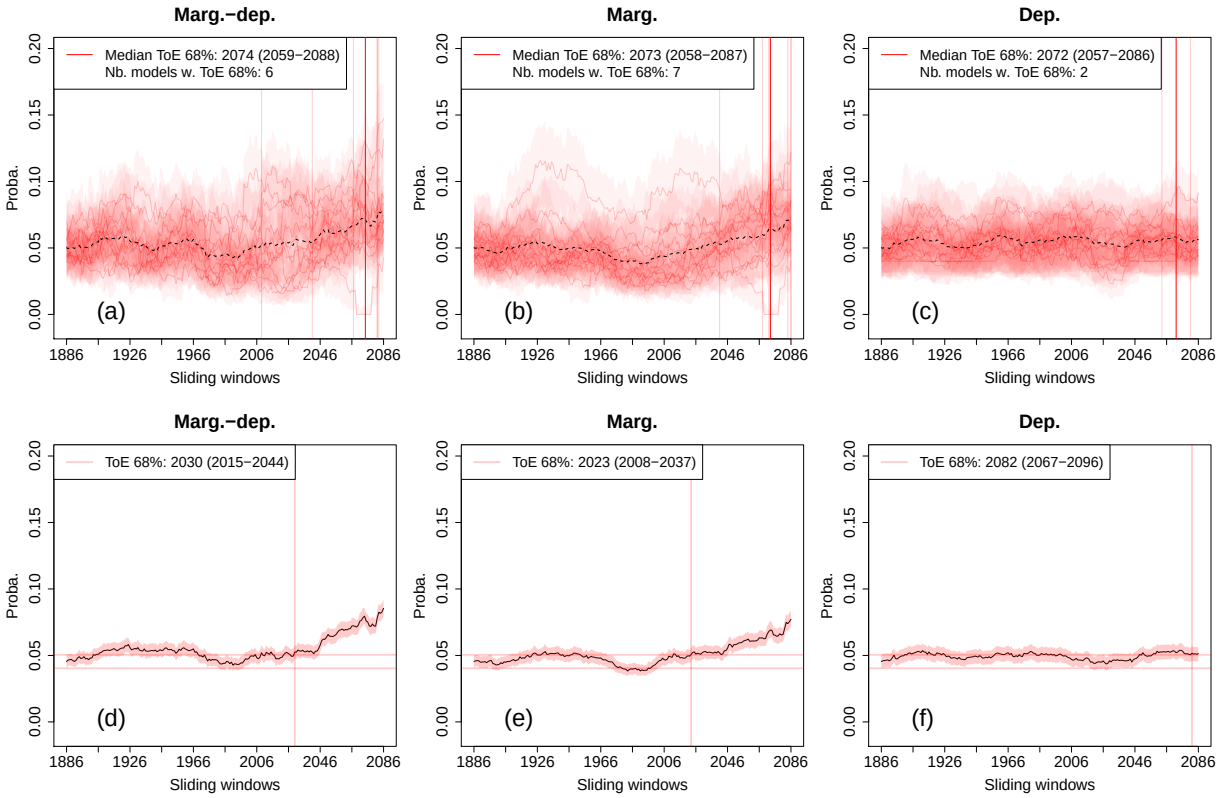

Figure 7. Probability changes and time of emergence (at 68%) of compound wind and precipitation extremes (exceeding the individual 80th percentiles of selected points of high values) for (a-c) Indiv- and (d-f) Full-Ensemble versions due to changes of (a,d) both marginal and dependence properties, (b,e) marginal properties only, and (c,f) dependence properties only. The shaded bands indicate 68% confidence intervals of the probabilities. For (a-c), individual time of emergence for the different models within the ensemble are displayed when defined (vertical light red lines), as well as the corresponding median time of emergence (vertical red line). For information purpose, multi-model mean exceedance probability time series are also plotted (black dotted lines).

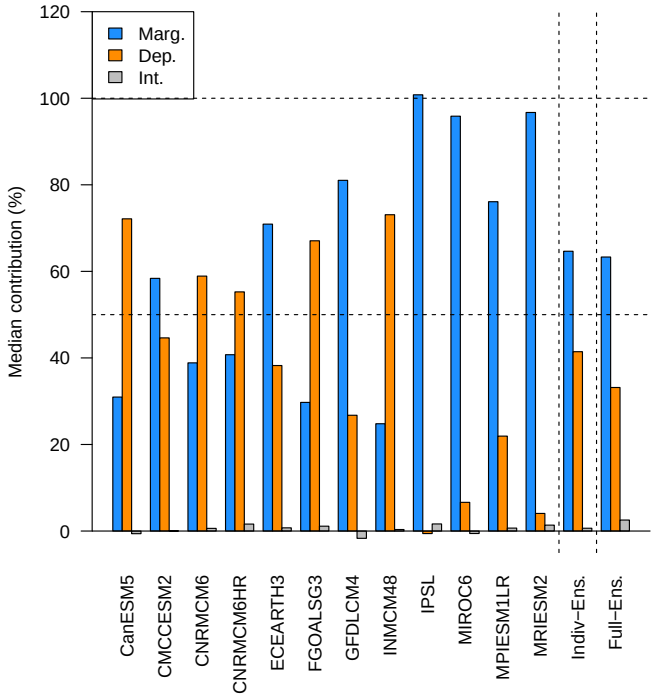

Figure 8. Median contribution, over all sliding windows, of the marginal, dependence and interaction terms to overall probability changes for the 12 individual CMIP6 models, and for Indiv- and Full-Ensemble versions.

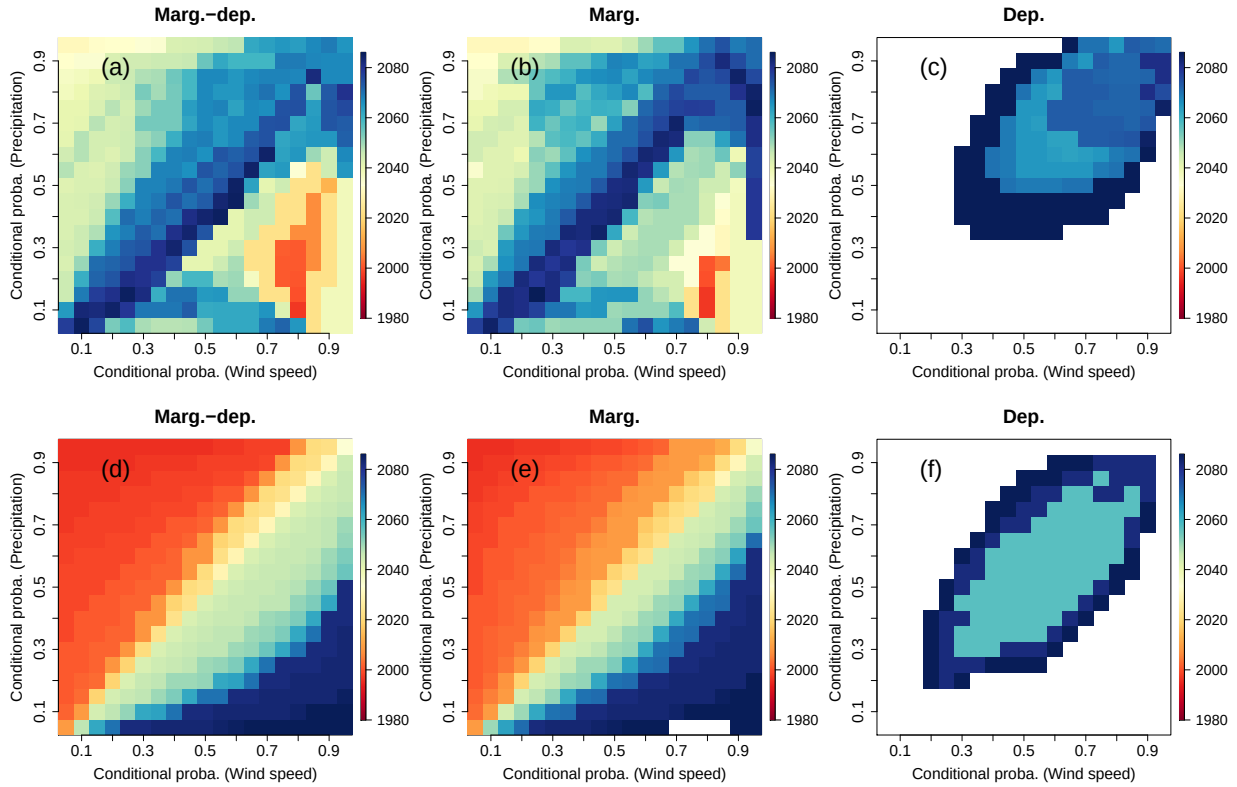

Figure 9. Time of Emergence (at 68% confidence level) matrices of compound wind and precipitation extremes due to changes of (a, d) both marginal and dependence properties, (b, e) marginal properties only, and (c, f) dependence properties only. Results are displayed for (a-c) the Indiv- and (d-f) Full-Ensemble versions for varying exceedance thresholds between the 5th and 95th percentile of compound wind and precipitation extremes data. For each subplot, white indicates that no time of emergence is detected.

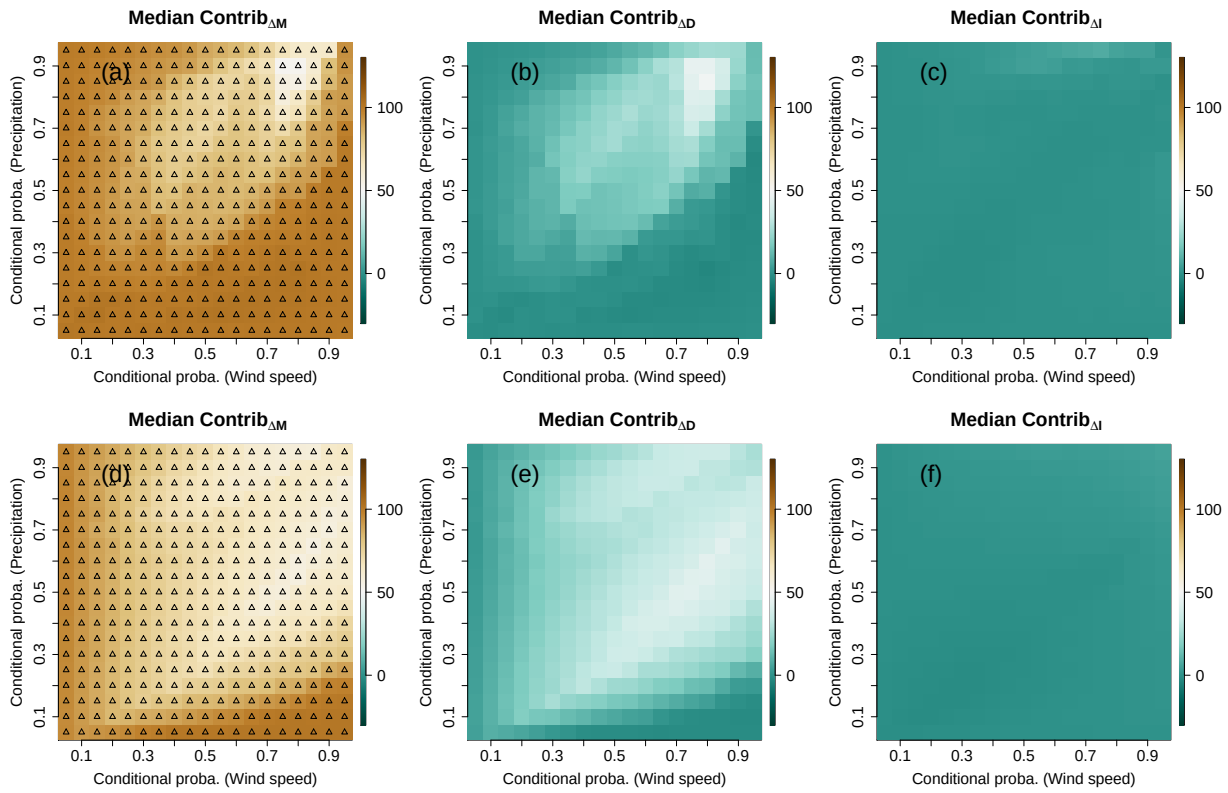

Figure 10. Median contributions of (a, d) marginal, (b, e) dependence and (c, f) interaction terms for (a-c) Indiv- and (d-f) Full-Ensemble versions. Results are presented for compound wind and precipitation extremes with varying exceedance thresholds between the 5th and 95th percentile. Upper triangles show where contribution ≥ 50 %.

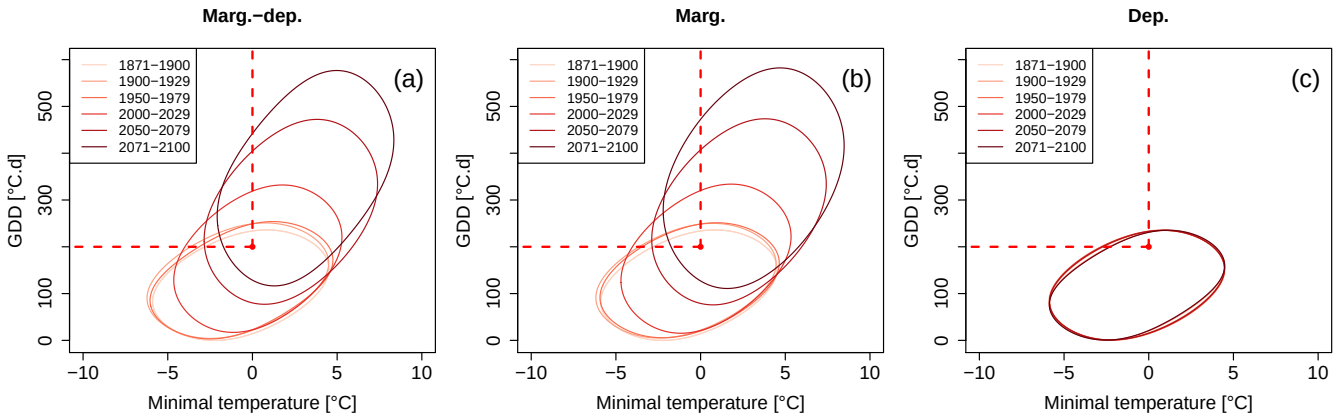

Figure 11. Changes of minimal temperature vs. GDD distributions in Central France for the Full-Ensemble version due to (a) marginal and dependence changes ("Marg.-dep."), (b) marginal changes while keeping dependence of the reference period ("Marg.") and (c) dependence changes while keeping marginal of the reference period ("Dep."). For the bivariate distributions, contour lines encompassing 90 % of all data points are shown. A selection of six 30-years sliding windows is presented using a color gradient from light (1871-1900) to dark (2071-2100).

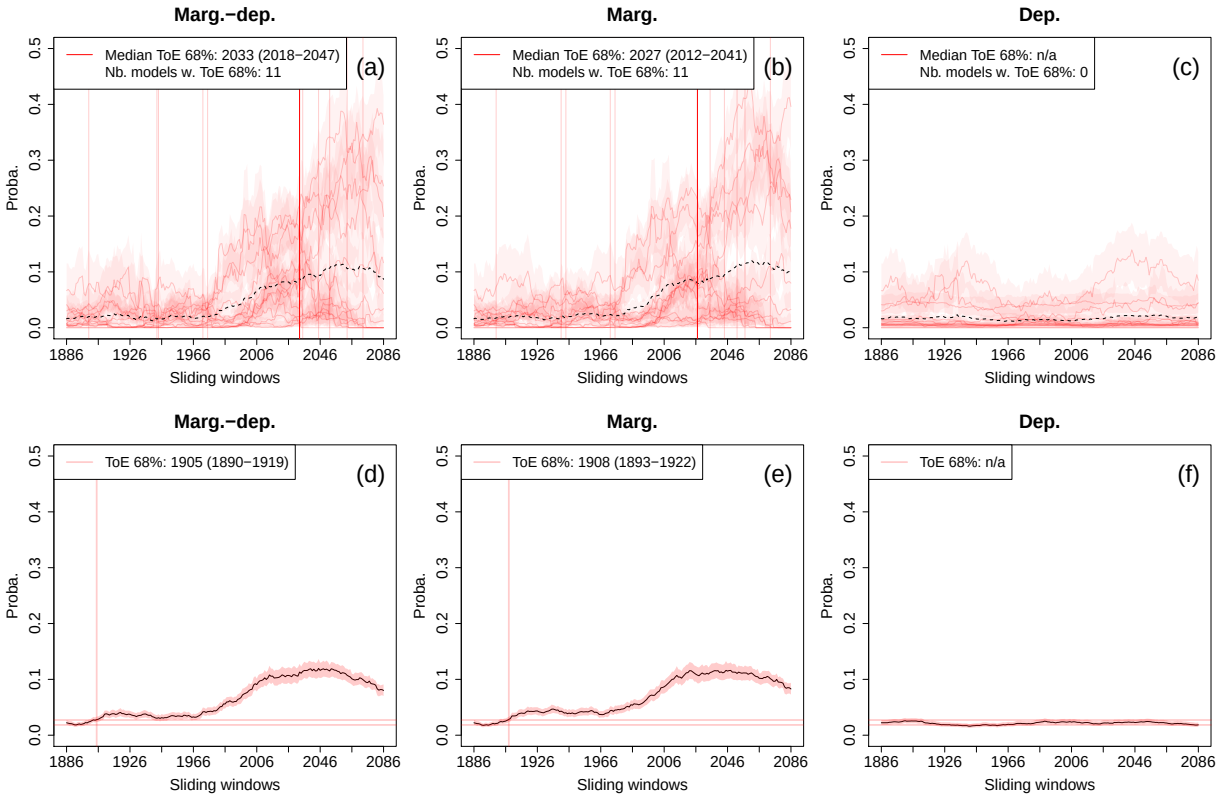

Figure 12. Probability changes and times of emergence (at 68%) of growing-period frosts (GDD $\geq$ 200 °C.d and minimal temperatures $\leq$ 0 °C) for (a-c) Indiv- and (d-f) Full-Ensemble versions due to changes of (a, d) both marginal and dependence properties, (b, e) marginal properties only, and (c, f) dependence properties only. The shaded envelops indicate 68% confidence intervals of the probabilities. For (a-c), individual time of emergence for the different models within the ensemble are displayed when defined (vertical light red lines), as well as the corresponding median time of emergence (vertical red line). For information purpose, multi-model mean exceedance probability time series are also plotted (black dashed lines). Not-applicable (n/a) is indicated when no time of emergence is detected.

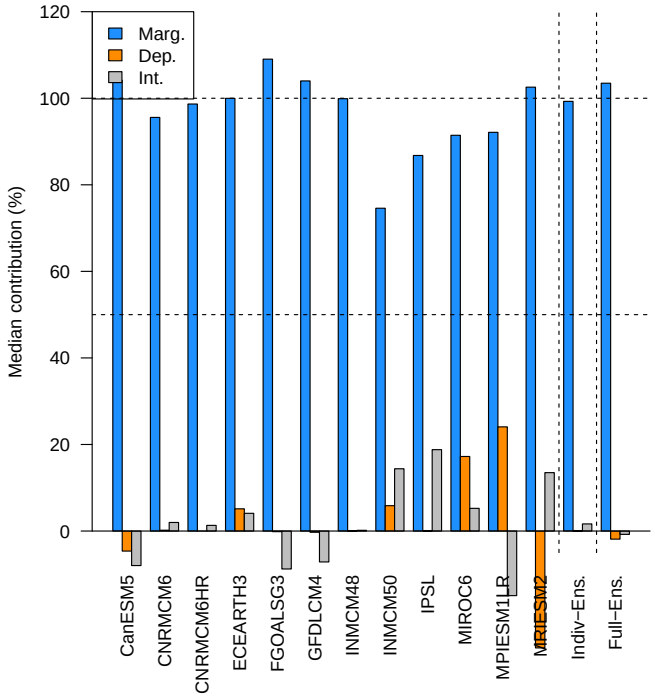

Figure 13. Median contribution, over all sliding windows, of the marginal, dependence and interaction terms to overall probability changes for the 12 individual CMIP6 models, and for Indiv- and Full-Ensemble versions.

Table 1. List of CMIP6 simulations used in this study, their run, approximate horizontal resolution and references.

Model	Institution	Spatial res. (lon. $\times$ lat.)	Data reference
CanESM5	Canadian Centre for Climate Modelling and Analysis, Canada	$2.81^\circ \times 2.81^\circ$	Swart et al. (2019)
FGOALS-g3	Chinese Academy of Sciences, China	$2.00^\circ \times 2.25^\circ$	Li (2019)
CNRM-CM6-1	Centre National de Recherches Meteorologiques, Meteo-France, France	$1.41^\circ \times 1.41^\circ$	Voldoire (2019)
CNRM-CM6-1-HR	Centre National de Recherches Meteorologiques, Meteo-France, France	$0.50^\circ \times 0.50^\circ$	Voldoire (2018)
GFDL-CM4	Geophysical Fluid Dynamics Laboratory, USA	$1.25^\circ \times 1^\circ$	Guo et al. (2018)
INM-CM4-8	Institute for Numerical Mathematics, Russia	$2^\circ \times 1.5^\circ$	Volodin et al. (2019a)
INM-CM5-0	Institute for Numerical Mathematics, Russia	$2^\circ \times 1.5^\circ$	Volodin et al. (2019b)
IPSL-CM6A-LR	Institut Pierre-Simon Laplace, France	$2.50^\circ \times 1.26^\circ$	Boucher et al. (2018)
MIROC6	JAMSTEC, AORI, NIES, R-CCS, Japan	$1.41^\circ \times 1.41^\circ$	Shiogama et al. (2019)
MPI-ESM1-2-LR	Max Planck Institute for Meteorology, Germany	$1.88^\circ \times 1.88^\circ$	Wieners et al. (2019)
MRI-ESM2-0	Meteorological Research Institute, Japan	$1.13^\circ \times 1.13^\circ$	Yukimoto et al. (2019)
CMCC-ESM2	Centro Euro-Mediterraneo per i Cambiamenti, Italy	$1.25^\circ \times 0.94^\circ$	Cherchi et al. (2019)
EC-Earth3	EC-Earth-Consortium	$0.70^\circ \times 0.70^\circ$	EC-Earth (2019)

4.3 Summary and conclusions

In this chapter, I propose a new methodology to detect when the changes in CE probabilities are significantly different from the probabilities associated with natural variability (Time of Emergence). Based on copula theory, the proposed methodology allows to separate and quantify the contribution of univariate and dependence properties to the changes of CEs probability. The methodology is applied on a 13-member multi-model ensemble (CMIP6) to two case studies:

- wind and precipitation extremes in Brittany (France).
- growing-period frost events over Central France.

For wind and precipitation extremes, the main results show that the probabilities emerge before the end of the 21st century (median of ToE: 2074) for 6 models in the ensemble. However, the importance of univariate and dependence changes to these increased probabilities can vary depending on climate models. While, for some models, the changes of univariate properties mainly drive probability changes of wind and precipitation extremes, other models present a balanced importance of both univariate and dependence properties to probability changes. It highlights here that, when analysing compound wind and precipitation extremes probability, the dependence structure is potentially important and thus has to be carefully taken into consideration.

For frost events occurring during the growing season, significant changes in probability and emergences are detected for 11 models. Also, a large majority of models agree on the contribution of univariate and dependence properties: univariate properties contribute dominantly to probability changes of growing-period frost events, while contribution of dependence properties is mostly negligible.

Our methodology provides a better understanding of how the statistical properties of the variables leading to CEs evolve and contribute to changes in their occurrence. By applying the methodology to analyse two different compound events, the results permitted to highlight that the importance of univariate and dependence properties to probability of compound events can differ from a compound hazard to another, and from a climate model to another. It thus emphasises the importance of taking into account changes in univariate and dependence properties, as well as their inter-model variability, for future risk assessments of compound events.

Although not mentioned in the study, quantifying the importance of univariate and dependence properties changes has also obvious relevance in a context of bias correction. Indeed, as we have seen in Chapter 2, before applying any bias correction method, end users have to identify the important statistical properties to be reproduced in the correction, such as changes of univariate and/or dependence properties. In this chapter has been highlighted that some models presented a strong importance of both univariate and dependence properties to probability changes of compound wind and precipitation extremes. This non-stationary property can (or not) be desired by practitioners to be kept in the corrections, which therefore eventually direct them towards the use of MBC methods assuming nonstationarity of univariate properties and dependence structures. The perspective of compound events adjustment using MBC methods will be further discussed in Chapter 5.

4.4 Perspectives

The work carried out in this chapter leads to several perspectives in addition to those mentioned in the study. I first recall some of the perspectives already discussed in the submitted article (subsection 6.2) before providing additional perspectives.

4.4.1 A look back at some of the perspectives of the article

A subjective definition of Time of Emergence

In this study, I defined the time of emergence of compound events probability as the first period for which the probability signal permanently goes out from the baseline natural variability (1871-1900 period). This definition of Time of Emergence is subjective and can be modified or improved as needed. For example, choosing 1871-1900 as baseline has been made as this period belongs to the pre-industrial era and thus can be considered relevant for characterising natural climate variability (e.g., Hawkins et al., 2020). One could consider other baseline periods depending on their applications: choosing a more recent baseline period can provide useful information for adaptation planning as it would relate with recent societal experience (Ossó et al., 2022). However, modifying the baseline period would of course change the estimated natural variability, which would affect the results for both ToE of compound events probabilities and contributions from univariate and dependence changes. Also, defining the time of emergence of a signal when it *permanently* excesses a specific threshold can be quite limited. Indeed, the climate system is highly complex and non-monotonic, and relaxing this constraint of "persistent" changes in the definition of ToE to instead prefer detecting "periods of emergence" (PoE) can be interesting to rather describe periods where signals of probability emerge significantly — but temporarily — from the estimated natural variability.

Considering large-ensembles for more robust results

As mentioned in the perspectives of the study, natural climate variability for compound events probabilities has been characterised by using uncertainty from statistical fitting procedures (see subsection 3.1 of the article). However, uncertainty arisen from natural climate variability, i.e., from the inherent chaotic behaviour of the Earth's climate, could have been defined in a more robust way. The accurate characterisation of these uncertainties is all the more important in decision-making processes (e.g., Raymond et al., 2022; Bevacqua et al., 2022). A more robust way to assess these uncertainties would be to consider large-ensemble simulations, i.e., multiple runs from different climate models, in future works. Applying the methodology to large-ensembles would permit to assess probabilities of hazards, time of emergence, contributions of marginal and dependence properties and their associated uncertainties due to internal climate variability and model-to-model differences.

Attributing univariate and dependence changes to climate change

In this study, ToE of compound events probabilities has been evaluated with respect to a baseline period (1871-1900) used to characterise the natural variability. The probabilities for future periods have then been estimated and compared with respect to the baseline period, allowing to assess the influence of future climate change on probabilities. The underlying philosophy of ToE is relatively similar to attribution, which is a research field in climate science that

seeks to determine the mechanisms responsible for recent global warming, such as anthropogenic greenhouse gas emissions. The attribution framework is however different from that of ToE, as it generally considers two worlds with different forcings: a "factual" world with anthropogenic climate change and a "counterfactual" world in which anthropogenic emissions had never occurred. Adapting and applying our methodology in an attribution framework is an interesting perspective as it would permit to not only attribute changes of compound events to a specific cause, but also to quantify the underlying contributions of the changes in univariate and dependence properties. This will provide an interesting tool that would allow to better characterise the statistical features of climate change, thus complementing the already existing multivariate event attribution framework (e.g., Kiriliouk and Naveau, 2020; Zscheischler and Lehner, 2021).

4.4.2 Additional perspectives

Do climate models realistically simulate compound events?

A natural follow-up of this work is the application of the methodology to reanalyses over the historical period. By comparing the results obtained with those from climate models for the same time period, it would permit to evaluate the capability of climate models to reproduce compound events probability, time of emergence and contributions of marginal and dependence properties. This work on climate simulations and reanalyses could then be extended to analyse other compound events at a larger geographical scale. Considering the global scale would permit to derive a more complete picture of the climatology of compound events and their future evolutions. This would complement the works from Ridder et al. (2020, 2021, 2022) by providing valuable information for policy-makers such as the identification of geographical regions prone to emergence of compound events. Also, for projections, considering alternative climate scenarios to that of the SSP-585 one (Riahi et al., 2017) would permit to assess the sensitivity of ToE results to anthropogenic emissions, which has obvious interest to mitigation policy. In a previous perspective has been mentioned the use of large-ensembles to derive more robust ToE and contributions results. In a context of climate model evaluation, using large-ensembles would also permit to enrich the analysis by attributing discrepancy of compound events probability to biases in internal variability and forced response (Suarez-Gutierrez et al., 2021). This work of model evaluation will permit to further investigate if climate models are able to realistically simulate compound events probabilities and their evolutions. This would provide help for the selection of appropriate models to carry different compound events analyses in different geographical regions for past, current and future climate periods (Krinner and Flanner, 2018).

Towards analysing compound events with high dimensionality

Using copulas, the methodology has been applied to study bivariate climate hazards probability. Extending the methodology to study climate hazards composed of more than two variables is of course an interesting perspective. Although multivariate parametric copulas have already been applied to analyse trivariate climate hazards (e.g., Tavakol et al., 2020), copulas present a lack of flexibility when modelling dependence between more than two variables, as dependencies can be different between the different pairs of variables (Aas et al., 2009). To address this problem, pair-copula constructions (PCCs) have been developed (Aas et al., 2009; Acar et al., 2012; Hobæk Haff, 2012). PCCs decompose the dependence structure into several bivariate copulas and give greater flexibility in modelling complex dependence structures. PCCs have already shown useful in several compound event studies (e.g., Bevacqua et al., 2017; Aghatise et al.,

2021; Li et al., 2021; Lan et al., 2022). Adapting our methodology to study ToE of climate hazards in high dimensionality using PCCs is thus an interesting perspective that would strengthen its use.

Résumé

Contexte et objectifs

Les catastrophes climatiques résultent souvent de la combinaison de plusieurs phénomènes climatiques, également appelés "événements composés" (CEs). Les propriétés univariées et multivariées des phénomènes climatiques formant les CEs sont des propriétés statistiques cruciales caractérisant leurs probabilités d'occurrence. Dans ce chapitre, nous proposons une nouvelle méthodologie pour détecter à partir de quand (quelle période) les changements de probabilités des CEs sont significativement différents par rapport aux probabilités associées à la variabilité naturelle.

Méthodes

La méthodologie proposée, basée sur la théorie des copules, permet de séparer et de quantifier la contribution des propriétés univariées et de dépendance aux changements de probabilité des CEs. La méthodologie est appliquée sur un ensemble multi-modèle de 13 membres (CMIP6) à deux études de cas :

- les extrêmes de vent et de précipitations en Bretagne (France).
- les épisodes de gel survenant pendant la saison de croissance sur le centre de la France, préconditionnés par des températures chaudes.

Résultats

Pour les extrêmes de vent et de précipitations, les résultats montrent que les probabilités émergent avant la fin du 21ème siècle pour 6 modèles de l'ensemble. Pour les épisodes de gel survenant en période de croissance, des changements significatifs de probabilité sont détectés pour 11 modèles. Cependant, la contribution des propriétés univariées et de dépendance à ces changements de probabilités peut être très différente d'un CE à l'autre, et d'un modèle à l'autre. Selon le CE, certains modèles donnent une forte importance à la fois aux propriétés marginales et aux propriétés de dépendance pour les changements de probabilité.

Notre méthodologie permet de mieux comprendre comment les propriétés statistiques des variables conduisant aux CEs évoluent et contribuent au changement de leurs occurrences. Les résultats soulignent l'importance de prendre en compte les changements des propriétés univariées et de dépendance, ainsi que leur variabilité inter-modèles, pour les évaluations futures des risques dus aux événements composés.

Chapter 5

Conclusion

In this chapter, the main results of this thesis will be recalled in the following subsection. I will then explore the possible research directions resulting from the body of work presented in this manuscript.

5.1 Conclusion

Climate being the result of various interactions between physical processes, it is crucial to improve the understanding of the dependence between climate variables. It is all the more crucial as dependence can be a key feature of rare compounding events having the most extreme socio-economic and environmental impacts, but also of more usual and less impacting climate events. **The objective of this thesis was to assess whether and how the use of multivariate statistical approaches can contribute to provide a better understanding and use of climate simulations through the consideration of dependencies between climate variables.** This thesis focused on designing and applying new multivariate statistical tools for, on the one hand, bias correction of climate simulations, and on the other hand, the investigation of compounding climate events with high impacts. These two specific topics were briefly introduced in Chapter 1.

My work has first focused on the intercomparison of existing multivariate bias correction methods to adjust dependence properties of climate simulations (Chapter 2). Four multivariate statistical methods with differences in terms of methodologies, statistical techniques and assumptions, have been applied to adjust simulated temperature and precipitation times series from a climate model with respect to reanalyses. A univariate BC method has also been applied to assess the benefits of considering multivariate bias correction. Implemented in different dimensional configurations, the MBC methods showed great ability to adjust the complete inter-variable, spatial, and inter-variable/spatial dependencies of simulated temperature and precipitation. However, none of the evaluated MBC methods were designed to adjust temporal properties, resulting in generating corrected outputs with weak temporal dependencies compared to reanalyses. In addition, instabilities affecting the quality of corrections were detected for some methods when applied in high-dimensional contexts. Also, evaluating the multivariate BC methods in a non-stationary context has shown that, when designed for this purpose, the methods are able to take simulated changes in univariate and multivariate properties into account reasonably well in the correction procedure. In general, this chapter concludes that **it**

is crucial for end users to carefully evaluate the suitability of the multivariate BC methods for their application purposes, before applying them and conducting analysis studies. The work carried out has allowed to draw general recommendations to help end users to choose which multivariate bias correction method depending on their needs.

Among the perspectives arising from Chapter 2 is the need for new MBC methods able to adjust multivariate dependence in high-dimensional contexts. This can be particularly valuable to provide adjustments of spatial dependencies between a high number of locations (e.g., grid cells or stations). In Chapter 3, I then proposed a new multivariate bias correction method based on Machine Learning techniques that are promising approaches to model complex dependence between a large number of statistical variables. More specifically, this new method based on a neural network algorithm named CycleGAN has been designed to adjust spatial dependencies of climate simulations. The new method was tested to adjust simulated temperature and precipitation time series with respect to reanalyses dataset. Developed in a proof-of-concept context, the aim of this work was more to explore the potential of machine learning techniques for multivariate bias correction than to provide a method outperforming others MBC alternatives. Still, the new method produced promising performances of adjustments of spatial dependencies for temperature and precipitation with respect to other MBC state-of-the-art techniques. The results obtained in this chapter suggest that **the use and adaptation of Machine Learning techniques is very promising for multivariate bias correction.** The work carried out in this chapter can serve as a starting point for the use of such statistical tools for this purpose.

Finally, Chapter 4 focused on investigating the multivariate distributions of compounding events potentially leading to high impacts, as well as their probability changes. I specifically focused on studying such changes in climate simulations, although assessing these changes for historical periods in reanalyses is also an interesting perspective. As the most significant impacts are likely to occur when changes emerge from natural climate variability, I proposed a new methodology to assess the time at which simulated compound events probabilities go out from the background variability (also named "Time of Emergence"). The multivariate modelling technique, based on copula functions, is used to estimate compound events probabilities. The occurrence probability of compound events being defined by both univariate properties and dependence structure of hazards forming the CEs, it is crucial to determine how much of probability changes are due to changing univariate properties, and how much are due to changing dependence structures. Hence, within this framework, I also proposed to quantify the contribution of univariate and dependence properties in the probability changes of multivariate climate events leading to compound events. Using this methodology, I analysed two different compounding climate events using simulations from 13 climate models (CMIP6): wind and precipitation extremes in Brittany and growing-period frosts over Central France. The results suggest that probabilities of these two different compounding climate events are likely to increase and potentially emerge before the end of the 21st century depending on the climate model considered. The results also highlight that not only univariate properties but also **dependence structures can have a strong importance on probability changes of compound events and Time of Emergence**. This importance can however differ from a compound hazard to another and from a climate model to another: while, for wind and precipitation extremes, a balanced importance of both univariate and dependence properties to probability changes is obtained, the contribution of dependence properties for growing-period frosts events is mostly negligible. Results from Chapter 4 illustrated once again, after Chapters 2 and 3, the added value of taking dependence properties into account to analyse climate events.

5.2 Perspectives

Throughout the manuscript, specific perspectives for each of the presented studies have been discussed in Chapters 2, 3 and 4. In this section, I will not repeat the same perspectives that have already been mentioned, but I will rather discuss new perspectives emerging from all the work carried out on bias correction of climate simulations (Chapters 2 and 3) and compound events (Chapter 4).

Assessing the ability of existing MBC methods to adjust simulated multivariate hazards

Although not evaluated in Chapter 4, biases in simulated multivariate hazards forming CEs are necessarily present, whether on univariate properties, on dependence structure, or both, which thus potentially affects the robustness of the results obtained in associated CE analyses. Adjusting correctly the univariate properties and dependence of climate hazards is therefore crucial to conduct appropriate investigations on present and future compound events. As the ability of MBC methods to improve univariate properties and various dependence structures of simulated climate variables has been shown in Chapter 2, their use is promising to increase the realism of specific compound events. In such a perspective, different methodological approaches can be considered to apply MBC methods. For example, existing MBCs can be applied to adjust the complete multivariate distributions of hazards (i.e., not only the extreme part) leading to CEs. This however explicitly assumes that the univariate and dependence properties of the subset of the multivariate distribution leading to CEs require the same correction as the rest of the distributions, which can be questionable. If this assumption is not considered valid, another possibility would be to apply MBC methods not to adjust the complete distribution, but rather the subset of the multivariate distribution leading to impacts. This possibility however assumes that multivariate subsets leading to CEs can be well identified, which is in practice difficult due to missing impact data.

Independently from the chosen methodological approach (i.e., MBC method and/or correcting whole the distribution or a part of it), the ability of the different existing MBC methods to increase the realism of specific compound events would need to be evaluated. In particular, this is questionable for compound events involving a high number of climate variables (e.g., spatially compounding events over a large domain), as instabilities for some MBC methods in high-dimensional context have been detected in Chapter 2. The ability of MBC methods to adjust compound events can also be questioned by the fact that the methods are not often able to account for temporal dependencies (with the exception of recent methods; Mehrotra and Sharma, 2019; Robin and Vrac, 2021; Vrac and Thao, 2020), while it can be key properties for several CEs. Also, some CEs can be composed of interacting extreme climate events, thus located in the tails of multivariate distributions (e.g., wind and precipitation extremes). Without even considering multivariate aspects, adjusting univariate extremes is already a challenging task, mainly due to the limited length of the time series (e.g., Kallache et al., 2011; Schmith et al., 2021). Correcting robustly multivariate extreme distributions is even more difficult as it would require larger amounts of climate data to fill the high-dimensional data space. More generally, **assessing the ability of existing MBC methods to improve the realism of different simulated multivariate hazards forming CEs is thus an important perspective towards robust analyses of simulated past, present and future compound events.**

Designing new MBCs for CEs

Depending on the capabilities of the existing MBC methods to increase the realism of CEs, **investigating the design of new MBC methods based on sophisticated statistical and machine learning tools can be envisaged.** In a second step, as practitioners can be interested not only in an appropriate representation of multivariate hazards but also in the rest of the multivariate distributions, research could then be envisaged to design new two-fold MBCs methods that consider separately the adjustments of the bulk of the multivariate distribution and the subset leading to CEs. Such new MBC methods would follow the philosophy of two-fold univariate bias correction methods (e.g., Vrac et al., 2007; Carreau and Bengio, 2009; Laflamme et al., 2016; Tani and Gobiet, 2019; Holthuijzen et al., 2022) for multivariate aspects. However, as mentioned above, this work would assume that multivariate subsets leading to CEs can be well identified, which is in practice difficult due to the lack of impact data. Still, developing two-fold MBCs represents an interesting perspective to improve the representation of many multivariate distributions for which compounding extremes are likely to generate huge impacts (e.g., due to simultaneous wind and precipitation extremes).

Influence of MBCs on compound events studies

To assess the potential benefits of using multivariate BC methods for CEs, a first natural step could be to derive corrections using the different BC methods used in Chapter 2 to adjust, over historical periods, simulated wind and precipitation extremes and growing-period frost events (GDD and minimal temperature data) from the CMIP6 models used in Chapter 4. To this end, the reference datasets (reanalyses or observations) used for bias correction will have to be carefully selected according to their ability to represent the targeted dependencies (e.g., Cortés-Hernández et al., 2016). By then computing multivariate hazards probabilities from adjusted outputs and reanalyses over historical periods, this would permit to give a first idea of the potential improvements provided by BC on the realism of multivariate hazards leading to CEs in a low-dimensional context. In particular, is there any added value in considering multivariate aspects in bias correction compared to only adjusting univariate properties? I expect it would depend on how much dependencies are important for the considered multivariate hazards, and how well raw climate models are already simulating the dependence, which would have to be investigated. The results of adjustments of CEs using BC methods would also permit to identify potential avenues for the development of new methods. Then, corrections of multivariate hazards leading to CEs could be performed for future periods, where no reanalyses are available but the evolutions of probability could be investigated. **An interesting perspective would be to assess the influence of adjustments on the time of emergence of these compound hazards probabilities**, e.g., by applying the methodology developed in Chapter 4 to bias-corrected outputs. In particular, do the different MBC methods modify ToE values similarly? And how contributions from univariate and dependence to probability changes are modified by MBC methods? This would potentially lead to provide more robust information for adaptation planning. These questions could then be generalised to other types of compounding hazards. Also, as already mentioned in the perspectives of Chapter 4, adapting the methodology developed for ToE to an attribution framework would be interesting in order to attribute changes of compound events probabilities to specific causes such as anthropogenic emissions. However, despite univariate extreme event attribution is now a mature field of research (e.g., Stott et al., 2016; Vautard et al., 2016; van der Wiel et al., 2017; Ribes et al., 2017; Otto et al., 2018; Philip et al., 2018; Yiou et al., 2020; van Oldenborgh et al., 2021, among many others), only a few

studies have focused on multivariate compound event attribution (e.g., Mazdiyasni et al., 2019; Kiriliouk and Naveau, 2020; Chiang et al., 2021; Verschuur et al., 2021; Zscheischler and Lehner, 2021). Conducting various compound events attribution studies in the context of future changes using 'raw' model simulations with and without anthropogenic forcings is of course an important perspective. Then, understanding the impact of statistical biases of the models on the attribution results would be essential to generate information adapted to the real world and not only to the simulated world. This could be partly investigated by first examining the influence of the different MBC methods on the results of compound events attribution studies.

Are existing MBCs able to take nonstationarity of CEs into account?

From the above perspective, we can easily understand that the notion of nonstationarity assumed (or not) in MBC methods would be crucial in ToE or compound events attribution results obtained from adjusted outputs, and more generally for the realism of simulated multivariate hazards forming CEs for future periods. In Chapter 2, we have seen that 3 out of the 4 tested MBC methods (MBC-n, dOTC and MRec) are designed to take simulated changes in multivariate properties into account in the correction procedure. **Investigating whether MBC methods assuming nonstationarity are able to capture simulated changes of CEs and to translate it into the multivariate corrections would permit to draw general recommendations for users to help them in the choice of BC methods for CEs.** This work would allow to get more understanding on the reliability of MBC methods and on the credibility of their corrections in a non-stationary context. Before applying MBC methods for CEs studies, the relevance of the simulated nonstationarity would however need to be investigated, raising key questions similar to those already formulated in Chapter 2, but this time specifically for multivariate hazards: Can we trust climate models in multivariate hazards evolutions (Vrac et al., 2021)? And should these changes be reproduced in the corrections (Vrac et al., 2022)? These are still open questions that would need to be answered on a case-by-case basis depending on the multivariate hazards before applying such corrections. The methodology developed in Chapter 4 could be helpful in such a perspective, as it would permit to produce a first analysis of univariate and dependence properties changes and their importance on probabilities of multivariate hazards. It then would permit to deduce the necessity of taking into account multi-dimensional changes in the bias correction procedure.

Studying the influence of large-scale circulations biases on the realism of CEs

Recent studies highlighted the importance of atmospheric circulation (e.g., Bevacqua et al., 2017; Faranda et al., 2020; Jézéquel et al., 2020; Blanchet et al., 2021) and large-scale climate modes (De Luca et al., 2020b; Singh et al., 2021) on compound and extreme events occurrences. When analysing simulated multivariate hazards in Chapter 4, these roles have not been investigated but it represents an interesting perspective. However, concerning atmospheric circulation, many climate models are not able to correctly represent the large-scale circulation patterns and their associated frequencies (Dawson et al., 2012; Dunn-Sigouin and Son, 2013), despite some progress between CMIP5 and CMIP6 models (e.g., Cannon, 2020). This is also true for simulated large-scale climate modes like El Niño-Southern Oscillation (e.g., Bellenger et al., 2014). To increase the understanding of simulated CEs, a first step would be **to investigate the contribution of biases in simulated large-scale circulation to biases in univariate and dependence properties of simulated multivariate hazards.** This work would

also participate to increase the knowledge of the influence of large-scale circulation patterns on statistical properties of multivariate hazards in climate models and reanalyses. In addition, not only the simulated circulation (e.g., occurrence probability and persistence of weather types) can be biased, but also the links between circulation patterns and meteorological variables (such as precipitation of wind), i.e, the univariate and multivariate properties of the meteorological variables conditional on the circulation pattern. The decomposition of these different sources of biases for multivariate hazards would provide a more complete picture of the performances of climate models in realistically simulating the climate system. A similar work has already been performed to investigate biases of temperature-precipitation relationships (Vrac et al., 2021) conditional on weather regimes, but was not particularly focused on high-impact events. Extending such investigations to high-impact or specific compound event would then allow to identify what contributes to the biases in multivariate hazard estimates, which could then be exploited to better adjust them using MBCs.

Towards adjustments of large-scale circulations

From all of the above-mentioned potential biases in large-scale circulation patterns, one question arises: can existing MBC methods be applied to adjust large-scale circulation structures? and to correct links between large-scale circulation structures and meteorological variables? And are these corrections able to improve the realism of circulation-driven CEs? Adjusting large-scale circulation structures over large domain would potentially be a problem for existing MBC methods presenting limited capability in high-dimensional contexts. **Developing new MBC methods to adjust large-scale circulation patterns based on Machine Learning techniques is a promising perspective to explore,** as Machine Learning tools like CNNs (Chapter 3) proved to be very efficient to model spatial fields in high-dimensional context. Also, a recent study led by Keisler (2022) reached impressive results to model spatio-temporal climate data using Graph Neural Networks, a generalised version of CNNs. As the circulation patterns and properties might also be modified by climate change, the question of nonstationarity would have to be explicitly taken into account when designing such new MBC tools.

Better accounting for uncertainties in compound events studies and bias correction

Uncertainties in climate studies are a crucial issue that needs to be well communicated to draw appropriate conclusions for compound event studies, and more generally when studying any impacts. Although sometimes partially taken into account in the manuscript, uncertainties inherent to bias correction and compound events studies deserve to be further investigated. In Chapter 4, by considering several climate models in the compound events study, I have partially taken into account the source of uncertainty associated to the different physical representations within climate models (inter-model differences). However, by only considering simulated projections from a single shared socioeconomic pathways (SSP-585) scenario, uncertainty due to the different scenarios has not been investigated. Using the SSP-585 scenario, we have seen that inter-model variability of simulated statistical properties of multivariate hazards can be important. It should be noted that in addition to this inter-model uncertainty, intra-model uncertainty can be added, which results from internal climate variability of the climate models. As already discussed in Chapter 4, this uncertainty can however be estimated by using large-ensemble simulations, i.e., multiple runs per climate model. Concerning the studies carried out

on bias correction (Chapters 2 and 3), none of the uncertainties mentioned above (emission scenario, model uncertainty and climate internal variability uncertainty) have been differentiated: in each study, a single climate model simulation to correct has been considered, thus implying that statistical biases of the model are potentially mixed up with its internal variability. The question is: should these different uncertainties be treated separately in bias correction procedure? Recently, Vaittinada Ayar et al. (2021) proposed a new univariate BC approach to take advantage of multiple runs of one model so that internal variability of simulated univariate properties is preserved in corrected outputs. Such an approach can be very insightful to better examine the different sources of uncertainties and their influences on subsequent analyses. Extending such approaches into the multivariate context to adjust dependence model biases while preserving internal variability of dependence structures is an interesting avenue to be explored. However, from a practical point of view, adjusting multiple runs can be potentially very costly in terms of resources and computing time, especially if various scenarios and different climate models are also considered. **Despite being difficult in practice, better accounting and communicating for all the different sources of uncertainties in climate studies is key to enable more insightful and informed decision-making.** At least, being the most transparent by providing information on the different sources of uncertainties when possible is recommended in future works, as it will greatly benefit the scientific community.

Towards more informed use of bias correction

Since the beginning of my thesis, it has been shown that MBC methods are able to adjust a broad range of statistical aspects including univariate, inter-variable, spatial and temporal properties. Thus, although still in its infancy, MBC methods offer many opportunities to end users for their applications. However, the MBC method adjusting perfectly univariate, inter-variable, spatial and temporal properties does not exist as adjustments of one statistical aspect necessarily modify, even slightly, the other statistical properties. Also, when many statistical properties are adjusted, this raises the question: what is finally kept from the climate model? This point is critically important as, by being based on statistical techniques, MBC methods lack the physical justification of climate models, which can sometimes limit the confidence in climate change projections involving BC. Regarding this issue, developing new physics-guided MBC methods that are able to use some physical processes to drive the correction procedure would be an interesting avenue to explore in order to provide physically consistent bias-corrected simulations (Pan et al., 2021). In any case, before applying any correction, end users have to ask themselves: Is the climate model good enough for my application? Is bias correction really needed? If so, what are the important statistical properties I want to adjust? Can a bias correction method provide relevant adjustments without deteriorating other important statistical and physical aspects? I really would like to stress that these questions should be answered by end-users before any use of BCs in order to avoid indiscriminate applications. **It would then permit to move towards more informed use of bias correction, which would contribute to an improved understanding of the climate and its evolution.**

All these perspectives represent a much larger amount of work than all of the work presented in this manuscript. This gives an insight into the fertility of the work carried out in this thesis, which sets up avenues for future research.

Bibliography

Aas, K., Czado, C., Frigessi, A., and Bakken, H.: Pair-copula constructions of multiple dependence, Insur.: Math. Econ., 44, 182–198, 2009.

Abatzoglou, J. T., Dobrowski, S. Z., and Parks, S. A.: Multivariate climate departures have outpaced univariate changes across global lands, Sci. Rep., 10, 2020.

Acar, E. F., Genest, C., and Nešlehová, J.: Beyond simplified pair-copula constructions, J. Multivar. Anal., 110, 74–90, https://doi.org/10.1016/j.jmva.2012.02.001, 2012.

Agassiz, L.: Des glaciers, des moraines, et des blocs erratiques, Bibliothèque universelle de Gèneve, 12, 369–394, 1837.

Aghatise, O., Khan, F., and Ahmed, S.: Reliability assessment of marine structures considering multidimensional dependency of the variables, Ocean Eng., 230, 109 021, https://doi.org/10.1016/j.oceaneng.2021.109021, 2021.

Alidoost, F., Su, Z., and Stein, A.: Evaluating the effects of climate extremes on crop yield, production and price using multivariate distributions: A new copula application, Weather Clim. Extrem., 26, 100 227, https://doi.org/10.1016/j.wace.2019.100227, 2019.

Allard, D., Ailliot, P., Monbet, V., and Naveau, P.: Stochastic weather generators: an overview of weather type models, J. Soc. Fr. Statistique, 156, 101–113, 2015.

Anderson, T.: An Introduction to Multivariate Statistical Analysis, 230 pp, John Wiley Sons, New York, 1958.

Aristotle: Meteorologica, English translation by E.W. Webster (vol. III of The works of Aristotle), Clarendon Press, 1931.

Arjovsky, M., Chintala, S., and Bottou, L.: Wasserstein GAN, 2017.

Arrhenius, S.: On the influence of carbonic acid in the air upon the temperature of the ground, Lond. Edinb. Dublin philos. mag. j. sci., 41, 237–276, https://doi.org/10.1080/14786449608620846, 1896.

Baño Medina, J.: Understanding Deep Learning Decisions in Statistical Downscaling Models, in: Proceedings of the 10th International Conference on Climate Informatics, CI2020, pp. 79–85, ACM, https://doi.org/10.1145/3429309.3429321, 2020.

Baño-Medina, J., Manzanas, R., and Gutiérrez, J. M.: Configuration and intercomparison of deep learning neural models for statistical downscaling, Geosci. Model Dev., 13, 2109–2124, https://doi.org/10.5194/gmd-13-2109-2020, 2020.

Babaousmail, H., Hou, R., Gnitou, G. T., and Ayugi, B.: Novel statistical downscaling emulator for precipitation projections using deep Convolutional Autoencoder over Northern Africa, J. Atmos. Sol.-Terr. Phys., 218, 105 614, https://doi.org/10.1016/j.jastp.2021.105614, 2021.

Bárdossy, A. and Pegram, G.: Multiscale spatial recorrelation of RCM precipitation to produce unbiased climate change scenarios over large areas and small, Water Resour. Res., 48, 9502–, https://doi.org/10.1029/2011WR011524, 2012.

Bardossy, A. and Plate, E. J.: Space-time model for daily rainfall using atmospheric circulation patterns, Water Resour. Res., 28, 1247–1259, https://doi.org/10.1029/91WR02589, 1992.

Bartok, B., Tobin, I., Vautard, R., Vrac, M., Jin, X., Levavasseur, G., Denvil, S., Dubus, L., Parey, S., Michelangeli, P.-A., Troccoli, A., and Saint-Drenan, Y.-M.: A climate projection dataset tailored for the European energy sector, Clim. Serv., 16, 100 138, https://doi.org/10.1016/j.cliser.2019.100138, 2019.

Barton, Y., Giannakaki, P., von Waldow, H., Chevalier, C., Pfahl, S., and Martius, O.: Clustering of Regional-Scale Extreme Precipitation Events in Southern Switzerland, Mon. Weather Rev., 144, 347–369, https://doi.org/10.1175/MWR-D-15-0205.1, 2016.

Bashkirova, D., Usman, B., and Saenko, K.: Unsupervised Video-to-Video Translation, https://doi.org/10.48550/ARXIV.1806.03698, 2018.

Bates, B., Kundzewicz, Z., Wu, S., Burkett, V., Doell, P., Gwary, D., Hanson, C., Heij, B., Jiménez, B., Kaser, G., Kitoh, A., Kovats, S., Kumar, P., Magadza, C., Martino, D., Mata, L., Medany, M., Miller, K., and Arnell, N.: Climate Change and Water. Technical Paper of the Intergovernmental Panel on Climate Change, Tech. rep., The Intergovernmental Panel on Climate Change, 2008.

Batté, L. and Doblas-Reyes, F. J.: Stochastic atmospheric perturbations in the EC-Earth3 global coupled model: Impact of SPPT on seasonal forecast quality, Clim. Dyn., 45, 3419–3439, 2015.

Bellenger, H., Guilyardi, E., Leloup, J., Lengaigne, M., and Vialard, J.: ENSO representation in climate models: from CMIP3 to CMIP5, Clim. Dyn., 42, 1999–2018, https://doi.org/10.1007/s00382-013-1783-z, 2014.

Bengio, Y., Lamblin, P., Popovici, D., Larochelle, H., and Montreal, U.: Greedy layer-wise training of deep networks, Adv. Neural Inf. Process. Syst., 19, 2007.

Berg, P., Feldmann, H., and Panitz, H.-J.: Bias correction of high resolution regional climate model data, J. Hydrol., 448–449, 80–92, https://doi.org/10.1016/j.jhydrol.2012.04.026, 2012.

Berner, J., Achatz, U., Batté, L., Bengtsson, L., de la Cámara, A., Christensen, H. M., Colangeli, M., Coleman, D. R. B., Crommelin, D., Dolaptchiev, S. I., Franzke, C. L. E., Friederichs, P., Imkeller, P., Järvinen, H., Juricke, S., Kitsios, V., Lott, F., Lucarini, V., Mahajan, S., Palmer, T. N., Penland, C., Sakradzija, M., von Storch, J.-S., Weisheimer, A., Weniger, M., Williams, P. D., and Yano, J.-I.: Stochastic Parameterization: Toward a New View of Weather and Climate Models, Bull. Am. Meteorol. Soc., 98, 565–588, https://doi.org/10.1175/BAMS-D-15-00268.1, 2017.

Beucler, T., Pritchard, M., Yuval, J., Gupta, A., Peng, L., Rasp, S., Ahmed, F., O'Gorman, P. A., Neelin, J. D., Lutsko, N. J., and Gentine, P.: Climate-Invariant Machine Learning, https://doi.org/10.48550/ARXIV.2112.08440, 2021.

Bevacqua, E., Maraun, D., Hobæk Haff, I., Widmann, M., and Vrac, M.: Multivariate statistical modelling of compound events via pair-copula constructions: analysis of floods in Ravenna (Italy), Hydrol. Earth Syst. Sci., 21, 2701–2723, https://doi.org/10.5194/hess-21-2701-2017, 2017.

Bevacqua, E., Maraun, D., Vousdoukas, M. I., Voukouvalas, E., Vrac, M., Mentaschi, L., and Widmann, M.: Higher probability of compound flooding from precipitation and storm surge in Europe under anthropogenic climate change, Sci. Adv., 5, https://doi.org/10.1126/sciadv.aaw5531, 2019.

Bevacqua, E., Zappa, G., Lehner, F., and Zscheischler, J.: Precipitation trends determine future occurrences of compound hot–dry events, Nat. Clim. Chang., https://doi.org/10.1038/s41558-022-01309-5, 2022.

Bjerknes, V.: Das Problem der Wettervorhersage, betrachtet vom Standpunkte der Mechanik und der Physik, Meteorol. Z., 21, 1–7, 1904.

Blanchet, J. and Davison, A. C.: Spatial modeling of extreme snow depth, The Annals of Applied Statistics, 5, https://doi.org/10.1214/11-aoas464, 2011.

Blanchet, J., Blanc, A., and Creutin, J.-D.: Explaining recent trends in extreme precipitation in the Southwestern Alps by changes in atmospheric influences, Weather. Clim. Extremes, 33, 100 356, https://doi.org/10.1016/j.wace.2021.100356, 2021.

Boulaguiem, Y., Zscheischler, J., Vignotto, E., van der Wiel, K., and Engelke, S.: Modeling and simulating spatial extremes by combining extreme value theory with generative adversarial networks, Environmental Data Science, 1, e5, https://doi.org/10.1017/eds.2022.4, 2022.

Breiman, L.: Random Forests, Machine Learning, 45, 5–32, https://doi.org/10.1023/A:1010933404324, 2001.

Breitner, S., Wolf, K., Devlin, R. B., Diaz-Sanchez, D., Peters, A., and Schneider, A.: Short-term effects of air temperature on mortality and effect modification by air pollution in three cities of Bavaria, Germany: A time-series analysis, Sci. Total Environ., 485-486, 49–61, https://doi.org/10.1016/j.scitotenv.2014.03.048, 2014.

Brunner, M. I., Viviroli, D., Sikorska, A. E., Vannier, O., Favre, A.-C., and Seibert, J.: Flood type specific construction of synthetic design hydrographs, Water Resour. Res., 53, 1390–1406, https://doi.org/10.1002/2016WR019535, 2017.

Brunner, M. I., Gilleland, E., Wood, A., Swain, D. L., and Clark, M.: Spatial Dependence of Floods Shaped by Spatiotemporal Variations in Meteorological and Land-Surface Processes, Geophys. Res. Lett., 47, e2020GL088 000, https://doi.org/10.1029/2020GL088000, 2020.

Caminade, C., Kovats, S., Rocklov, J., Tompkins, A. M., Morse, A. P., Colón-González, F. J., Stenlund, H., Martens, P., and Lloyd, S. J.: Impact of climate change on global malaria distribution, Proc. Natl. Acad. Sci. U.S.A., 111, 3286–3291, https://doi.org/10.1073/pnas.1302089111, 2014.

Cannon, A., Sobie, S., and Murdock, T.: Bias Correction of GCM Precipitation by Quantile Mapping: How Well Do Methods Preserve Changes in Quantiles and Extremes?, J. Climate, 28, 6938–6959, https://doi.org/10.1175/JCLI-D-14-00754.1, 2015.

Cannon, A. J.: Multivariate quantile mapping bias correction: an N-dimensional probability density function transform for climate model simulations of multiple variables, Clim. Dynam., 50, 31–49, https://doi.org/10.1007/s00382-017-3580-6, 2018.

Cannon, A. J.: Reductions in daily continental-scale atmospheric circulation biases between generations of global climate models: CMIP5 to CMIP6, Environ. Res. Lett., 15, 064 006, https://doi.org/10.1088/1748-9326/ab7e4f, 2020.

Carranza-García, M., García-Gutiérrez, J., and Riquelme, J. C.: A Framework for Evaluating Land Use and Land Cover Classification Using Convolutional Neural Networks, Remote Sens., 11, https://doi.org/10.3390/rs11030274, 2019.

Carreau, J. and Bengio, Y.: A hybrid Pareto model for asymmetric fat-tailed data: the univariate case, Extremes, 12, 53–76, https://doi.org/10.1007/s10687-008-0068-0, 2009.

Chamberlin, T. C.: An Attempt to Frame a Working Hypothesis of the Cause of Glacial Periods on an Atmospheric Basis, J. Geol., 7, 545–584, 1899.

Chapman, W. E., Subramanian, A. C., Delle Monache, L., Xie, S. P., and Ralph, F. M.: Improving Atmospheric River Forecasts With Machine Learning, Geophys. Res. Lett., 46, 10 627–10 635, https://doi.org/10.1029/2019GL083662, 2019.

Charney, J. G., FjÖrtoft, R., and Neumann, J. V.: Numerical Integration of the Barotropic Vorticity Equation, Tellus, 2, 237–254, https://doi.org/10.3402/tellusa.v2i4.8607, 1950.

Chemison, A., Ramstein, G., Tompkins, A. M., Defrance, D., Camus, G., Charra, M., and Caminade, C.: Impact of an accelerated melting of Greenland on malaria distribution over Africa, Nat. Commun., 12, 3971, https://doi.org/10.1038/s41467-021-24134-4, 2021.

Chen, J., Brissette, F. P., Chaumont, D., and Braun, M.: Finding appropriate bias correction methods in downscaling precipitation for hydrologic impact studies over North America, Water Resour. Res., 49, 4187–4205, https://doi.org/10.1002/wrcr.20331, 2013.

Chen, Y., Pan, Y., Yao, T., Tian, X., and Mei, T.: Mocycle-GAN: Unpaired Video-to-Video Translation, https://doi.org/10.48550/ARXIV.1908.09514, 2019.

Cheng, Y. and Kan, H.: Effect of the interaction between outdoor air pollution and extreme temperature on daily mortality in Shanghai, China, J. Epidemiol., 22, 28–36, https://doi.org/10.2188/jea.je20110049, 2012.

Chiang, F., Greve, P., Mazdiyasni, O., Wada, Y., and AghaKouchak, A.: A Multivariate Conditional Probability Ratio Framework for the Detection and Attribution of Compound Climate Extremes, Geophys. Res. Lett., 48, e2021GL094 361, https://doi.org/10.1029/2021GL094361, 2021.

Christensen, J., Hewitson, B., Busuioc, A., Chen, A., Gao, X., Held, I., Jones, R., Kolli, R., Kwon, W.-T., Laprise, R., Rueda, V., Mearns, L., Menéndez, C., Räisänen, J., Rinke, A., Sarr, A., and Whetton, P.: Regional climate projections. Climate change 2007: The physical science basis, Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, pp. 847–940, 2007.

Christensen, J. H., Boberg, F., Christensen, O. B., and Lucas-Picher, P.: On the need for bias correction of regional climate change projections of temperature and precipitation, Geophys. Res. Lett., 35, L20 709, https://doi.org/10.1029/2008GL035694, 2008.

Ciais, P., Reichstein, M., Viovy, N., Granier, A., Ogée, J., Allard, V., Aubinet, M., Buchmann, N., Bernhofer, C., Carrara, A., Chevallier, F., De Noblet, N., Friend, A. D., Friedlingstein, P., Grünwald, T., Heinesch, B., Keronen, P., Knohl, A., Krinner, G., Loustau, D., Manca, G., Matteucci, G., Miglietta, F., Ourcival, J. M., Papale, D., Pilegaard, K., Rambal, S., Seufert, G., Soussana, J. F., Sanz, M. J., Schulze, E. D., Vesala, T., and Valentini, R.: Europe-wide reduction in primary productivity caused by the heat and drought in 2003, Nature, 437, 529–533, https://doi.org/10.1038/nature03972, 2005.

Coffel, E. D., Horton, R. M., and de Sherbinin, A.: Temperature and humidity based projections of a rapid rise in global heat stress exposure during the 21st century, Environ. Res. Lett., 13, 014 001, https://doi.org/10.1088/1748-9326/aaa00e, 2017.

Cohen, J., Coumou, D., Hwang, J., Mackey, L., Orenstein, P., Totz, S., and Tziperman, E.: S2S reboot: An argument for greater inclusion of machine learning in subseasonal to seasonal forecasts, Wiley Interdiscip Rev Clim Change, 10, e00 567, https://doi.org/10.1002/wcc.567, 2019.

Coles, S.: An introduction to statistical modeling of extreme values, Springer Series in Statistics, Springer-Verlag, London, 2001.

Collins, M., Knutti, R., Arblaster, J., Dufresne, J.-L., Fichefet, T., Friedlingstein, P., Gao, X., Gutowski, W., Johns, T., Krinner, G., Shongwe, M., Tebaldi, C., Weaver, A., and Wehner, M.: Long-term Climate Change: Projections, Commitments and Irreversibility, book section 12, p. 1029–1136, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, https://doi.org/10.1017/CBO9781107415324.024, 2013.

Cortes, C. and Vapnik, V.: Support-vector networks, Machine Learning, 20, 273–297, https://doi.org/10.1007/BF00994018, 1995.

Cortés-Hernández, V. E., Zheng, F., Evans, J., Lambert, M., Sharma, A., and Westra, S.: Evaluating regional climate models for simulating sub-daily rainfall extremes, Clim. Dynam., 47, 1613–1628, https://doi.org/10.1007/s00382-015-2923-4, 2016.

Corti, S., Molteni, F., and Palmer, T. N.: Signature of recent climate change in frequencies of natural atmospheric circulation regimes, Nature, 398, 799–802, https://doi.org/10.1038/19745, 1999.

Couasnon, A., Sebastian, A., and Morales-Nápoles, O.: A Copula-Based Bayesian Network for Modeling Compound Flood Hazard from Riverine and Coastal Interactions at the Catchment Scale: An Application to the Houston Ship Channel, Texas, Water, 10, https://doi.org/10.3390/w10091190, 2018.

Cramer, W., Yohe, G. W., Auffhammer, M., Huggel, C., Molau, U., Dias, M. A. F. S., Solow, A., Stone, D. A., and Tibig, L.: Detection and attribution of observed impacts, pp. 979—-1037, Cambridge University Press, 2014.

Dai, A. and Wigley, T. M. L.: Global patterns of ENSO-induced precipitation, Geophys. Res. Lett., 27, 1283–1286, https://doi.org/10.1029/1999GL011140, 2000.

Daniely, A., Frostig, R., and Singer, Y.: Toward Deeper Understanding of Neural Networks: The Power of Initialization and a Dual View on Expressivity, https://doi.org/10.48550/ARXIV.1602.05897, 2016.

Dawson, A., Palmer, T. N., and Corti, S.: Simulating regime structures in weather and climate prediction models, Geophys. Res. Lett., 39, https://doi.org/10.1029/2012GL053284, 2012.

de Bezenac, E., Pajot, A., and Gallinari, P.: Deep Learning for Physical Processes: Incorporating Prior Scientific Knowledge, https://doi.org/10.48550/ARXIV.1711.07970, 2017.

de Bézenac, E., Ayed, I., and Gallinari, P.: CycleGAN Through the Lens of (Dynamical) Optimal Transport, in: Machine Learning and Knowledge Discovery in Databases. Research Track, edited by Oliver, N., Pérez-Cruz, F., Kramer, S., Read, J., and Lozano, J. A., pp. 132–147, Springer International Publishing, Cham, 2021.

De Luca, P., Messori, G., Pons, F. M. E., and Faranda, D.: Dynamical systems theory sheds new light on compound climate extremes in Europe and Eastern North America, Q. J. Roy. Meteor. Soc., 146, 1636–1650, https://doi.org/10.1002/qj.3757, 2020a.

De Luca, P., Messori, G., Wilby, R. L., Mazzoleni, M., and Di Baldassarre, G.: Concurrent wet and dry hydrological extremes at the global scale, Earth Syst. Dynam., 11, 251–266, https://doi.org/10.5194/esd-11-251-2020, 2020b.

Defrance, D., Ramstein, G., Charbit, S., Vrac, M., Famien, A. M., Sultan, B., Swingedouw, D., Dumas, C., Gemenne, F., Alvarez-Solas, J., and Vanderlinden, J.-P.: Consequences of rapid ice sheet melting on the Sahelian population vulnerability, Proc. Natl. Acad. Sci. U.S.A., 114, 6533–6538, https://doi.org/10.1073/pnas.1619358114, 2017.

Dekens, L., Parey, S., Grandjacques, M., and Dacunha-Castelle, D.: Multivariate distribution correction of climate model outputs: A generalization of quantile mapping approaches: Multivariate distribution correction of climate model outputs, Environmetrics, 28, e2454, https://doi.org/10.1002/env.2454, 2017.

Denton, E., Chintala, S., Szlam, A., and Fergus, R.: Deep Generative Image Models using a Laplacian Pyramid of Adversarial Networks, 2015.

Didier, D., Baudry, J., Bernatchez, P., Dumont, D., Sadegh, M., Bismuth, E., Bandet, M., Dugas, S., and Sévigny, C.: Multihazard simulation for coastal flood mapping: Bathtub versus numerical modelling in an open estuary, Eastern Canada, J. Flood Risk Manag., 12, e12 505, https://doi.org/10.1111/jfr3.12505, 2019.

Diffenbaugh, N. and Scherer, M.: Observational and model evidence of global emergence of permanent, unprecedented heat in the 20th and 21st centuries, Clim. Change, 107, 615–624, https://doi.org/10.1007/s10584-011-0112-y, 2011.

Doury, A., Somot, S., Gadat, S., Ribes, A., and Corre, L.: Regional Climate Model Emulator based on Deep Learning: Concept and First Evaluation of a Novel Hybrid Downscaling Approach, https://doi.org/10.21203/rs.3.rs-725819/v1, preprint on webpage at https://www.researchsquare.com/article/rs-725819/v1, 2022.

Dunn-Sigouin, E. and Son, S.-W.: Northern Hemisphere blocking frequency and duration in the CMIP5 models, J. Geophys. Res. Atmos., 118, 1179–1188, https://doi.org/10.1002/jgrd.50143, 2013.

Déqué, M.: Frequency of precipitation and temperature extremes over France in an anthropogenic scenario: Model results and statistical correction according to observed values, Global Planet. Change, 57, 16–26, https://doi.org/10.1016/j.gloplacha.2006.11.030, 2007.

Ehret, U., Zehe, E., Wulfmeyer, V., Warrach-Sagi, K., and Liebert, J.: HESS Opinions "Should we apply bias correction to global and regional climate model data?", Hydrol. Earth Syst. Sci., 16, 3391–3404, https://doi.org/10.5194/hess-16-3391-2012, 2012.

Elith, J., H. Graham, C., P. Anderson, R., Dudík, M., Ferrier, S., Guisan, A., J. Hijmans, R., Huettmann, F., R. Leathwick, J., Lehmann, A., Li, J., G. Lohmann, L., A. Loiselle, B., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., McC. M. Overton, J., Townsend Peterson, A., J. Phillips, S., Richardson, K., Scachetti-Pereira, R., E. Schapire, R., Soberón, J., Williams, S., S. Wisz, M., and E. Zimmermann, N.: Novel methods improve prediction of species' distributions from occurrence data, Ecography, 29, 129–151, https://doi.org/10.1111/j.2006.0906-7590.04596.x, 2006.

Famien, A. M., Janicot, S., Ochou, A. D., Vrac, M., Defrance, D., Sultan, B., and Noel, T.: A bias-corrected CMIP5 dataset for Africa using the CDF-t method : a contribution to agricultural impact studies, Earth Syst. Dynam., 9, 313–338, https://doi.org/10.5194/esd-9-313-2018, 2018.

Faranda, D., Vrac, M., Yiou, P., Jézéquel, A., and Thao, S.: Changes in Future Synoptic Circulation Patterns: Consequences for Extreme Event Attribution, Geophys. Res. Lett., 47, e2020GL088 002, https://doi.org/10.1029/2020GL088002, 2020.

Favre, A.-C., El Adlouni, S., Perreault, L., Thiémonge, N., and Bobée, B.: Multivariate hydrological frequency analysis using copulas, Water Resour. Res., 40, https://doi.org/10.1029/2003WR002456, 2004.

Fischer, E. M., Sedláček, J., Hawkins, E., and Knutti, R.: Models agree on forced response pattern of precipitation and temperature extremes, Geophys. Res. Lett., 41, 8554–8562, https://doi.org/10.1002/2014GL062018, 2014.

Fix, E. and Hodges, J. L.: Discriminatory Analysis. Nonparametric Discrimination: Consistency Properties, Int. Stat. Rev., 57, 238–247, 1989.

Foote, E. N.: Circumstances Affecting the Heat of the Sun's Rays, Am. J. Sci. Arts, 22, 383–384, 1856.

Fourier, J.: Remarques générales sur les températures du globe terrestre et des espaces planétaires, Annal. Chim. Phys., 27, 136–167, 1824.

Frame, D., Joshi, M., Hawkins, E., Harrington, L., and Róiste, M.: Population-based emergence of unfamiliar climates, Nat. Clim. Chang., 7, 407–411, https://doi.org/10.1038/nclimate3297, 2017.

Gabellani, S., Boni, G., Ferraris, L., von Hardenberg, J., and Provenzale, A.: Propagation of uncertainty from rainfall to runoff: A case study with a stochastic rainfall generator, Adv. Water Resour., 30, 2061–2071, https://doi.org/10.1016/j.advwatres.2006.11.015, 2007.

Gaetani, M., Janicot, S., Vrac, M., Famien, A. M., and Sultan, B.: Robust assessment of the time of emergence of precipitation change in West Africa, Sci. Rep., 10, https://doi.org/10.1038/s41598-020-63782-2, 2020.

Galton, F.: Natural Inheritance, MacMillan, London, 1889.

Garg, S., Rasp, S., and Thuerey, N.: WeatherBench Probability: A benchmark dataset for probabilistic medium-range weather forecasting along with deep learning baseline models, https://doi.org/10.48550/ARXIV.2205.00865, 2022.

Gauss, K. F.: Theory of the Combination of Observations, Gottingen, 1823.

Gauthier, J.: Conditional generative adversarial nets for convolutional face generation, Class Project for Stanford CS231N: Convolutional Neural Networks for Visual Recognition, Winter semester, 2014(5):2, 2014.

Ghizzoni, T., Roth, G., and Rudari, R.: Multivariate skew-t approach to the design of accumulation risk scenarios for the flooding hazard, Adv. Water Resour., 33, 1243–1255, https://doi.org/10.1016/j.advwatres.2010.08.003, special Issue on Novel Insights in Hydrological Modelling, 2010.

Ghizzoni, T., Roth, G., and Rudari, R.: Multisite flooding hazard assessment in the Upper Mississippi River, J. Hydrol., 412-413, 101–113, https://doi.org/10.1016/j.jhydrol.2011.06.004, hydrology Conference 2010, 2012.

Giorgi, F. and Bi, X.: Time of emergence (TOE) of GHG-forced precipitation change hot-spots, Geophys. Res. Lett., 36, https://doi.org/10.1029/2009GL037593, 2009.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y.: Generative Adversarial Nets, Advances in Neural Information Processing Systems, 3, https://doi.org/10.1145/3422622, 2014.

Grimm, R., Behrens, T., Maerker, M., and Elsenbeer, H.: Soil organic carbon concentrations and stocks on Barro Colorado Island — Digital soil mapping using Random Forests analysis, Geoderma, 146, 102–113, https://doi.org/10.1016/j.geoderma.2008.05.008, 2008.

Gudmundsson, L., Bremnes, J. B., Haugen, J. E., and Engen-Skaugen, T.: Technical Note: Downscaling RCM precipitation to the station scale using statistical transformations – a comparison of methods, Hydrol. Earth Syst. Sci., 16, 3383–3390, https://doi.org/10.5194/hess-16-3383-2012, 2012.

Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., and Courville, A. C.: Improved Training of Wasserstein GANs, in: Advances in Neural Information Processing Systems, edited by Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R., vol. 30, Curran Associates, Inc., 2017.

Guo, Q., Chen, J., Zhang, X., Shen, M., Chen, H., and Guo, S.: A new two-stage multivariate quantile mapping method for bias correcting climate model outputs, Clim. Dynam., 53, 3603–3623, https://doi.org/10.1007/s00382-019-04729-w, 2019.

Guo, Q., Chen, J., Zhang, X. J., Xu, C.-Y., and Chen, H.: Impacts of Using State-of-the-Art Multivariate Bias Correction Methods on Hydrological Modeling Over North America, Water Resour. Res., 56, e2019WR026 659, https://doi.org/10.1029/2019WR026659, 2020.

Haddad, Z. and Rosenfeld, D.: Optimality of empirical Z-R relations, Q. J. Roy. Meteor. Soc., 123, 1283–1293, https://doi.org/10.1002/qj.49712354107, 1997.

Hadley, G.: Concerning the cause of the general trade-winds, Phil. Trans., 39, 436–44, 1735.

Han, Y., Zhang, G. J., Huang, X., and Wang, Y.: A Moist Physics Parameterization Based on Deep Learning, J. Adv. Model. Earth Syst., 12, e2020MS002 076, https://doi.org/10.1029/2020MS002076, 2020.

Hao, Z., Hao, F., Xia, Y., Singh, V. P., and Zhang, X.: A monitoring and prediction system for compound dry and hot events, Environ. Res. Lett., 14, 114 034, https://doi.org/10.1088/1748-9326/ab4df5, 2019.

Hawkins, E. and Sutton, R.: Time of emergence of climate signals, Geophys. Res. Lett., 39, https://doi.org/10.1029/2011GL050087, 2012.

Hawkins, E., Frame, D., Harrington, L., Joshi, M., King, A., Rojas, M., and Sutton, R.: Observed Emergence of the Climate Change Signal: From the Familiar to the Unknown, Geophys. Res. Lett., 47, e2019GL086 259, https://doi.org/10.1029/2019GL086259, 2020.

He, K., Zhang, X., Ren, S., and Sun, J.: Deep Residual Learning for Image Recognition, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778, https://doi.org/10.1109/CVPR.2016.90, 2016.

Hengl, T., Mendes de Jesus, J., Heuvelink, G. B. M., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M. N., Geng, X., Bauer-Marschallinger, B., Guevara, M. A., Vargas, R., MacMillan, R. A., Batjes, N. H., Leenaars, J. G. B., Ribeiro, E., Wheeler, I., Mantel, S., and Kempen, B.: SoilGrids250m: Global gridded soil information based on machine learning, PLOS ONE, 12, 1–40, https://doi.org/10.1371/journal.pone.0169748, 2017.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N.: The ERA5 global reanalysis, Q. J. R. Meteorol. Soc., 146, 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

Hewage, P., Trovati, M., Pereira, E., and Behera, A.: Deep learning-based effective fine-grained weather forecasting model, Pattern Anal. Appl., 24, 343–366, https://doi.org/10.1007/s10044-020-00898-1, 2021.

Hilbert, D. W. and Ostendorf, B.: The utility of artificial neural networks for modelling the distribution of vegetation in past, present and future climates, Ecol. Modell., 146, 311–327, https://doi.org/10.1016/S0304-3800(01)00323-4, 2001.

Hillier, J., Matthews, T., Wilby, R., and Murphy, C.: Multi-hazard dependencies can increase or decrease risk, Nat. Clim. Chang., 10, 1–4, https://doi.org/10.1038/s41558-020-0832-y, 2020.

Hinton, G. E., Osindero, S., and Teh, Y.-W.: A fast learning algorithm for deep belief nets, Neural Comput., 18, 1527–1554, 2006.

Hobæk Haff, I.: Comparison of estimators for pair-copula constructions, J. Multivar. Anal., 110, 91–105, https://doi.org/10.1016/j.jmva.2011.08.013, special Issue on Copula Modeling and Dependence, 2012.

Holthuijzen, M., Beckage, B., Clemins, P. J., Higdon, D., and Winter, J. M.: Robust bias-correction of precipitation extremes using a novel hybrid empirical quantile-mapping method, Theor. Appl. Climatol., https://doi.org/10.1007/s00704-022-04035-2, 2022.

Huth, R.: Sensitivity of local daily temperature change estimates to the selection of downscaling models and predictors, J. Climate, 17, 640–652, 2004.

IPCC: Annex II : Glossary. 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 [K. Mach et al.], IPCC, Geneva, Switzerland, 151 pp, pp. 117–130, 2014.

IPCC, 2021: *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* [Masson-Delmotte, V., P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (eds.)]., Cambridge University Press. In Press.

Isola, P., Zhu, J.-Y., Zhou, T., and Efros, A. A.: Image-to-Image Translation with Conditional Adversarial Networks, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5967–5976, https://doi.org/10.1109/CVPR.2017.632, 2017.

Jahan, N. and Gan, T. Y.: Modelling the vegetation–climate relationship in a boreal mixedwood forest of Alberta using normalized difference and enhanced vegetation indices, Int. J. Remote Sens., 32, 313–335, https://doi.org/10.1080/01431160903464146, 2011.

Jézéquel, A.: Approches statistique et épistémologique de l'attribution d'événements extrêmes, URL http://www.theses.fr/2018SACLV055, thèse de doctorat dirigée par Yiou, Pascal Océan, atmosphère, climat et observations spatiales Université Paris-Saclay (ComUE) 2018, 2018.

Jézéquel, A., Bevacqua, E., d'Andrea, F., Thao, S., Vautard, R., Vrac, M., and Yiou, P.: Conditional and residual trends of singular hot days in Europe, Environ. Res. Lett., 15, 064 018, https://doi.org/10.1088/1748-9326/ab76dd, 2020.

Johnson, M.: Multivariate Statistical Simulation, 230 pp, Wiley, 1987.

Johnson, N. and Kotz, S.: Distributions in Statistics-4. Continuous Multivariate Distributions, 333 pp, New York, Wiley, 1972.

Jonsdottir, H., Madsen, H., and Palsson, O. P.: Parameter estimation in stochastic rainfall-runoff models, J. Hydrol., 326, 379–393, https://doi.org/10.1016/j.jhydrol.2005.11.004, 2006.

Kallache, M., Vrac, M., Naveau, P., and Michelangeli, P.-A.: Non-stationary probabilistic downscaling of extreme precipitation, J. Geophys. Res. Atmos., 116, D05 113, https://doi.org/10.1029/2010JD014892, 2011.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Leetmaa, A., Reynolds, R., Jenne, R., and Joseph, D.: The NCEP/NCAR 40-Year Reanalysis Project, Bull. Am. Meteorol. Soc., 77, 437–472, https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2, 1996.

Keisler, R.: Forecasting Global Weather with Graph Neural Networks, https://doi.org/10.48550/ARXIV.2202.07575, 2022.

Kessler, W. S.: EOF Representations of the Madden–Julian Oscillation and Its Connection with ENSO, J. Clim., 14, 3055–3061, https://doi.org/10.1175/1520-0442(2001)014<3055:EROTMJ>2.0.CO;2, 2001.

Kew, S. F., Selten, F. M., Lenderink, G., and Hazeleger, W.: The simultaneous occurrence of surge and discharge extremes for the Rhine delta, Nat. Hazards Earth Syst. Sci., 13, 2017–2029, https://doi.org/10.5194/nhess-13-2017-2013, 2013.

Kharin, V. and Zwiers, F.: Changes in the Extremes in an Ensemble of Transient Climate Simulation With a Coupled Atmosphere-Ocean GCM, J. Climate, 13, 3760–3788, https://doi.org/10.1175/1520-0442(2000)013<3760:CITEIA>2.0.CO;2, 2000.

Kharin, V., Zwiers, F., Zhang, X., and Wehner, M.: Changes in Temperature and Precipitation Extremes in the CMIP5 Ensemble, Clim. Change, 119, https://doi.org/10.1007/s10584-013-0705-8, 2013.

Kim, T., Cha, M., Kim, H., Lee, J. K., and Kim, J.: Learning to Discover Cross-Domain Relations with Generative Adversarial Networks, 2017.

King, A. D., Donat, M. G., Fischer, E. M., Hawkins, E., Alexander, L. V., Karoly, D. J., Dittus, A. J., Lewis, S. C., and Perkins, S. E.: The timing of anthropogenic emergence in simulated climate extremes, Environ. Res. Lett., 10, 094 015, https://doi.org/10.1088/1748-9326/10/9/094015, 2015.

Kingma, D. P. and Welling, M.: Auto-Encoding Variational Bayes, https://doi.org/10.48550/ARXIV.1312.6114, 2013.

Kiriliouk, A. and Naveau, P.: Climate extreme event attribution using multivariate peaks-over-thresholds modeling and counterfactual theory, Ann. Appl. Stat., 14, 1342–1358, https://doi.org/10.1214/20-AOAS1355, 2020.

Kopp, J., Rivoire, P., Ali, S. M., Barton, Y., and Martius, O.: A novel method to identify subseasonal clustering episodes of extreme precipitation events and their contributions to large accumulation periods, Hydrol. Earth. Syst. Sci., 25, 5153–5174, https://doi.org/10.5194/hess-25-5153-2021, 2021.

Krinner, G. and Flanner, M. G.: Striking stationarity of large-scale climate model bias patterns under strong climate change, Proc. Natl. Acad. Sci. U.S.A., 115, 9462–9466, https://doi.org/10.1073/pnas.1807912115, 2018.

Krizhevsky, A., Sutskever, I., and Hinton, G. E.: ImageNet Classification with Deep Convolutional Neural Networks, Commun. ACM, 60, 84–90, https://doi.org/10.1145/3065386, 2017.

Kundzewicz, Z. W., Mata, L. J., Arnell, N. W., Döll, P., Kabat, P., Jiménez, B., Miller, K. A., Oki, T., Sen, Z., and Shiklomanov, I. A.: Freshwater resources and their management, in: *Climate Change 2007: Impacts, Adaptation and Vulnerability – Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge University Press, Cambridge, UK and New York, NY, USA, 2007.

Köppen, W. P.: Grundriss der klimakunde, 1931.

Laflamme, E. M., Linder, E., and Pan, Y.: Statistical downscaling of regional climate model output to achieve projections of precipitation extremes, Weather. Clim. Extremes, 12, 15–23, https://doi.org/10.1016/j.wace.2015.12.001, 2016.

Lan, M., Gardoni, P., Luo, R., Zhu, J., and Lo, S.: Risk-driven statistical modeling for hurricane-induced compound events: Design event implementation for industrial areas subjected to coastal floods and winds, Ocean Eng., 251, 111 159, 2022.

Laplace, P. S.: Mémoire sur les intégrales définies et leur application aux probabilités, Mémoires de l'Institut Imperial de France, pp. 279–347, 1811.

Laux, P., Rötter, R. P., Webber, H., Dieng, D., Rahimi, J., Wei, J., Faye, B., Srivastava, A. K., Bliefernicht, J., Adeyeri, O., Arnault, J., and Kunstmann, H.: To bias correct or not to bias correct? An agricultural impact modelers' perspective on regional climate model data, Agric. For. Meteorol., 304–305, 108 406, https://doi.org/10.1016/j.agrformet.2021.108406, 2021.

Lawrence, S., Giles, C., Tsoi, A. C., and Back, A.: Face recognition: a convolutional neural-network approach, IEEE Trans. Neural Netw. Learn. Syst., 8, 98–113, https://doi.org/10.1109/72.554195, 1997.

Le Treut, H., Somerville, R., Cubasch, U., Ding, Y., Mauritzen, C., Mokssit, A., Peterson, T., and Prather, M.: Historical Overview of Climate Change, pp. 589–662, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2007.

Lecun, Y. and Bengio, Y.: Convolutional Networks for Images, Speech, and Time-Series, MIT Press, 1995.

Lee, J., Weger, R., Sengupta, S., and Welch, R.: A neural network approach to cloud classification, IEEE Trans. Geosci. Remote Sens., 28, 846–855, https://doi.org/10.1109/36.58972, 1990.

Legasa, M., Manzanas, R., Calviño, A., and Gutiérrez, J.: A Posteriori Random Forests for Stochastic Downscaling of Precipitation by Predicting Probability Distributions, Water Resour. Res., 58, https://doi.org/10.1029/2021WR030272, 2022.

Leinonen, J., Nerini, D., and Berne, A.: Stochastic Super-Resolution for Downscaling Time-Evolving Atmospheric Fields with a Generative Adversarial Network, IEEE Trans. Geosci. Remote Sens., pp. 1–13, https://doi.org/10.1109/TGRS.2020.3032790, 2020.

Lenderink, G., Buishand, A., and van Deursen, W.: Estimates of future discharges of the river Rhine using two scenario methodologies: direct versus delta approach, Hydrol. Earth Syst. Sci., 11, 1145–1159, https://doi.org/10.5194/hess-11-1145-2007, 2007.

Lenggenhager, S. and Martius, O.: Quantifying the link between heavy precipitation and Northern Hemisphere blocking—A Lagrangian analysis, Atmos. Sci. Lett., 21, e999, https://doi.org/10.1002/asl.999, 2020.

Leonard, M., Westra, S., Phatak, A., Lambert, M., Hurk, B., Mcinnes, K., Risbey, J., Schuster, S., Jakob, D., and Stafford Smith, M.: A compound event framework for understanding extreme impacts, Wiley Interdiscip. Rev. Clim. Change, 2014.

Li, H., Li, Y., Huang, G., and Sun, J.: Quantifying effects of compound dry-hot extremes on vegetation in Xinjiang (China) using a vine-copula conditional probability model, Agric. For. Meteorol., 311, 108 658, https://doi.org/10.1016/j.agrformet.2021.108658, 2021.

Li, W., Gao, X., Hao, Z., and Sun, R.: Using deep learning for precipitation forecasting based on spatio-temporal information: a case study, Clim. Dyn., 58, 443–457, https://doi.org/10.1007/s00382-021-05916-4, 2022.

Liu, M.-Y., Breuel, T., and Kautz, J.: Unsupervised Image-to-Image Translation Networks, https://doi.org/10.48550/ARXIV.1703.00848, 2017.

Liu, Y., Racah, E., Prabhat, Correa, J., Khosrowshahi, A., Lavers, D., Kunkel, K., Wehner, M., and Collins, W.: Application of Deep Convolutional Neural Networks for Detecting Extreme Weather in Climate Datasets, 2016.

Lobell, D. B. and Burke, M. B.: Why are agricultural impacts of climate change so uncertain? The importance of temperature relative to precipitation, Environ. Res. Lett., 3, 034 007, https://doi.org/10.1088/1748-9326/3/3/034007, 2008.

MacQueen, J.: Some methods for classification and analysis of multivariate observations, Proc. Fifth Berkeley Symp. on Math. Statist. and Prob., 1, 281–297, 1967.

Maher, P., Vallis, G. K., Sherwood, S. C., Webb, M. J., and Sansom, P. G.: The Impact of Parameterized Convection on Climatological Precipitation in Atmospheric Global Climate Models, Geophys. Res. Lett., 45, 3728–3736, https://doi.org/10.1002/2017GL076826, 2018.

Mahlstein, I., Knutti, R., Solomon, S., and Portmann, R. W.: Early onset of significant local warming in low latitude countries, Environ. Res. Lett., 6, 034 009, https://doi.org/10.1088/1748-9326/6/3/034009, 2011.

Mahlstein, I., Hegerl, G., and Solomon, S.: Emerging local warming signals in observational data, Geophys. Res. Lett., 39, https://doi.org/10.1029/2012GL053952, 2012.

Mallakpour, I., Villarini, G., Jones, M., and Smith, J.: On the use of Cox regression to examine the temporal clustering of flooding and heavy precipitation across the central United States, Glob. Planet Change, 155, 98–108, https://doi.org/10.1016/j.gloplacha.2017.07.001, 2017.

Manabe, S. and Bryan, K.: Climate Calculations with a Combined Ocean-Atmosphere Model, J. Atmos. Sci., 26, 786–789, https://doi.org/10.1175/1520-0469(1969)026<0786:CCWACO>2.0.CO;2, 1969.

Manning, C., Widmann, M., Bevacqua, E., Loon, A. F. V., Maraun, D., and Vrac, M.: Increased probability of compound long-duration dry and hot events in Europe during summer (1950–2013), Environ. Res. Lett., 14, 094 006, https://doi.org/10.1088/1748-9326/ab23bf, 2019.

Manning, C., Bevacqua, E., Widmann, M., Maraun, D., and Loon, A.: Large discrepancies in the representation of compound long-duration dry and hot spells over Europe in CMIP5, https://doi.org/10.5194/wcd-2022-15, 2022.

Mao, G., Wang, M., Liu, J., Wang, Z., Wang, K., Meng, Y., Zhong, R., Wang, H., and Li, Y.: Comprehensive comparison of artificial neural networks and long short-term memory networks for rainfall-runoff simulation, Phys. Chem. Earth, 123, 103 026, https://doi.org/10.1016/j.pce.2021.103026, 2021.

Maraun, D.: Bias Correction, Quantile Mapping, and Downscaling: Revisiting the Inflation Issue, J. Climate, 26, 2137–2143, https://doi.org/10.1175/JCLI-D-12-00821.1, 2013a.

Maraun, D.: When will trends in European mean and heavy daily precipitation emerge?, Environ. Res. Lett., 8, 014 004, https://doi.org/10.1088/1748-9326/8/1/014004, 2013b.

Maraun, D.: Bias Correcting Climate Change Simulations - A Critical Review, Curr. Clim. Chang. Reports, 2, 211–220, https://doi.org/10.1007/s40641-016-0050-x, 2016.

Maraun, D. and Widmann, M.: Statistical Downscaling and Bias Correction for Climate Research, Cambridge University Press, https://doi.org/10.1017/9781107588783, 2018.

Maraun, D., Wetterhall, F., Ireson, A. M., Chandler, R. E., Kendon, E. J., Widmann, M., Brienen, S., Rust, H. W., Sauter, T., Themeßl, M., Venema, V. K. C., Chun, K. P., Goodess, C. M., Jones, R. G., Onof, C., Vrac, M., and Thiele-Eich, I.: Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user, Rev. Geophys., 48, https://doi.org/10.1029/2009RG000314, 2010.

Maraun, D., Shepherd, T. G., Widmann, M., Zappa, G., Walton, D., Gutiérrez, J. M., Hagemann, S., Richter, I., Soares, P. M. M., Hall, A., and Mearns, L. O.: Towards process-informed bias correction of climate change simulations, Nat. Clim. Change, 7, 764–773, https://doi.org/10.1038/nclimate3418, 2017.

Maraun, D., Truhetz, H., and Schaffer, A.: Regional Climate Model Biases, Their Dependence on Synoptic Circulation Biases and the Potential for Bias Adjustment: A Process-Oriented Evaluation of the Austrian Regional Climate Projections, J. Geophys. Res. Atmos., 126, e2020JD032 824, https://doi.org/10.1029/2020JD032824, 2021.

Martius, O., Pfahl, S., and Chevalier, C.: A global quantification of compound precipitation and wind extremes, Geophys. Res. Lett., 43, 7709–7717, https://doi.org/10.1002/2016GL070017, 2016.

Mazdiyasni, O., Sadegh, M., Chiang, F., and AghaKouchak, A.: Heat wave Intensity Duration Frequency Curve: A Multivariate Approach for Hazard and Attribution Analysis, Sci. Rep., 9, 14 117, https://doi.org/10.1038/s41598-019-50643-w, 2019.

McCulloch, W. S. and Pitts, W.: A logical calculus of the ideas immanent in nervous activity, Bull. Math. Biophys., 5, 115–133, https://doi.org/10.1007/BF02478259, 1943.

Mehrotra, R. and Sharma, A.: A Resampling Approach for Correcting Systematic Spatiotemporal Biases for Multiple Variables in a Changing Climate, Water Resour. Res., 55, 754–770, https://doi.org/10.1029/2018WR023270, 2019.

Meyer, J., Kohn, I., Stahl, K., Hakala, K., Seibert, J., and Cannon, A. J.: Effects of univariate and multivariate bias correction on hydrological impact projections in alpine catchments, Hydrol. Earth Syst. Sci., 23, 1339–1354, https://doi.org/10.5194/hess-23-1339-2019, 2019.

Michelangeli, P.-A., Vrac, M., and Loukos, H.: Probabilistic downscaling approaches: Application to wind cumulative distribution functions, Geophys. Res. Lett., 36, L11 708, https://doi.org/10.1029/2009GL038401, 2009.

Milankovitch, M.: Mathematische Klimalehre und Astronomische Theorie der Klimaschwankungen, Borntraeger, Berlin, 1930.

Mirza, M. and Osindero, S.: Conditional Generative Adversarial Nets, 2014.

Moftakhari, H. R., Salvadori, G., AghaKouchak, A., Sanders, B. F., and Matthew, R. A.: Compounding effects of sea level rise and fluvial flooding, PNAS, 114, 9785–9790, https://doi.org/10.1073/pnas.1620325114, 2017.

Moghim, S. and Bras, R. L.: Bias Correction of Climate Modeled Temperature and Precipitation Using Artificial Neural Networks, J. Hydrometeorol., 18, 1867–1884, https://doi.org/10.1175/JHM-D-16-0247.1, 2017.

Nahar, J., Johnson, F., and Sharma, A.: Addressing Spatial Dependence Bias in Climate Model Simulations–An Independent Component Analysis Approach, Water Resour. Res., 54, 827–841, https://doi.org/10.1002/2017WR021293, 2018.

Nguyen, H., Mehrotra, R., and Sharma, A.: Correcting systematic biases across multiple atmospheric variables in the frequency domain, Clim. Dynam., 52, 1283–1298, https://doi.org/10.1007/s00382-018-4191-6, 2019.

Nicolet, G., Eckert, N., Morin, S., and Blanchet, J.: Decreasing spatial dependence in extreme snowfall in the French Alps since 1958 under climate change, J. Geophys. Res. Atmos., 121, 8297–8310, https://doi.org/10.1002/2016JD025427, 2016.

Oesting, M. and Stein, A.: Spatial modeling of drought events using max-stable processes, Stoch. Environ. Res. Risk Assess., 32, https://doi.org/10.1007/s00477-017-1406-z, 2018.

Ossó, A., Allan, R., Hawkins, E., Shaffrey, L., and Maraun, D.: Emerging new climate extremes over Europe, Clim. Dyn., 58, https://doi.org/10.1007/s00382-021-05917-3, 2022.

Otto, F. E. L., Wolski, P., Lehner, F., Tebaldi, C., van Oldenborgh, G. J., Hogesteeger, S., Singh, R., Holden, P., Fučkar, N. S., Odoulami, R. C., and New, M.: Anthropogenic influence on the drivers of the Western Cape drought 2015–2017, Environ. Res. Lett., 13, 124 010, https://doi.org/10.1088/1748-9326/aae9f9, 2018.

Owen, L. E., Catto, J. L., Stephenson, D. B., and Dunstone, N. J.: Compound precipitation and wind extremes over Europe and their relationship to extratropical cyclones, Weather. Clim. Extremes, 33, 100 342, https://doi.org/10.1016/j.wace.2021.100342, 2021.

Pan, B., Anderson, G., Gonçalves, A., Lucas, D., Bonfils, C., Lee, J., Tian, Y., and Ma, H.-Y.: Learning to Correct Climate Projection Biases, J. Adv. Model. Earth Syst., https://doi.org/10.1029/2021MS002509, 2021.

Panofsky, H. and Brier, G.: Some applications of statistics to meteorology, Earth and Mineral Sciences Continuing Education, College of Earth and Mineral Sciences, 103 pp., 1968.

Papadimitriou, L. V., Koutroulis, A. G., Grillakis, M. G., and Tsanis, I. K.: The effect of GCM biases on global runoff simulations of a land surface model, Hydrol. Earth Syst. Sci., 21, 4379–4401, https://doi.org/10.5194/hess-21-4379-2017, 2017.

Pearson, K.: LIII. On lines and planes of closest fit to systems of points in space, London, Edinburgh Dublin Philos. Mag. J. Sci., 2, 559–572, https://doi.org/10.1080/14786440109462720, 1901.

Pearson, K. and Henrici, O. M. F. E.: VII. Mathematical contributions to the theory of evolution.—III. Regression, heredity, and panmixia, Philos. Trans. Royal Soc. A, 187, 253–318, https://doi.org/10.1098/rsta.1896.0007, 1896.

Petzka, H., Fischer, A., and Lukovnicov, D.: On the regularization of Wasserstein GANs, https://doi.org/10.48550/ARXIV.1709.08894, 2017.

Pfleiderer, P., Menke, I., and Schleussner, C.-F.: Increasing risks of apple tree frost damage under climate change, Clim. Change, 157, https://doi.org/10.1007/s10584-019-02570-y, 2019.

Philip, S., Kew, S. F., van Oldenborgh, G. J., Aalbers, E., Vautard, R., Otto, F., Haustein, K., Habets, F., and Singh, R.: Validation of a Rapid Attribution of the May/June 2016 Flood-Inducing Precipitation in France to Climate Change, J. Hydrometeorol., 19, 1881–1898, https://doi.org/10.1175/JHM-D-18-0074.1, 2018.

Phillips, S. J., Anderson, R. P., and Schapire, R. E.: Maximum entropy modeling of species geographic distributions, Ecol. Modell., 190, 231–259, https://doi.org/10.1016/j.ecolmodel.2005.03.026, 2006.

Pohl, E., Grenier, C., Vrac, M., and Kageyama, M.: Emerging climate signals in the Lena River catchment: a non-parametric statistical approach, Hydrol. Earth Syst. Sci., 24, 2817–2839, https://doi.org/10.5194/hess-24-2817-2020, 2020.

Qian, Q., Jia, X., Lin, H., and Zhang, R.: Seasonal Forecast of Nonmonsoonal Winter Precipitation over the Eurasian Continent Using Machine-Learning Models, J. Clim., 34, 7113–7129, https://doi.org/10.1175/JCLI-D-21-0113.1, 2021.

Quesada-Chacón, D., Barfus, K., and Bernhofer, C.: Climate change projections and extremes for Costa Rica using tailored predictors from CORDEX model output through statistical downscaling with artificial neural networks, Int. J. Climatol., 41, 211–232, https://doi.org/10.1002/joc.6616, 2021.

Racah, E., Beckham, C., Maharaj, T., Kahou, S. E., Prabhat, and Pal, C.: ExtremeWeather: A large-scale climate dataset for semi-supervised detection, localization, and understanding of extreme weather events, 2017.

Radford, A., Metz, L., and Chintala, S.: Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks, 2016.

Ranzato, M., Poultney, C., Chopra, S., and Lecun, Y.: Efficient learning of sparse representations with an energy-based model, in: NIPS, pp. 1137–1144, MIT Press, 2006.

Rasp, S., Dueben, P. D., Scher, S., Weyn, J. A., Mouatadid, S., and Thuerey, N.: WeatherBench: A Benchmark Data Set for Data-Driven Weather Forecasting, J. Adv. Model. Earth Syst., 12, e2020MS002 203, https://doi.org/10.1029/2020MS002203, e2020MS002203 10.1029/2020MS002203, 2020.

Raymond, C., Matthews, T., and Horton, R. M.: The emergence of heat and humidity too severe for human tolerance, Sci. Adv., 6, https://doi.org/10.1126/sciadv.aaw1838, 2020.

Raymond, C., Suarez-Gutierrez, L., Kornhuber, K., Pascolini-Campbell, M., Sillmann, J., and Waliser, D. E.: Increasing spatiotemporal proximity of heat and precipitation extremes in a warming world quantified by a large model ensemble, Environ. Res. Lett., 17, 035 005, https://doi.org/10.1088/1748-9326/ac5712, 2022.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., and Prabhat, M.: Deep learning and process understanding for data-driven Earth system science, Nature, 566, 195–204, https://doi.org/10.1038/s41586-019-0912-1, 2019.

Riahi, K., van Vuuren, D. P., Kriegler, E., Edmonds, J., O'Neill, B. C., Fujimori, S., Bauer, N., Calvin, K., Dellink, R., Fricko, O., Lutz, W., Popp, A., Cuaresma, J. C., KC, S., Leimbach, M., Jiang, L., Kram, T., Rao, S., Emmerling, J., Ebi, K., Hasegawa, T., Havlik, P., Humpenöder, F., Da Silva, L. A., Smith, S., Stehfest, E., Bosetti, V., Eom, J., Gernaat, D., Masui, T., Rogelj, J., Strefler, J., Drouet, L., Krey, V., Luderer, G., Harmsen, M., Takahashi, K., Baumstark, L., Doelman, J. C., Kainuma, M., Klimont, Z., Marangoni, G., Lotze-Campen, H., Obersteiner, M., Tabeau, A., and Tavoni, M.: The Shared Socioeconomic Pathways and their energy, land use, and greenhouse gas emissions implications: An overview, Global Environmental Change, 42, 153–168, https://doi.org/10.1016/j.gloenvcha.2016.05.009, 2017.

Ribeiro, A. F. S., Russo, A., Gouveia, C. M., Páscoa, P., and Zscheischler, J.: Risk of crop failure due to compound dry and hot extremes estimated with nested copulas, Biogeosciences, 17, 4815–4830, https://doi.org/10.5194/bg-17-4815-2020, 2020.

Ribes, A., Zwiers, F. W., Azaïs, J.-M., and Naveau, P.: A new statistical approach to climate change detection and attribution, Clim. Dyn., 48, 367–386, https://doi.org/10.1007/s00382-016-3079-6, 2017.

Richardson, L. F.: Weather prediction by numerical methods, Cambridge: Cambridge University Press, 1922.

Ridder, N., de Vries, H., and Drijfhout, S.: The role of atmospheric rivers in compound events consisting of heavy precipitation and high storm surges along the Dutch coast, Nat. Hazards Earth Syst. Sci., 18, 3311–3326, https://doi.org/10.5194/nhess-18-3311-2018, 2018.

Ridder, N., Pitman, A., and Ukkola, A.: Do CMIP6 Climate Models simulate Global or Regional Compound Events skilfully?, Geophys. Res. Lett., 48, https://doi.org/10.1029/2020GL091152, 2021.

Ridder, N., Ukkola, A., Pitman, A., and Perkins-Kirkpatrick, S.: Increased occurrence of high impact compound events under climate change, NPJ Clim. Atmos. Sci., 5, https://doi.org/10.1038/s41612-021-00224-4, 2022.

Ridder, N. N., Pitman, A. J., Westra, S., Ukkola, A., Do, H. X., Bador, M., Hirsch, A. L., Evans, J. P., Di Luca, A., and Zscheischler, J.: Global hotspots for the occurrence of compound events, Nat. Commun., 11, 5956, https://doi.org/10.1038/s41467-020-19639-3, 2020.

Robin, Y. and Vrac, M.: Is time a variable like the others in multivariate statistical downscaling and bias correction?, ESDD, 2021, 1–32, https://doi.org/10.5194/esd-2021-12, 2021.

Robin, Y., Vrac, M., Naveau, P., and Yiou, P.: Multivariate stochastic bias corrections with optimal transport, Hydrol. Earth Syst. Sci., 23, 773–786, https://doi.org/10.5194/hess-23-773-2019, 2019.

Rohde, R., Muller, R., Jacobsen, R., Muller, E., and Wickham, C.: A New Estimate of the Average Earth Surface Land Temperature Spanning 1753 to 2011, Geoinformatics geostat., 01, https://doi.org/10.4172/2327-4581.1000101, 2013.

Rosenblatt, F.: The perceptron: a probabilistic model for information storage and organization in the brain., Psychol. Rev., 65 6, 386–408, 1958.

Rudin, C.: Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead, https://doi.org/10.48550/ARXIV.1811.10154, 2018.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J.: Learning representations by back-propagating errors, Nature, 323, 533–536, https://doi.org/10.1038/323533a0, 1986.

Rummukainen, M.: State-of-the-art with regional climate models, WIREs Climate Change, 1, 82–96, https://doi.org/10.1002/wcc.8, 2010.

Russo, F., Lombardo, F., Napolitano, F., and Gorgucci, E.: Rainfall stochastic modeling for runoff forecasting, Phys. Chem. Earth, 31, 1252–1261, https://doi.org/10.1016/j.pce.2006.06.002, 2006.

Rust, H. W., Vrac, M., Sultan, B., and Lengaigne, M.: Mapping Weather-Type Influence on Senegal Precipitation Based on a Spatial–Temporal Statistical Model, J. Climate, 26, 8189–8209, https://doi.org/10.1175/JCLI-D-12-00302.1, 2013.

Räty, O., Räisänen, J., Bosshard, T., and Donnelly, C.: Intercomparison of Univariate and Joint Bias Correction Methods in Changing Climate From a Hydrological Perspective, Climate, 6, 33, https://doi.org/10.3390/cli6020033, 2018.

Sachindra, D., Ahmed, K., Rashid, M. M., Shahid, S., and Perera, B.: Statistical downscaling of precipitation using machine learning techniques, Atmos. Res., 212, 240–258, https://doi.org/10.1016/j.atmosres.2018.05.022, 2018.

Salman, A. G., Kanigoro, B., and Heryadi, Y.: Weather forecasting using deep learning techniques, in: 2015 International Conference on Advanced Computer Science and Information Systems (ICACSIS), pp. 281–285, https://doi.org/10.1109/ICACSIS.2015.7415154, 2015.

Salvadori, G. and De Michele, C.: Frequency analysis via copulas: Theoretical aspects and applications to hydrological events, Water Resour. Res., 40, https://doi.org/10.1029/2004WR003133, 2004.

Sarhadi, A., Ausín, M. C., Wiper, M. P., Touma, D., and Diffenbaugh, N. S.: Multidimensional risk in a nonstationary climate: Joint probability of increasingly severe warm and dry conditions, Sci. Adv., 4, eaau3487, https://doi.org/10.1126/sciadv.aau3487, 2018.

Scher, S. and Messori, G.: Predicting weather forecast uncertainty with machine learning, Q. J. R. Meteorol. Soc., 144, 2830–2841, https://doi.org/10.1002/qj.3410, 2018.

Scher, S. and Messori, G.: Weather and climate forecasting with neural networks: using general circulation models (GCMs) with different complexity as a study ground, Geosci. Model Dev., 12, 2797–2809, https://doi.org/10.5194/gmd-12-2797-2019, 2019.

Scherrer, S., Croci-Maspoli, M., Schwierz, C., and Appenzeller, C.: Two-dimensional indices of atmospheric blocking and their statistical relationship with winter climate patterns in the Euro-Atlantic region, Int. J. Climatol., 26, 233–249, https://doi.org/10.1002/joc.1250, 2006.

Schmidhuber, J.: Deep learning in neural networks: An overview, Neural Netw., 61, 85 – 117, https://doi.org/10.1016/j.neunet.2014.09.003, 2015.

Schmith, T., Thejll, P., Berg, P., Boberg, F., Christensen, O. B., Christiansen, B., Christensen, J. H., Madsen, M. S., and Steger, C.: Identifying robust bias adjustment methods for European extreme precipitation in a multi-model pseudo-reality setting, Hydrol. Earth Syst. Sci., 25, 273–290, https://doi.org/10.5194/hess-25-273-2021, 2021.

Schnell, J. L. and Prather, M. J.: Co-occurrence of extremes in surface ozone, particulate matter, and temperature over eastern North America, PNAS, 114, 2854–2859, https://doi.org/10.1073/pnas.1614453114, 2017.

Seneviratne, S., Nicholls, N., Easterling, D., Goodess, C., Kanae, S., Kossin, J., Luo, Y., Marengo, J., McInnes, K., Rahimi, M., Reichstein, M., Sorteberg, A., Vera, C., and Zhang, X.: Changes in climate extremes and their impacts on the natural physical environment. In: *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation* [Field, C.B., V. Barros, T.F. Stocker, D. Qin, D.J. Dokken, K.L. Ebi, M.D. Mastrandrea, K.J. Mach, G.-K. Plattner, S.K. Allen, M. Tignor, and P.M. Midgley (eds.)], pp. 109–230, 2012.

Serinaldi, F. and Kilsby, C. G.: A Blueprint for Full Collective Flood Risk Estimation: Demonstration for European River Flooding, Risk Anal., 37, 1958–1976, https://doi.org/10.1111/risa.12747, 2017.

Sgubin, G., Swingedouw, D., Dayon, G., Garcia de Cortazar-Atauri, I., Ollat, N., Page, C., and van Leeuwen, C.: The risk of tardive frost damage in French vineyards in a changing climate, Agric. For. Meteorol., 250–251, 226–242, https://doi.org/10.1016/j.agrformet.2017.12.253, 2018.

Shepherd, T. G.: A Common Framework for Approaches to Extreme Event Attribution, Curr. Clim. Change Rep., 2, 28–38, https://doi.org/10.1007/s40641-016-0033-y, 2016.

Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W., and Woo, W.: Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting, 2015.

Shojaeezadeh, S. A., Nikoo, M. R., Mirchi, A., Mallakpour, I., AghaKouchak, A., and Sadegh, M.: Probabilistic hazard assessment of contaminated sediment in rivers, Sci. Total Environ., 703, 134 875, https://doi.org/10.1016/j.scitotenv.2019.134875, 2020.

Singh, J., Ashfaq, M., Skinner, C. B., Anderson, W. B., and Singh, D.: Amplified risk of spatially compounding droughts during co-occurrences of modes of natural ocean variability, NPJ Clim. Atmos. Sci., 4, 7, https://doi.org/10.1038/s41612-021-00161-2, 2021.

Sklar, A.: Fonctions de Répartition à n Dimensions et Leurs Marges, Publications de l'Institut Statistique de l'Université de Paris, 8, 229–231, 1959.

Smith, J. and Eli, R. N.: Neural-Network Models of Rainfall-Runoff Process, J. Water Resour. Plan. Manag., 121, 499–508, https://doi.org/10.1061/(ASCE)0733-9496(1995)121:6(499), 1995.

Specht, D.: A general regression neural network, IEEE Trans. Neural Netw. Learn. Syst., 2, 568–576, https://doi.org/10.1109/72.97934, 1991.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R.: Dropout: A Simple Way to Prevent Neural Networks from Overfitting, J. Mach. Learn. Res., 15, 1929–1958, 2014.

Steinhaus, H.: Sur la division des corps mat eriels en parties, Bull. Int. Acad. Polon. Sci., Classe III, vol. IV, no. 12, 801–804, 1956.

Storch, H. and Zwiers, F.: Statistical Analysis in Climate Research, Cambridge University Press, 1999.

Stott, P. A., Stone, D. A., and Allen, M. R.: Human contribution to the European heatwave of 2003, Nature, 432, 610–614, https://doi.org/10.1038/nature03089, 2004.

Stott, P. A., Christidis, N., Otto, F. E. L., Sun, Y., Vanderlinden, J.-P., van Oldenborgh, G. J., Vautard, R., von Storch, H., Walton, P., Yiou, P., and Zwiers, F. W.: Attribution of extreme weather and climate-related events, Wiley Interdiscip. Rev. Clim. Change, 7, 23–41, https://doi.org/10.1002/wcc.380, 2016.

Suarez-Gutierrez, L., Milinski, S., and Maher, N.: Exploiting large ensembles for a better yet simpler climate model evaluation, Clim. Dynam., 57, 2557–2580, https://doi.org/10.1007/s00382-021-05821-w, 2021.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A.: Going Deeper with Convolutions, in: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1–9, https://doi.org/10.1109/CVPR.2015.7298594, 2015.

Tani, S. and Gobiet, A.: Quantile mapping for improving precipitation extremes from regional climate models, J. Agrometeorol., 21, 434–443, https://doi.org/10.54386/jam.v21i4.278, 2019.

Tao, F., Rötter, R. P., Palosuo, T., Gregorio Hernández Díaz-Ambrona, C., Mínguez, M. I., Semenov, M. A., Kersebaum, K. C., Nendel, C., Specka, X., Hoffmann, H., Ewert, F., Dambreville, A., Martre, P., Rodríguez, L., Ruiz-Ramos, M., Gaiser, T., Höhn, J. G., Salo, T., Ferrise, R., Bindi, M., Cammarano, D., and Schulman, A. H.: Contribution of crop model structure, parameters and climate projections to uncertainty in climate change impact assessments, Glob. Change Biol. Bioenergy, 24, 1291–1307, https://doi.org/10.1111/gcb.14019, 2018.

Tavakol, A., Rahmani, V., and Jr, J.: Probability of compound climate extremes in a changing climate: A copula-based study of hot, dry, and windy events in the central United States, Environ. Res. Lett., 15, https://doi.org/10.1088/1748-9326/abb1ef, 2020.

Taylor, K., Ronald, S., and Meehl, G.: An overview of CMIP5 and the Experiment Design, Bull. Am. Meteorol. Soc., 93, 485–498, https://doi.org/10.1175/BAMS-D-11-00094.1, 2011.

Teixeira, J. and Reynolds, C. A.: Stochastic Nature of Physical Parameterizations in Ensemble Prediction: A Stochastic Convection Approach, Mon. Weather Rev., 136, 483–496, https://doi.org/10.1175/2007MWR1870.1, 2008.

Tian, B., Shaikh, M., Azimi-Sadjadi, M., Haar, T., and Reinke, D.: A study of cloud classification with neural networks using spectral and textural features, IEEE Trans. Neural Netw. Learn. Syst., 10, 138–151, https://doi.org/10.1109/72.737500, 1999.

Tompson, J., Schlachter, K., Sprechmann, P., and Perlin, K.: Accelerating Eulerian Fluid Simulation With Convolutional Networks, CoRR, abs/1607.03597, URL http://arxiv.org/abs/1607.03597, 2016.

Tramblay, Y., Ruelland, D., Somot, S., Bouaicha, R., and Servat, E.: High-resolution Med-CORDEX regional climate model simulations for hydrological impact studies: a first evaluation of the ALADIN-Climate model in Morocco, Hydrol. Earth Syst. Sci., 17, 3721–3739, https://doi.org/10.5194/hess-17-3721-2013, 2013.

Vaittinada Ayar, P., Vrac, M., Bastin, S., Carreau, J., Déqué, M., and Gallardo, C.: Intercomparison of statistical and dynamical downscaling models under the EURO- and MED-CORDEX initiative framework: present climate evaluations, Clim. Dynam., 46, 1301–1329, https://doi.org/10.1007/s00382-015-2647-5, 2016.

Vaittinada Ayar, P., Vrac, M., and Mailhot, A.: Ensemble bias correction of climate simulations: preserving internal variability, Scientific Reports, 11, 3098, https://doi.org/10.1038/s41598-021-82715-1, 2021.

Van de Velde, J., Demuzere, M., De Baets, B., and Verhoest, N. E. C.: Impact of bias nonstationarity on the performance of uni- and multivariate bias-adjusting methods: a case study on data from Uccle, Belgium, Hydrology and Earth System Sciences, 26, 2319–2344, https://doi.org/10.5194/hess-26-2319-2022, 2022.

van den Hurk, B., van Meijgaard, E., de Valk, P., van Heeringen, K.-J., and Gooijer, J.: Analysis of a compounding surge and precipitation event in the Netherlands, Environ. Res. Lett., 10, 035 001, https://doi.org/10.1088/1748-9326/10/3/035001, 2015.

van der Wiel, K., Kapnick, S. B., van Oldenborgh, G. J., Whan, K., Philip, S., Vecchi, G. A., Singh, R. K., Arrighi, J., and Cullen, H.: Rapid attribution of the August 2016 flood-inducing extreme precipitation in south Louisiana to climate change, Hydrol. Earth Syst. Sci., 21, 897–921, https://doi.org/10.5194/hess-21-897-2017, 2017.

van Oldenborgh, G. J., van der Wiel, K., Kew, S., Philip, S., Otto, F., Vautard, R., King, A., Lott, F., Arrighi, J., Singh, R., and van Aalst, M.: Pathways and pitfalls in extreme event attribution, Clim. Change, 166, 13, https://doi.org/10.1007/s10584-021-03071-7, 2021.

Vandal, T., Kodra, E., Ganguly, S., Michaelis, A., Nemani, R., and Ganguly, A. R.: DeepSD: Generating High Resolution Climate Change Projections through Single Image Super-Resolution, in: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1663–1672, https://doi.org/10.1145/3097983.3098004, 2017.

Vandal, T., Kodra, E., and Ganguly, A. R.: Intercomparison of machine learning methods for statistical downscaling: the case of daily and extreme precipitation, Theor. Appl. Climatol., 137, 557–570, https://doi.org/10.1007/s00704-018-2613-3, 2019.

Vautard, R., Yiou, P., Otto, F., Stott, P., Christidis, N., van Oldenborgh, G. J., and Schaller, N.: Attribution of human-induced dynamical and thermodynamical contributions in extreme weather events, Environ. Res. Lett., 11, 114 009, https://doi.org/10.1088/1748-9326/11/11/114009, 2016.

Vautard, R., van Oldenborgh, G. J., Bonnet, R., Li, S., Philip, S., Soubeyroux, J. M., Dubuisson, B., Viovy, N., Reichstein, M., and Otto, F. E. L.: Human influence on growing period frosts like the early april 2021 in Central France, 2021.

Verschuur, J., Li, S., Wolski, P., and Otto, F.: Climate change as a driver of food insecurity in the 2007 Lesotho-South Africa drought, Sci. Rep., 11, https://doi.org/10.1038/s41598-021-83375-x, 2021.

Vigaud, N., Vrac, M., and Caballero, Y.: Probabilistic downscaling of GCM scenarios over southern India, Int. J. Climatol., 33, 1248–1263, https://doi.org/10.1002/joc.3509, 2013.

Villarini, G., Smith, J. A., Baeck, M. L., Vitolo, R., Stephenson, D. B., and Krajewski, W. F.: On the frequency of heavy rainfall for the Midwest of the United States, J. Hydrol., 400, 103–120, https://doi.org/10.1016/j.jhydrol.2011.01.027, 2011.

Vislocky, R. L. and Fritsch, J. M.: An Automated, Observations-Based System for Short-Term Prediction of Ceiling and Visibility, WAF, 12, 31–43, https://doi.org/10.1175/1520-0434(1997)012<0031:AAOBSF>2.0.CO;2, 1997.

Vlasenko, A., Matthias, V., and Callies, U.: Simulation of chemical transport model estimates by means of a neural network using meteorological data, Atmos. Environ., 254, 118 236, https://doi.org/10.1016/j.atmosenv.2021.118236, 2021.

Vrac, M.: Multivariate bias adjustment of high-dimensional climate simulations: the Rank Resampling for Distributions and Dependences (R^2D^2) bias correction, Hydrol. Earth Syst. Sci., 22, 3175–3196, https://doi.org/10.5194/hess-22-3175-2018, 2018.

Vrac, M. and Thao, S.: R^2D^2 v2.0: Accounting for temporal dependences in multivariate bias correction via analogue ranks resampling, Geosci. Model Dev., 2020, 1–29, https://doi.org/10.5194/gmd-2020-132, 2020.

Vrac, M., Chédin, A., and Diday, E.: Clustering a Global Field of Atmospheric Profiles by Mixture Decomposition of Copulas, J. Atmos. Ocean Technol., 22, 1445–1459, https://doi.org/10.1175/JTECH1795.1, 2005.

Vrac, M., Marbaix, P., Paillard, D., and Naveau, P.: Non-linear statistical downscaling of present and LGM precipitation and temperatures over Europe, Clim. Past, 3, 669–682, https://doi.org/10.5194/cp-3-669-2007, 2007.

Vrac, M., Drobinski, P., Merlo, A., Herrmann, M., Lavaysse, C., Li, L., and Somot, S.: Dynamical and statistical downscaling of the French Mediterranean climate: uncertainty assessment, Nat. Hazards Earth Syst. Sci., 12, 2769–2784, https://doi.org/10.5194/nhess-12-2769-2012, 2012.

Vrac, M., Thao, S., and Yiou, P.: Changes in temperature-precipitation correlations over Europe: Are climate models reliable?, https://doi.org/10.21203/rs.3.rs-1008080/v1, 2021.

Vrac, M., Thao, S., and Yiou, P.: Should multivariate bias corrections of climate simulations account for changes of rank correlation over time?, https://doi.org/10.1002/essoar.10510318.1, 2022.

Wahl, T., Jain, S., Bender, J., Meyers, S., and Luther, M.: Increasing risk of compound flooding from storm surge and rainfall for major US cities, Nat. Clim. Chang., 5, 1093–1097, https://doi.org/10.1038/nclimate2736, 2015.

Walker, G. T.: Correlation in seasonal variations of weather - a further study of world weather, Mon. Weather Rev., 53, 252–254, https://doi.org/10.1175/1520-0493(1925)53<252:CISVOW>2.0.CO;2, 1925.

Wallis, T. W. and Griffiths, J. F.: Simulated meteorological input for agricultural models, Agric. For. Meteorol., 88, 241–258, https://doi.org/10.1016/S0168-1923(97)00035-X, 1997.

Walton, D., Berg, N., Pierce, D., Maurer, E., Hall, A., Lin, Y.-H., Rahimi, S., and Cayan, D.: Understanding Differences in California Climate Projections Produced by Dynamical and Statistical Downscaling, J. Geophys. Res. Atmos., 125, e2020JD032 812, https://doi.org/10.1029/2020JD032812, 2020.

Wang, Z., Yan, J., and Zhang, X.: Incorporating spatial dependence in regional frequency analysis, Water Resour. Res., 50, 9570–9585, https://doi.org/10.1002/2013WR014849, 2014.

Wang, Z., She, Q., and Ward, T. E.: Generative Adversarial Networks in Computer Vision: A Survey and Taxonomy, 2020.

Ward, P. J., Couasnon, A., Eilander, D., Haigh, I. D., Hendry, A., Muis, S., Veldkamp, T. I. E., Winsemius, H. C., and Wahl, T.: Dependence between high sea-level and high river discharge increases flood hazard in global deltas and estuaries, Environ. Res. Lett., 13, 084 012, https://doi.org/10.1088/1748-9326/aad400, 2018.

Watson-Parris, D., Rao, Y., Olivié, D., Seland, , Nowack, P. J., Camps-Valls, G., Stier, P., Bouabid, S., Dewey, M., Fons, E., and et al.: ClimateBench: A benchmark dataset for data-driven climate projections, Earth Space Sci. Open Archive, p. 33, https://doi.org/10.1002/essoar.10509765.2, 2021.

Werbos, P.: Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Science. Thesis (Ph. D.). Appl. Math. Harvard University, 1974.

Westra, S., Brown, C., Lall, U., and Sharma, A.: Modeling multivariable hydrological series: Principal component analysis or independent component analysis?, Weather Clim. Extrem., 43, https://doi.org/10.1029/2006WR005617, 2007.

Weyn, J. A., Durran, D. R., and Caruana, R.: Improving Data-Driven Global Weather Prediction Using Deep Convolutional Neural Networks on a Cubed Sphere, J. Adv. Model. Earth Syst., 12, e2020MS002 109, https://doi.org/10.1029/2020MS002109, 2020.

Wheeler, T. and von Braun, J.: Climate Change Impacts on Global Food Security, Science, 341, 508–513, https://doi.org/10.1126/science.1239402, 2013.

Wilby, R. L., Hay, L. E., Gutowski Jr., W. J., Arritt, R. W., Takle, E. S., Pan, Z., Leavesley, G. H., and Clark, M. P.: Hydrological responses to dynamically and statistically downscaled climate model output, Geophys. Res. Lett., 27, 1199–1202, https://doi.org/10.1029/1999GL006078, 2000.

Wilcke, R. A. I., Mendlik, T., and Gobiet, A.: Multi-variable error correction of regional climate models, Clim. Change, 120, 871–887, https://doi.org/10.1007/s10584-013-0845-x, 2013.

Wilks, D. S.: Statistical Methods in the Atmosphere Science, Academic Press, 2006.

Wilks, D. S. and Wilby, R. L.: The weather generation game: a review of stochastic weather models, Prog. Phys. Geogr., 23, 329–357, https://doi.org/10.1177/030913339902300302, 1999.

Wu, X., Hao, Z., Tang, Q., Singh, V. P., Zhang, X., and Hao, F.: Projected increase in compound dry and hot events over global land areas, Int. J. Climatol., 41, 393–403, https://doi.org/10.1002/joc.6626, 2021a.

Wu, X., Hao, Z., Hao, F., Zhang, X., Singh, V. P., and Sun, C.: Influence of Large-Scale Circulation Patterns on Compound Dry and Hot Events in China, J. Geophys. Res. Atmos., 126, e2020JD033 918, https://doi.org/10.1029/2020JD033918, 2021b.

Xu, C.-Y.: From GCMs to river flow: A review of downscaling methods and hydrologic modelling approaches, Prog. Phys. Geog., 23, 229–249, https://doi.org/10.1177/030913339902300204, 1999.

Xue, Q., Hu, Y., and Tompkins, W.: Neural-network-based adaptive matched filtering for QRS detection, IEEE Trans. Biomed. Eng., 39, 317–329, https://doi.org/10.1109/10.126604, 1992.

Yang, W., Gardelin, M., Olsson, J., and Bosshard, T.: Multi-variable bias correction: Application of forest fire risk in present and future climate in Sweden, Nat. Hazards Earth Syst. Sci., 15, https://doi.org/10.5194/nhess-15-2037-2015, 2015.

Yang, Z. and Villarini, G.: Examining the capability of reanalyses in capturing the temporal clustering of heavy precipitation across Europe, Clim. Dyn., 53, 1845–1857, https://doi.org/10.1007/s00382-019-04742-z, 2019.

Yang, Z. and Villarini, G.: Evaluation of the capability of global climate models in reproducing the temporal clustering in heavy precipitation over Europe, Int. J. Climatol., 41, 131–145, https://doi.org/10.1002/joc.6612, 2021.

Yiou, P., Cattiaux, J., Ribes, A., Vautard, R., and Vrac, M.: Recent Trends in the Recurrence of North Atlantic Atmospheric Circulation Patterns, Complexity, 2018, 3140 915, https://doi.org/10.1155/2018/3140915, 2018.

Yiou, P., Cattiaux, J., Faranda, D., Kadygrov, N., Jézéquel, A., Naveau, P., Ribes, A., Robin, Y., Thao, S., van Oldenborgh, G. J., and Vrac, M.: Analyses of the Northern European Summer Heatwave of 2018, Bull. Am. Meteorol. Soc., 101, S35 – S40, https://doi.org/10.1175/BAMS-D-19-0170.1, 2020.

Yoo, D., Kim, N., Park, S., Paek, A. S., and Kweon, I. S.: Pixel-Level Domain Transfer, 2016.

Zaytar, M. A. and Amrani, C. E.: Sequence to Sequence Weather Forecasting with Long Short-Term Memory Recurrent Neural Networks, Int. J. Comput. Appl., 143, 7–11, https://doi.org/10.5120/ijca2016910497, 2016.

Zhang, J., Liu, P., Zhang, F., and Song, Q.: CloudNet: Ground-Based Cloud Classification With Deep Convolutional Neural Network, Geophys. Res. Lett., 45, 8665–8672, https://doi.org/10.1029/2018GL077787, 2018.

Zhao, W. and Du, S.: Learning multiscale and deep representations for classifying remotely sensed imagery, ISPRS J. Photogramm. Remote Sens., 113, 155–165, https://doi.org/10.1016/j.isprsjprs.2016.01.004, 2016.

Zhao, Y. and Chen, C.: Unpaired Image-to-Image Translation via Latent Energy Transport, https://doi.org/10.48550/ARXIV.2012.00649, 2020.

Zheng, F., Westra, S., Leonard, M., and Sisson, S. A.: Modeling dependence between extreme rainfall and storm surge to estimate coastal flooding risk, Water Resour. Res., 50, 2050–2071, https://doi.org/10.1002/2013WR014616, 2014.

Zhou, P. and Liu, Z.: Likelihood of concurrent climate extremes and variations over China, Environ. Res. Lett., 13, 094 023, https://doi.org/10.1088/1748-9326/aade9e, 2018.

Zhu, J.-Y., Park, T., Isola, P., and Efros, A. A.: Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks, 2017.

Zintgraf, L. M., Cohen, T. S., Adel, T., and Welling, M.: Visualizing Deep Neural Network Decisions: Prediction Difference Analysis, https://doi.org/10.48550/ARXIV.1702.04595, 2017.

Zorita, E. and von Storch, H.: The Analog Method as a Simple Statistical Downscaling Technique: Comparison with More Complicated Methods, J. Clim., year =.

Zscheischler, J. and Lehner, F.: Attributing compound events to anthropogenic climate change, Bull. Amer. Meteor. Soc., pp. 1–45, https://doi.org/10.1175/BAMS-D-21-0116.1, 2021.

Zscheischler, J. and Seneviratne, S.: Dependence of drivers affects risks associated with compound events, Sci. Adv, 3, e1700 263, https://doi.org/10.1126/sciadv.1700263, 2017.

Zscheischler, J., Michalak, A. M., Schwalm, C., Mahecha, M. D., Huntzinger, D. N., Reichstein, M., Berthier, G., Ciais, P., Cook, R. B., El-Masri, B., Huang, M., Ito, A., Jain, A., King, A., Lei, H., Lu, C., Mao, J., Peng, S., Poulter, B., Ricciuto, D., Shi, X., Tao, B., Tian, H., Viovy, N., Wang, W., Wei, Y., Yang, J., and Zeng, N.: Impact of large-scale climate extremes on biospheric carbon fluxes: An intercomparison based on MsTMIP data, Glob. Biogeochem. Cycles, 28, 585–600, https://doi.org/10.1002/2014GB004826, 2014.

Zscheischler, J., Westra, S., Hurk, B., Seneviratne, S., Ward, P., Pitman, A., AghaKouchak, A., Bresch, D., Leonard, M., Wahl, T., and Zhang, X.: Future climate risk from compound events, Nat. Clim. Chang., pp. 469–477, https://doi.org/10.1038/s41558-018-0156-3, 2018.

Zscheischler, J., Fischer, E., and Lange, S.: The effect of univariate bias adjustment on multivariate hazard estimates, Earth Syst. Dynam., 10, 31–43, https://doi.org/10.5194/esd-10-31-2019, 2019.

Zscheischler, J., Martius, O., Westra, S., Bevacqua, E., Raymond, C., Horton, R., Hurk, B., AghaKouchak, A., Jézéquel, A., Mahecha, M., Maraun, D., Ramos, A., Ridder, N., Thiery, W., and Vignotto, E.: A typology of compound weather and climate events, Nat. Rev. Earth Environ., 1, 1–5, 2020.

Zscheischler, J., Naveau, P., Martius, O., Engelke, S., and Raible, C.: Evaluating the dependence structure of compound precipitation and wind speed extremes, https://doi.org/10.5194/esd-2020-31, 2021.

Appendix

Appendix A

Supplement of the article "Multivariate bias corrections of climate simulations: which benefits for which losses?"

Appendix A. Supplement of the article "Multivariate bias corrections of climate simulations: which benefits for which losses?"

Supplement of Earth Syst. Dynam., 11, 537–562, 2020
https://doi.org/10.5194/esd-11-537-2020-supplement

Earth System Dynamics
Open Access
EGU

Supplement of

Multivariate bias corrections of climate simulations: which benefits for which losses?

Bastien François et al.

Correspondence to: Bastien François (bastien.francois@lsce.ipsl.fr)

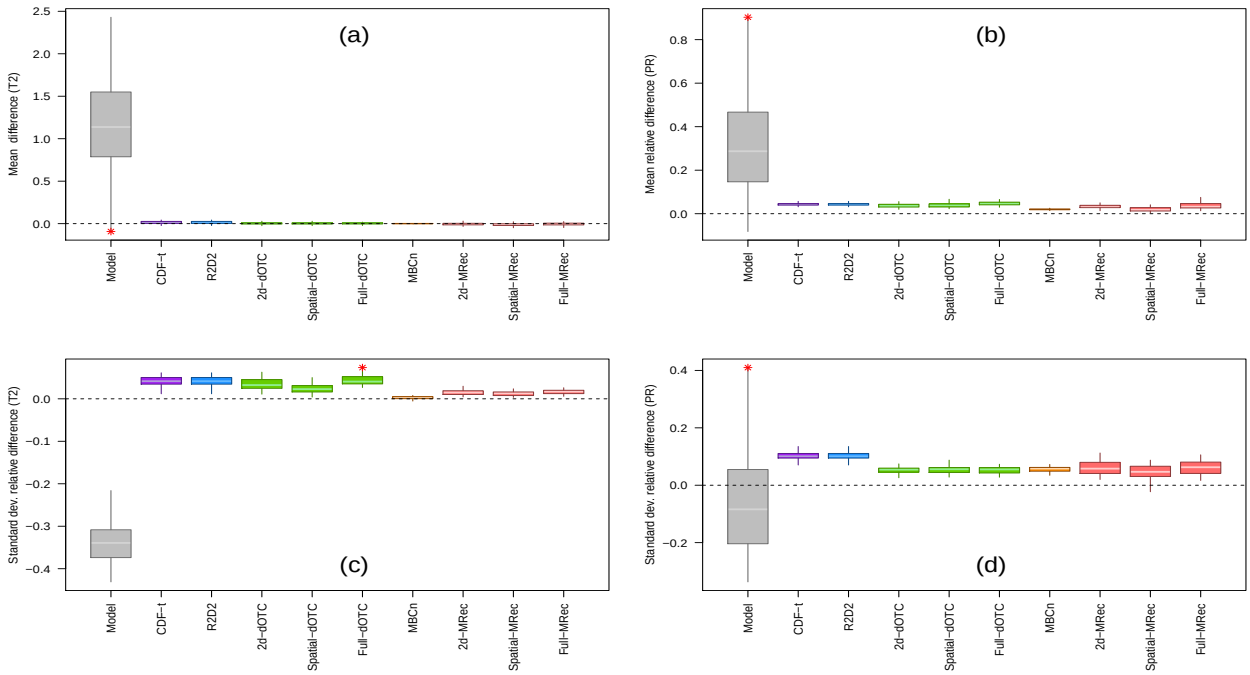

Figure S1. Boxplots of mean **(a and b)** and standard deviation **(c and d)** differences for Temperature (T2, a and c) and Precipitation (PR, b and d) during winter over the 1979-2016 period for the Brittany region (SAFRAN reference). Results are shown for: plain IPSL; CDF-t; R^2D^2; dOTC (2d-, Spatial- and Full-versions); MBC-n and MRec (2d-, Spatial- and Full-versions) outputs. Red asterisks indicate values lying outside the plotted range.

Appendix A. Supplement of the article "Multivariate bias corrections of climate simulations: which benefits for which losses?"

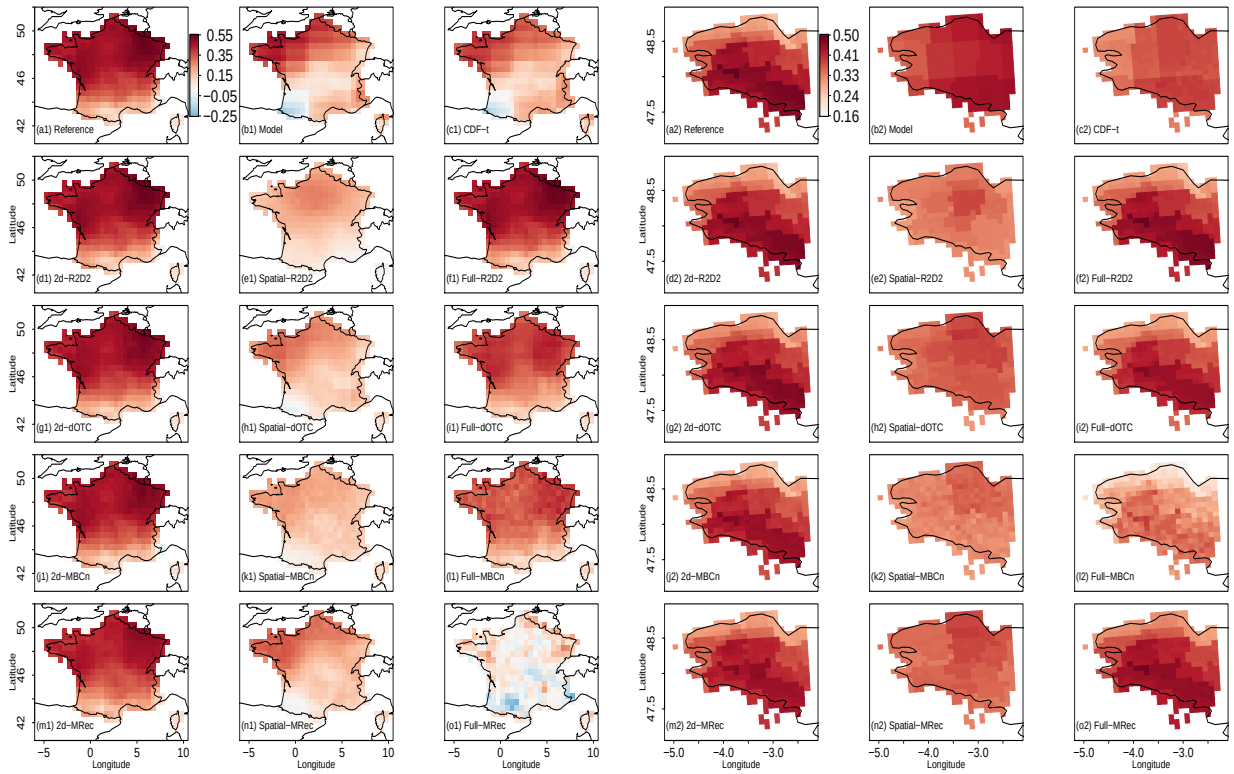

Figure S2. Maps of temperature vs. precipitation Spearman correlation maps computed at each grid cell using WFDEI reference **(a1-o1)** and SAFRAN reference **(a2-o2)** during winter over the 1979-2016 period. Results are shown for: Reference; plain IPSL; CDF-t; R^2D^2; dOTC; MBC-n and MRec outputs for respectively 2d-, Spatial- and Full- versions. Note that the color scales between **(a1-o1)** and **(a2-o2)** are not the same to better emphasize intensities of values of the two regions.

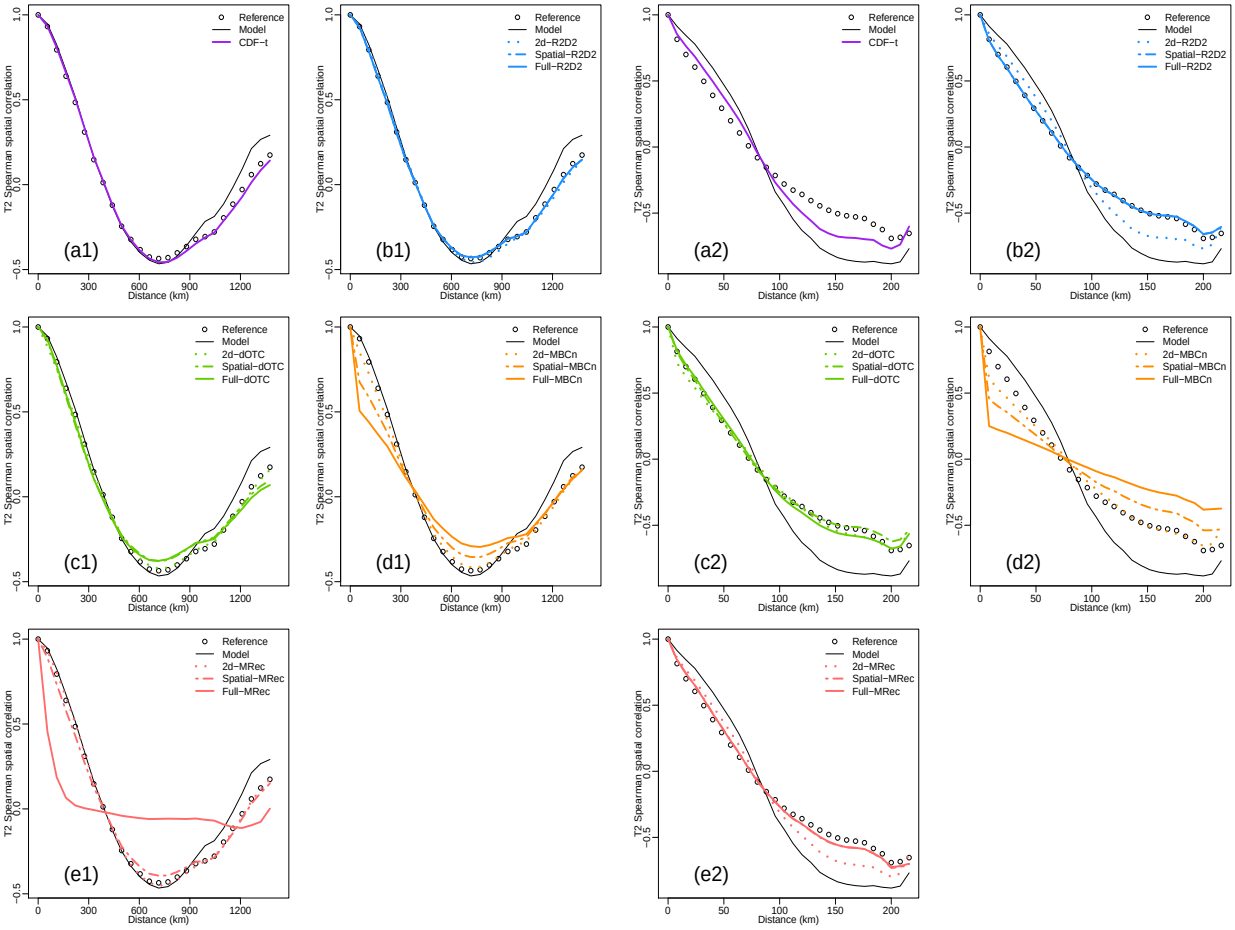

Figure S3. Correlograms for temperature using WFDEI reference **(a1-e1)** and SAFRAN reference **(a2-e2)** during winter over the 1979-2016 period. Results are shown for Reference (circles) and plain IPSL (black line). Results are displayed for: CDF-t; R^2D^2; dOTC; MBC-n and MRec outputs for respectively 2d- (dotted), Spatial- (dashed) and Full-versions (solid lines).

Appendix A. Supplement of the article "Multivariate bias corrections of climate simulations: which benefits for which losses?"

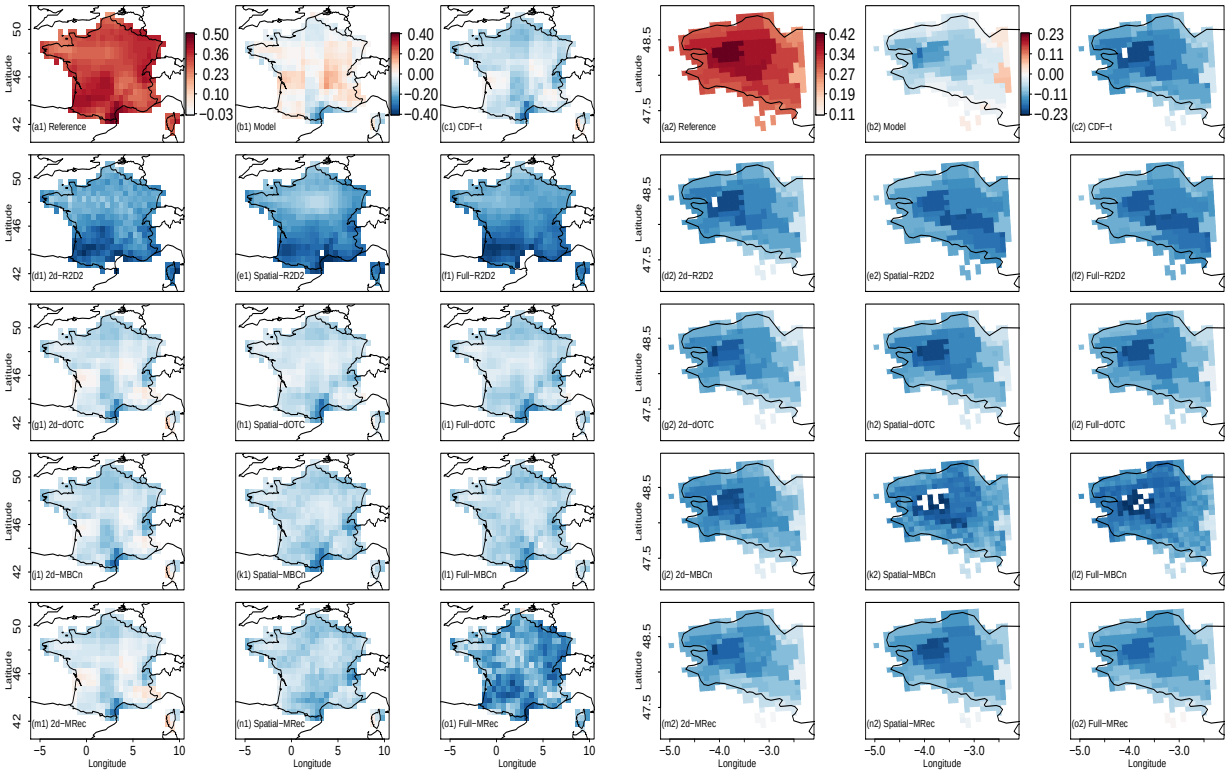

Figure S4. Differences of order 1 autocorrelation for precipitation using WFDEI reference **(a1-o1)** and SAFRAN reference **(a2-o2)** during winter over the 1979-2016 period. Results are shown for: Reference; plain IPSL; CDF-t; R^2D^2; dOTC; MBC-n and MRec outputs for respectively 2d-, Spatial- and Full- versions. Note that the color scales between **(a1-o1)** and **(a2-o2)** are not the same to better emphasize intensities of values of the two regions.

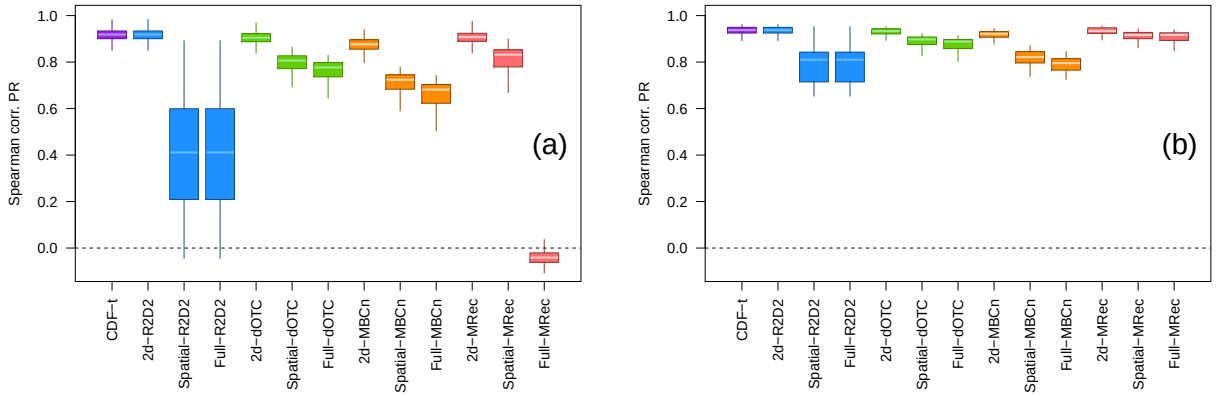

Figure S5. Boxplots of rank correlation computed at each grid cell between the bias corrected and the raw climate model time series for precipitation using WFDEI for France **(a)** and SAFRAN for Brittany **(b)** region during winter over the 1979-2016 period. For both boxplots, results are shown for: CDF-t; R^2D^2; dOTC; MBC-n and MRec outputs for 2d-, Spatial- and Full-versions.

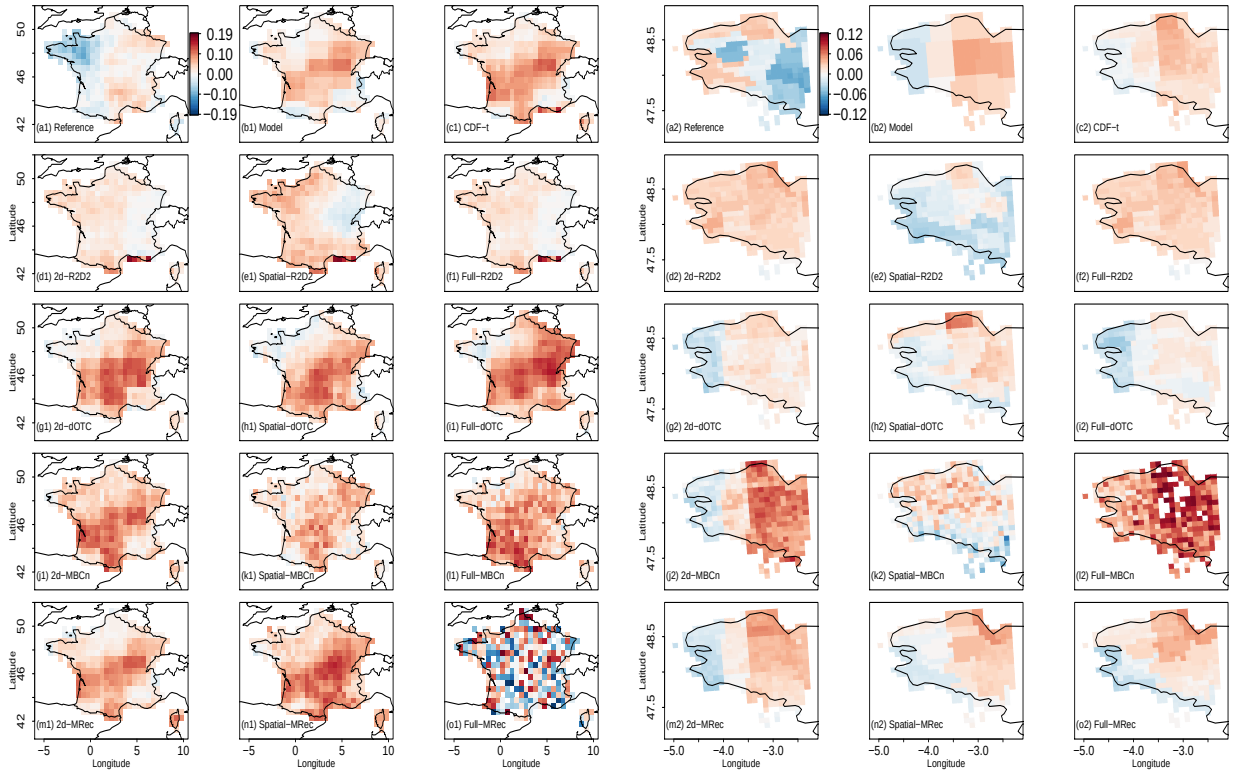

Figure S6. Differences of temperature vs. precipitation Spearman correlation computed at each grid cell between the 1979-1997 and 1998-2016 periods during summer. WFDEI **(a1-o1)** and SAFRAN **(a2-o2)** data are used for the bias correction. Note that the color scales between **(a1-o1)** and **(a2-o2)** are not the same to better emphasize intensities of values of the two regions.

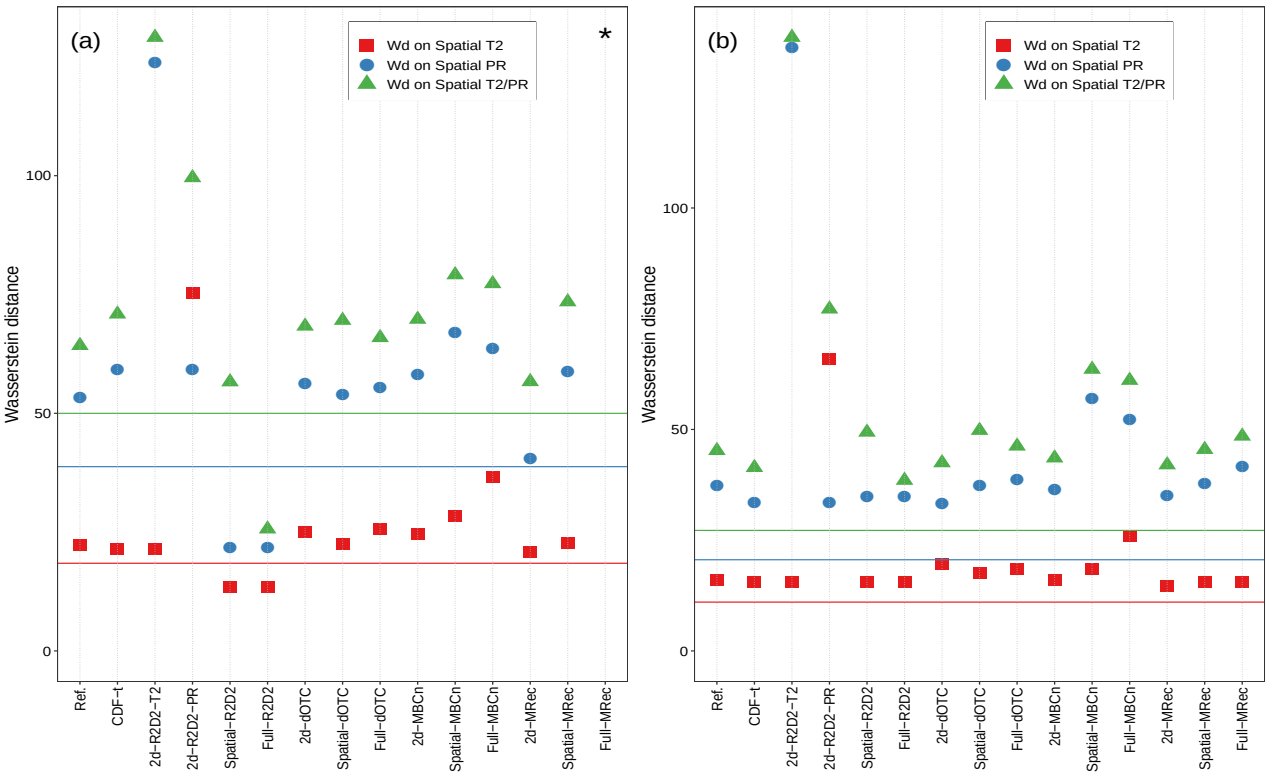

Figure S7. Values of the three Wasserstein distances between 1979-1997 and 1998-2016 periods during winter for temperature (square), precipitation (circle) and both temperature and precipitation (triangle) for the region of France **(a)** and Brittany **(b)**. Results are presented for: Reference; plain IPSL (lines); CDF-t and the different multivariate BC outputs. 2d-R^2D^2-T2 (resp. 2d-R^2D^2-PR) indicates results for 2d-R^2D^2 with temperature (resp. precipitation) used as reference dimension. Black asterisks indicate values lying outside the plotted range.

Appendix A. Supplement of the article "Multivariate bias corrections of climate simulations: which benefits for which losses?"

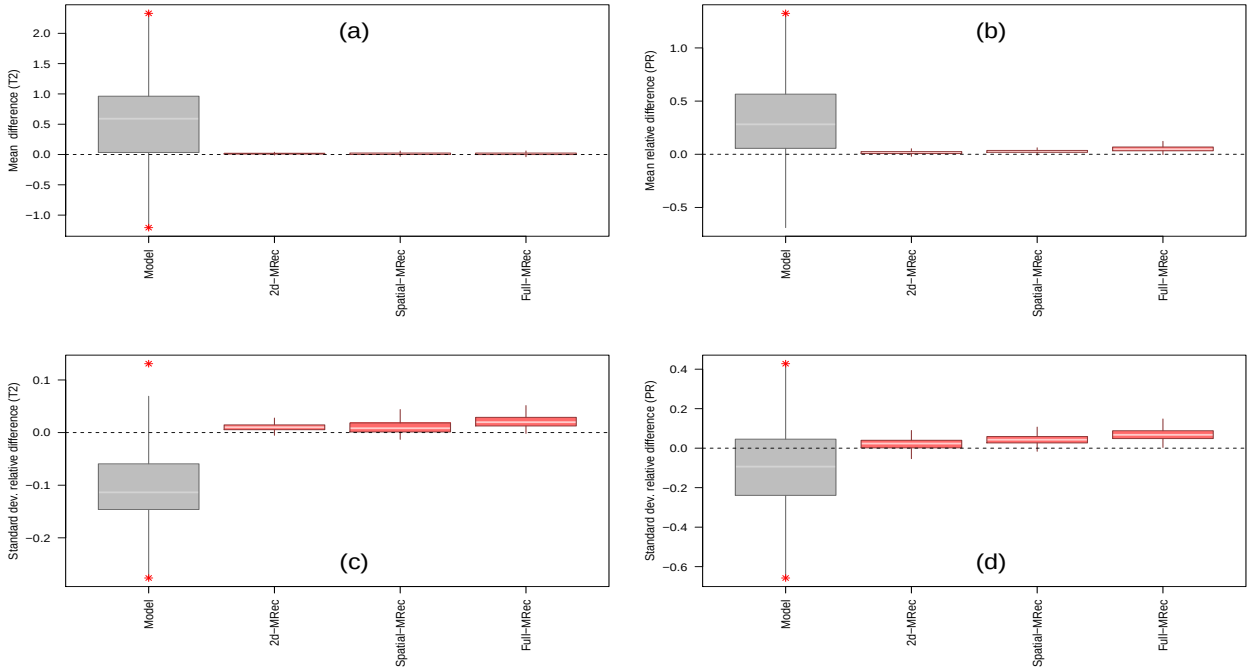

Figure S8. Boxplots of mean **(a and b)** and standard deviation **(c and d)** differences for Temperature (T2, a and c) and Precipitation (PR, b and d) during winter over the 1979-2016 period for the France region (WFDEI reference) with seasonal BC. Results are shown for: plain IPSL and MRec (2d-, Spatial- and Full-versions) outputs. Red asterisks indicate values lying outside the plotted range.

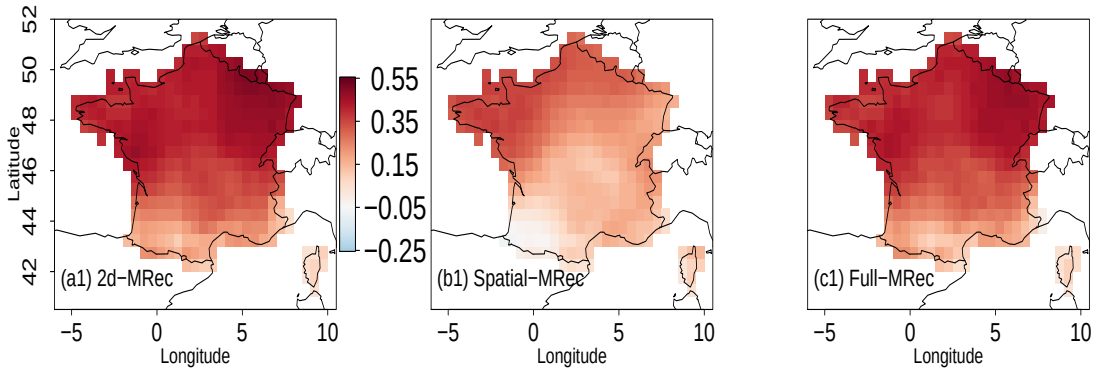

Figure S9. Temperature vs. precipitation Spearman correlation maps computed at each grid cell using WFDEI reference **(a1-c1)** during winter over the 1979-2016 period with seasonal BC. Results are shown for MRec **(a1-c1)** outputs for respectively 2d-, Spatial- and Full-versions.

Appendix A. Supplement of the article "Multivariate bias corrections of climate simulations: which benefits for which losses?"

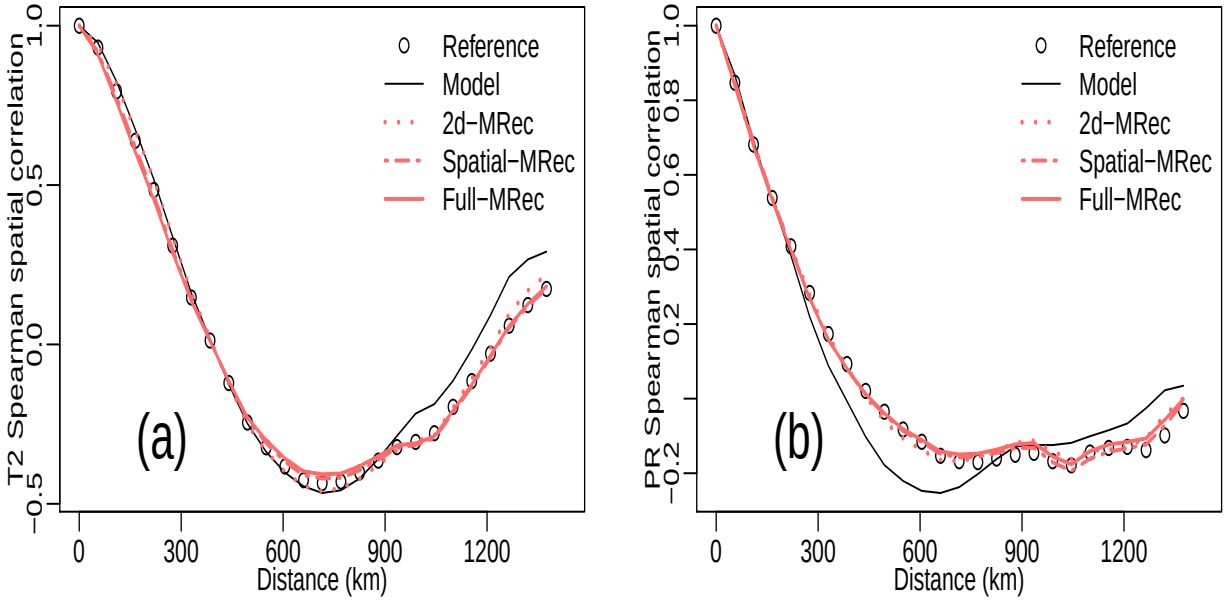

Figure S10. Correlograms for temperature **(a)** and precipitation **(b)** using WFDEI reference for France during winter over the 1979-2016 period with seasonal BC. Results are shown for: Reference (circles), plain IPSL (black line) and MRec outputs for respectively 2d- (dotted), Spatial- (dashed) and Full-versions (solid lines).

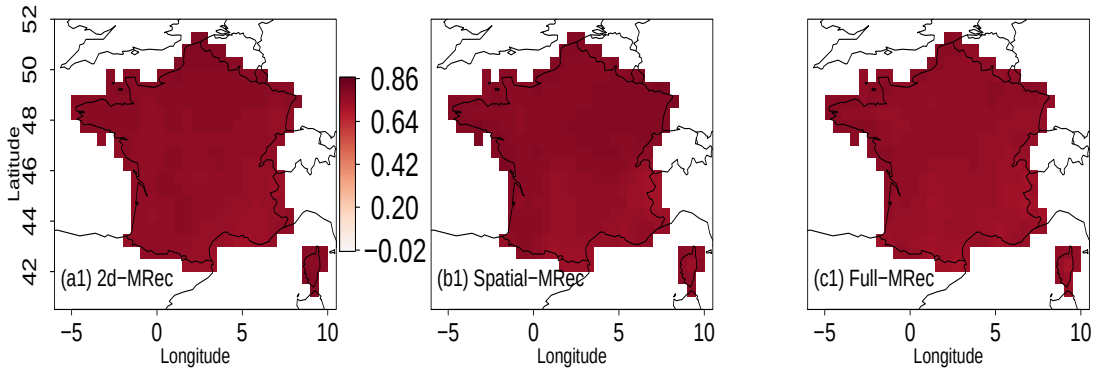

Figure S11. Order 1 Pearson autocorrelation for temperature using WFDEI reference during winter over the 1979-2016 period for the seasonal BC. Results are shown for MRec outputs for 2d-, Spatial- and Full-versions.

Appendix A. Supplement of the article "Multivariate bias corrections of climate simulations: which benefits for which losses?"

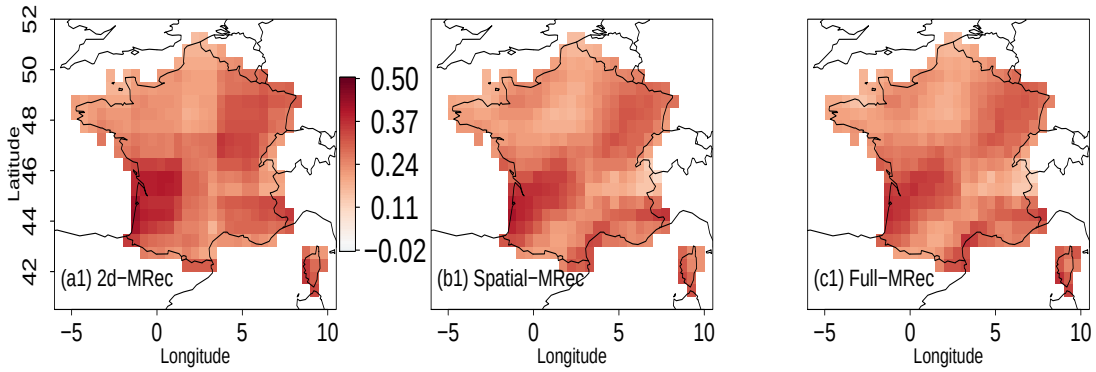

Figure S12. Order 1 Pearson autocorrelation for precipitation using WFDEI reference during winter over the 1979-2016 period for the seasonal BC. Results are shown for MRec outputs for 2d-, Spatial- and Full-versions.

Appendix B

Supplement of the article "Adjusting spatial dependence of climate model outputs with cycle-consistent adversarial networks"

Appendix B. Supplement of the article "Adjusting spatial dependence of climate model outputs with cycle-consistent adversarial networks"

Climate Dynamics manuscript No.
(will be inserted by the editor)

Adjusting spatial dependence of climate model outputs with Cycle-Consistent Adversarial Networks

Bastien François · Soulivanh Thao · Mathieu Vrac

Received: date / Accepted: date

Bastien François
Laboratoire des Sciences du Climat et l'Environnement (LSCE-IPSL) CNRS/CEA/UVSQ, UMR8212, Université Paris-Saclay, Gif-sur-Yvette, France
Tel.: +33-169086000
Fax: +33-169087716
E-mail: bastien.francois@lsce.ipsl.fr

Soulivanh Thao
Laboratoire des Sciences du Climat et l'Environnement (LSCE-IPSL) CNRS/CEA/UVSQ, UMR8212, Université Paris-Saclay, Gif-sur-Yvette, France

Mathieu Vrac
Laboratoire des Sciences du Climat et l'Environnement (LSCE-IPSL) CNRS/CEA/UVSQ, UMR8212, Université Paris-Saclay, Gif-sur-Yvette, France

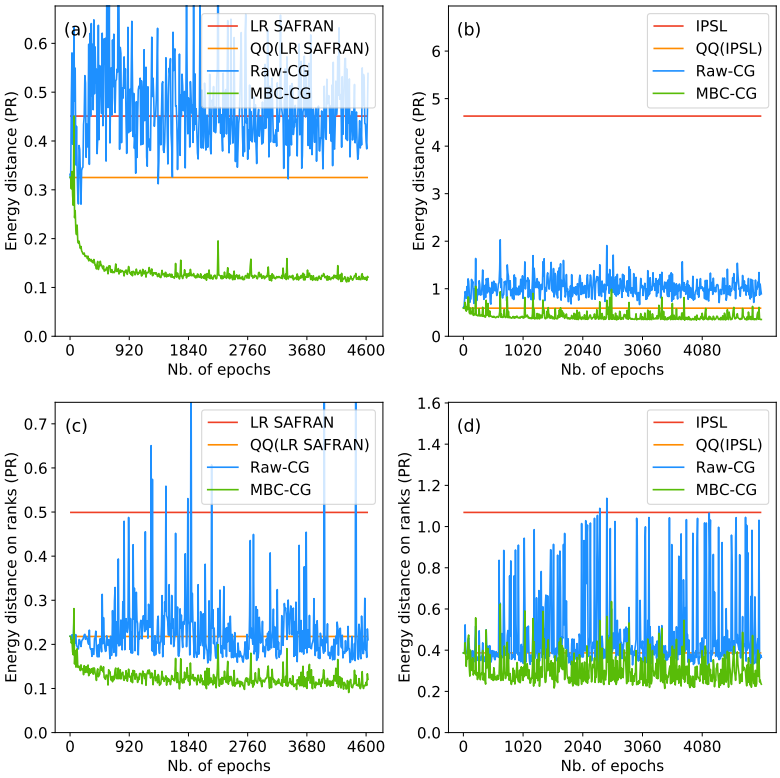

Fig. S1: Values of the energy distances with respect to SAFRAN reference for precipitation computed on **(a-b)** physical values and **(c-d)** ranks during the training of MBC-CycleGAN. Results are shown for the different datasets involved in **(a, c)** the Perfect Prognosis approach and **(b, d)** the MOS approach.

Appendix B. Supplement of the article "Adjusting spatial dependence of climate model outputs with cycle-consistent adversarial networks"

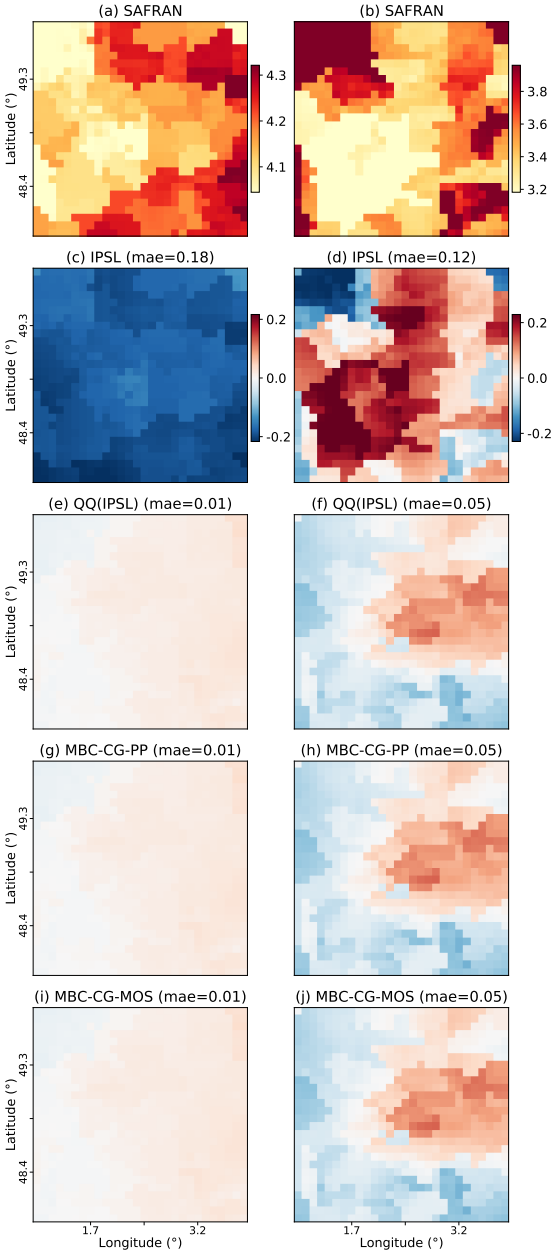

Fig. S2: Standard deviation relative differences for **(c, e, g, i)** temperature and **(d, f, h, j)** precipitation computed at each grid cell between SAFRAN reference and the different datasets (plain IPSL, QQ, MBC-CycleGAN-PP and MBC-CycleGAN-MOS outputs) during winter over the projection period. Note that the color scales between panels **(c, e, g, i)** and **(d, f, h, j)** are not the same to better emphasize intensities of values for the two physical variables. Maps of daily standard deviation for SAFRAN references are also shown for **(a)** temperature and **(b)** precipitation.

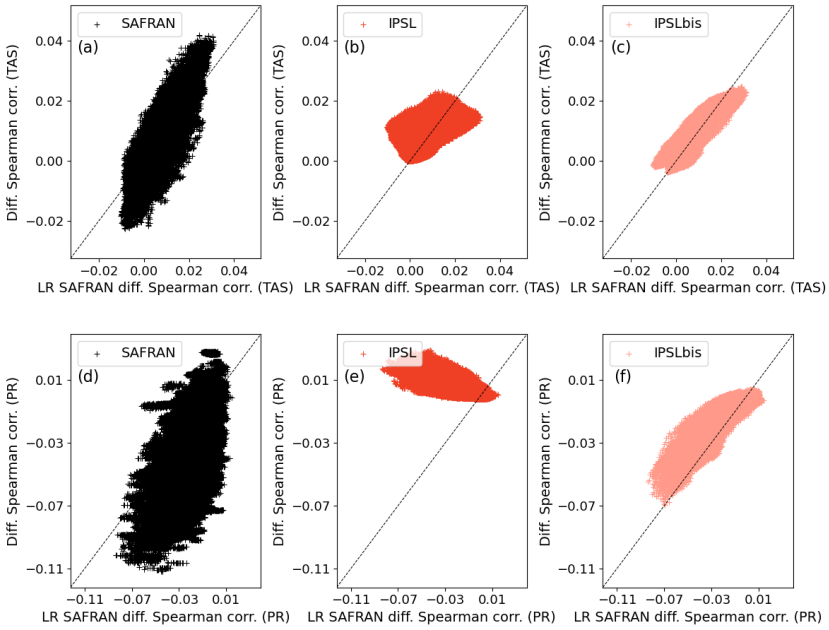

Fig. S3: Scatterplot showing winter differences of Spearman spatial correlations for **(a-c)** temperature and **(d-f)** precipitation between the calibration and the projection period. Results are shown for SAFRAN, IPSL and IPSLbis outputs that are compared with LR SAFRAN differences.

Appendix B. Supplement of the article "Adjusting spatial dependence of climate model outputs with cycle-consistent adversarial networks"

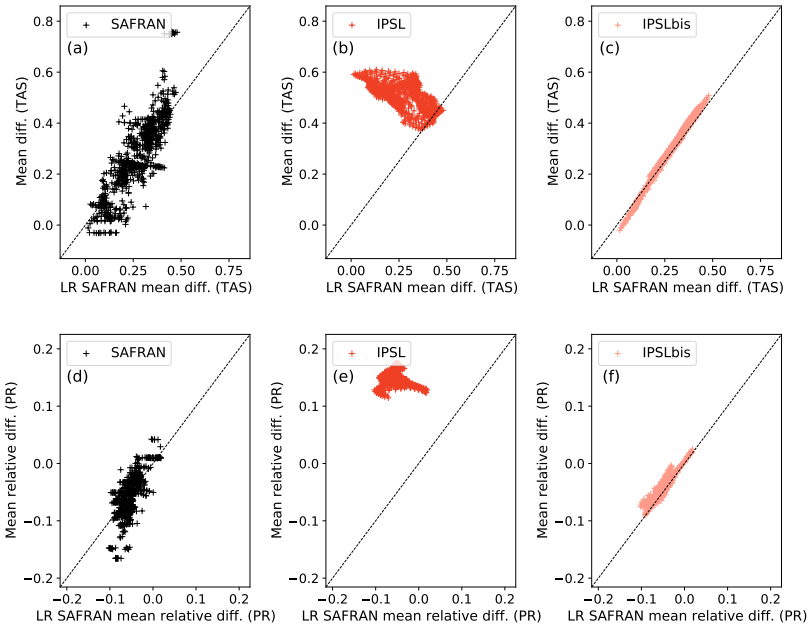

Fig. S4: Scatterplot showing winter mean differences for **(a-c)** temperature and **(d-f)** precipitation between the calibration and the projection period. Results are shown for SAFRAN, IPSL and IPSLbis outputs that are compared with LR SAFRAN differences.

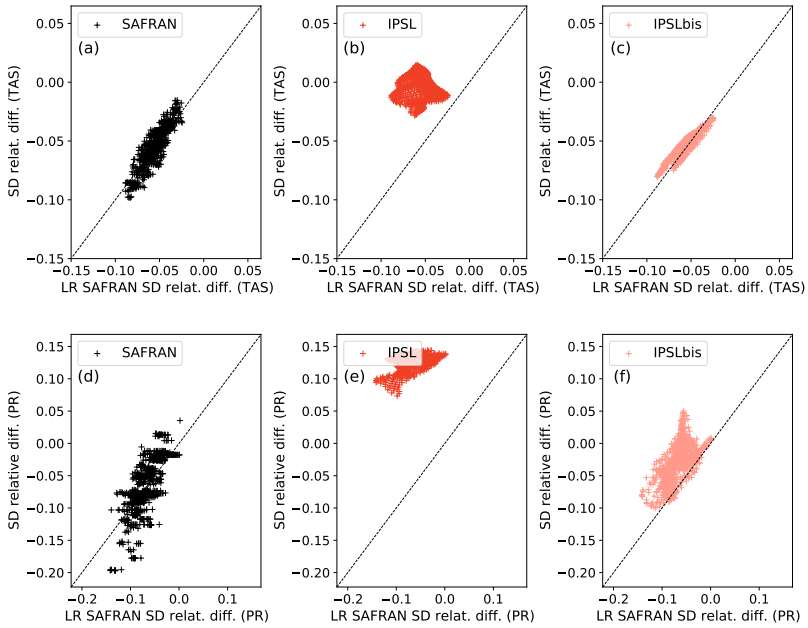

Fig. S5: Scatterplot showing winter relative differences of standard deviations for **(a-c)** temperature and **(d-f)** precipitation between the calibration and the projection period. Results are shown for SAFRAN, IPSL and IPSLbis outputs that are compared with LR SAFRAN relative differences.

Appendix B. Supplement of the article "Adjusting spatial dependence of climate model outputs with cycle-consistent adversarial networks"

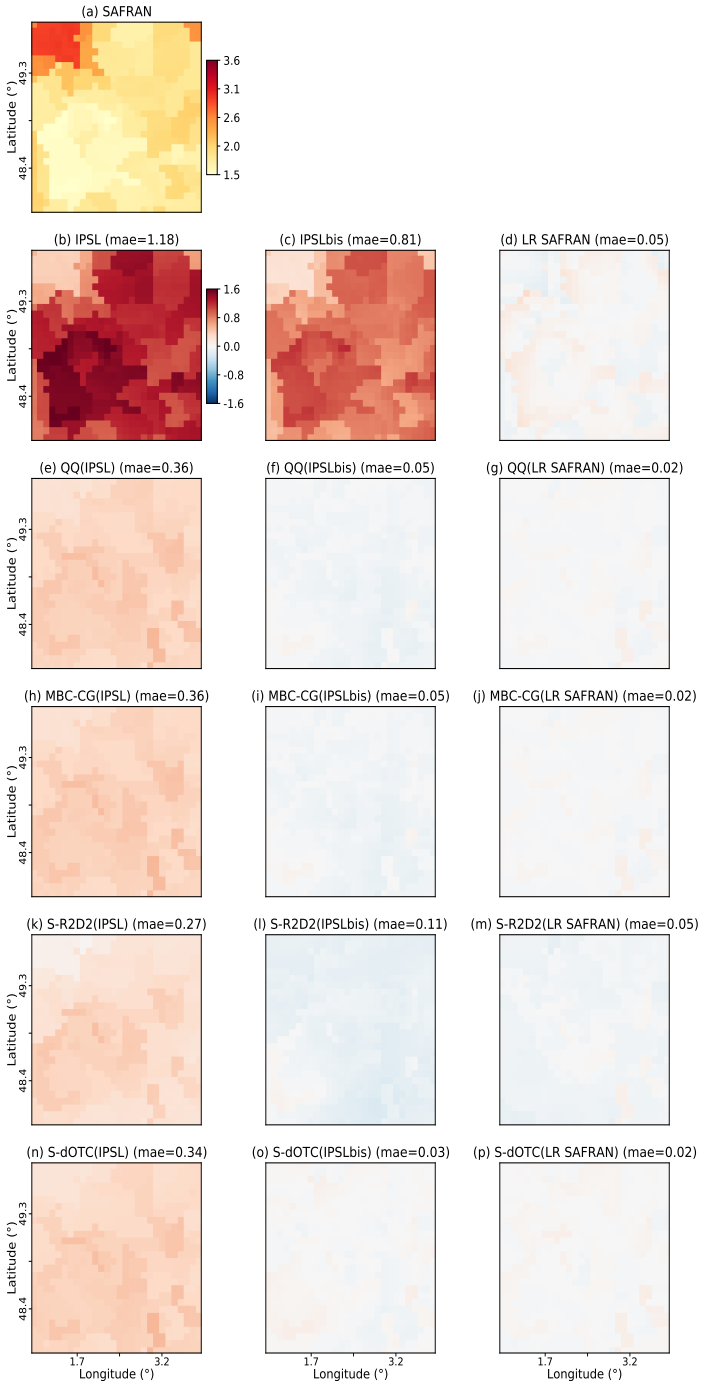

Fig. S6: Relative mean differences for precipitation with SAFRAN reference for BC methods using as inputs **(b, e, h, k, n)** IPSL, **(c, f, i, l, o)** IPSLbis and **(d, g, j, m, p)** LR SAFRAN data. Results are shown during winter over the projection period for IPSL, IPSLbis, LR SAFRAN, QQ, MBC-CycleGAN, Spatial-R^2D^2 and Spatial-dOTC datasets. The map of daily mean for SAFRAN references is also shown for precipitation **(a)**.

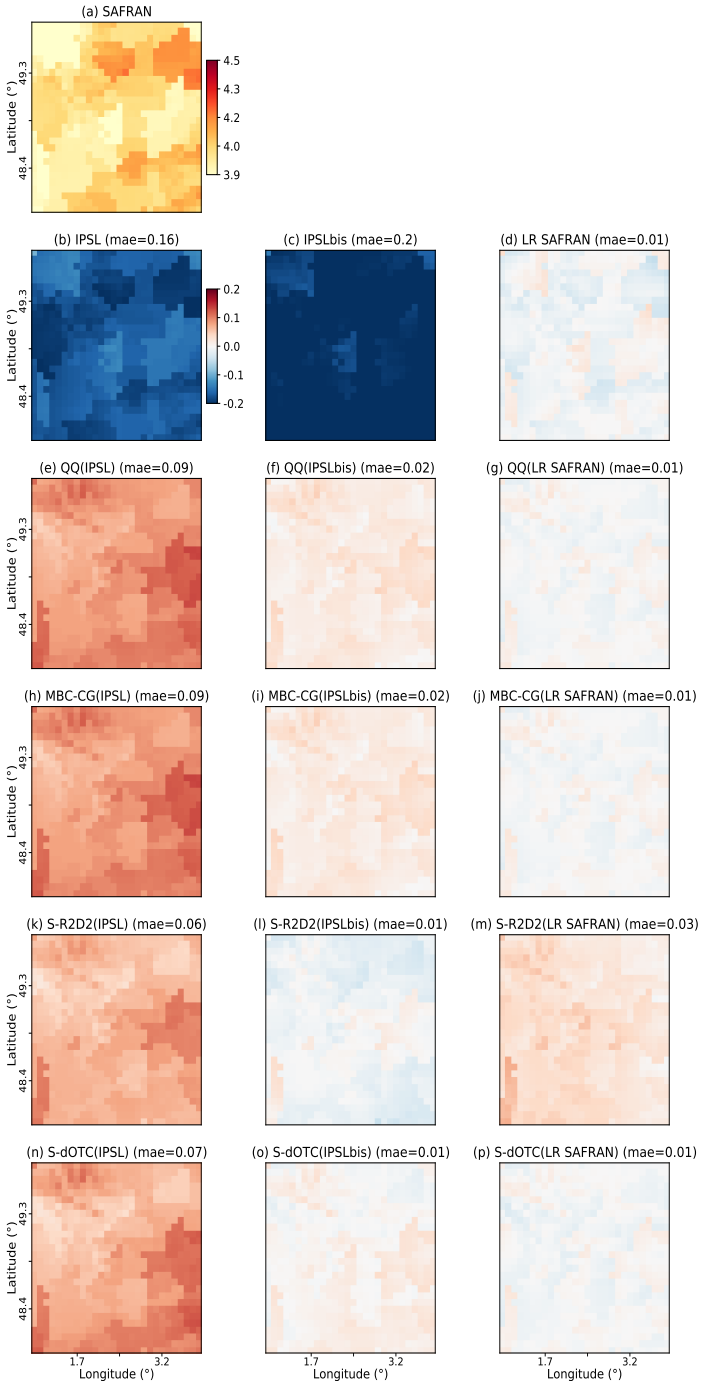

Fig. S7: Standard deviation relative differences for temperature with SAFRAN reference for BC methods using as inputs **(b, e, h, k, n)** IPSL, **(c, f, i, l, o)** IPSLbis and **(d, g, j, m, p)** LR SAFRAN data. Results are shown during winter over the projection period IPSL, IPSLbis, LR SAFRAN, QQ, MBC-CycleGAN, Spatial-R^2D^2 and Spatial-dOTC datasets. The map of daily standard deviation for SAFRAN references is also shown for temperature **(a)**.

Appendix B. Supplement of the article "Adjusting spatial dependence of climate model outputs with cycle-consistent adversarial networks"

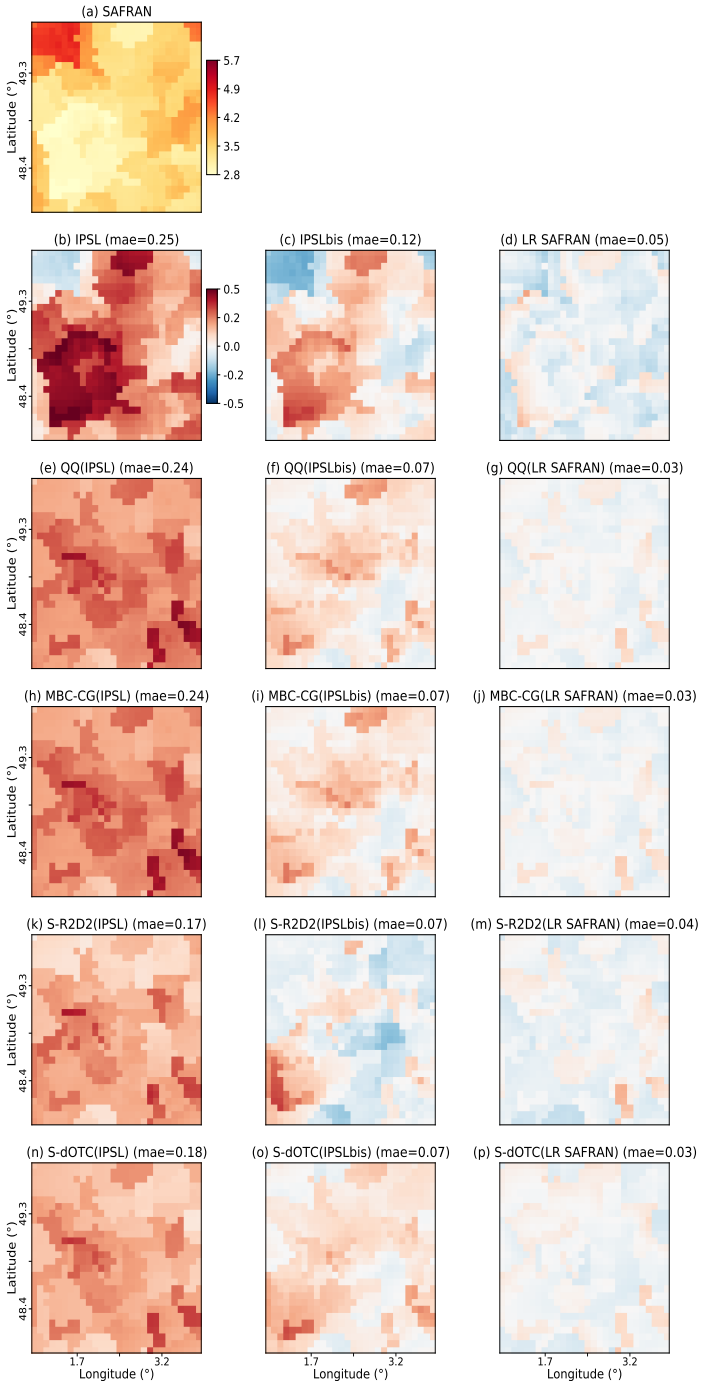

Fig. S8: Standard deviation relative differences for precipitation with SAFRAN reference for BC methods using as inputs **(b, e, h, k, n)** IPSL, **(c, f, i, l, o)** IPSLbis and **(d, g, j, m, p)** LR SAFRAN data. Results are shown during winter over the projection period for IPSL, IPSLbis, LR SAFRAN, QQ, MBC-CycleGAN, Spatial-R^2D^2 and Spatial-dOTC datasets. The map of daily standard deviation for SAFRAN references is also shown for precipitation **(a)**.

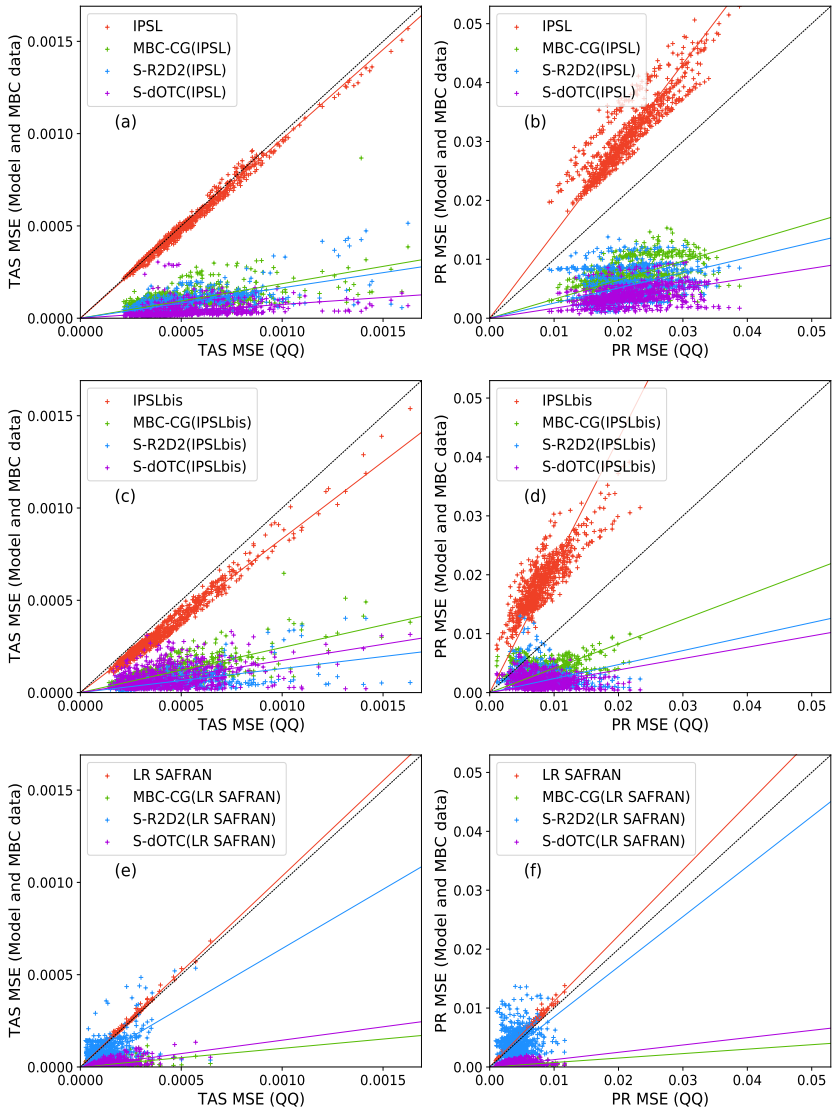

Fig. S9: Scatterplots of mean squared errors of Pearson spatial correlations computed at each grid cell for **(a, c, e)** temperature and **(b, d, f)** precipitation over the projection period. Results are shown for IPSL, IPSLbis, LR SAFRAN, MBC-CycleGAN, Spatial-R^{2}D^2 and Spatial-dOTC datasets and compared with QQ outputs.

Appendix B. Supplement of the article "Adjusting spatial dependence of climate model outputs with cycle-consistent adversarial networks"

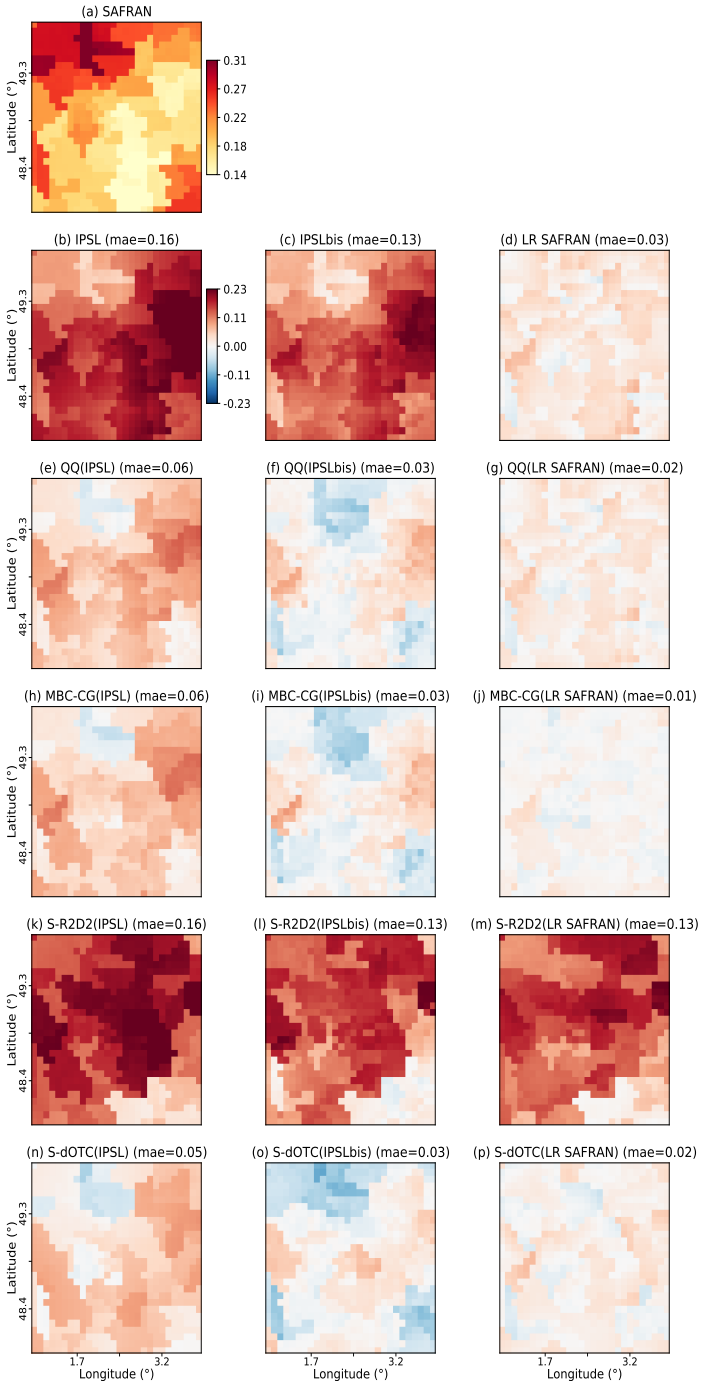

Fig. S10: Differences of order 1 Pearson autocorrelation for precipitation with SAFRAN reference for BC methods using as inputs **(b, e, h, k, n)** IPSL, **(c, f, i, l, o)** IPSLbis and **(d, g, j, m, p)** LR SAFRAN data. Results are shown during winter over the projection period for reference, IPSL, IPSLbis, LR SAFRAN, QQ, MBC-CycleGAN, Spatial-R^2D^2 and Spatial-dOTC datasets. The map of order 1 Pearson autocorrelation for SAFRAN references is also shown for precipitation **(a)**.

Appendix C

Supplement of the submitted article "Time of Emergence of compound events: contribution of univariate and dependence properties"

Time of Emergence of compound events: contribution of univariate and dependence properties

Bastien François[1] and Mathieu Vrac[1]

[1]Laboratoire des Sciences du Climat et l'Environnement (LSCE-IPSL) CNRS/CEA/UVSQ, UMR8212, Université Paris-Saclay, Gif-sur-Yvette, France

Correspondence: B. François (bastien.francois@lsce.ipsl.fr)

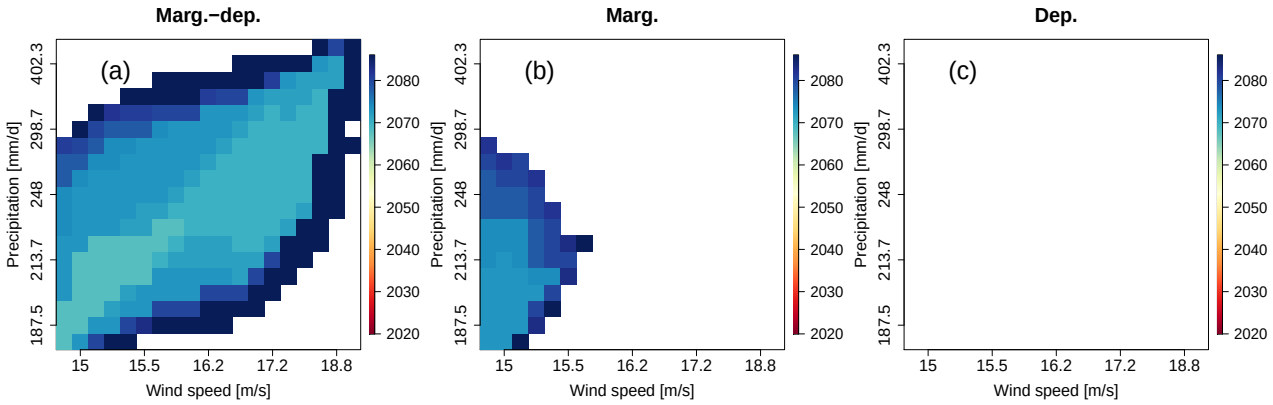

Figure S1. Same as Fig. 6 but for 95% confidence level: CNRM-CM6 (a-c) time of emergence at 95% confidence level for compound wind and precipitation extremes due to changes of (a) both marginal and dependence properties, (b) marginal properties only, and (c) dependence properties only. Results are presented for varying exceedance thresholds between the 5th and 95th percentile of compound wind and precipitation extremes data. White indicates that no time of emergence is detected.

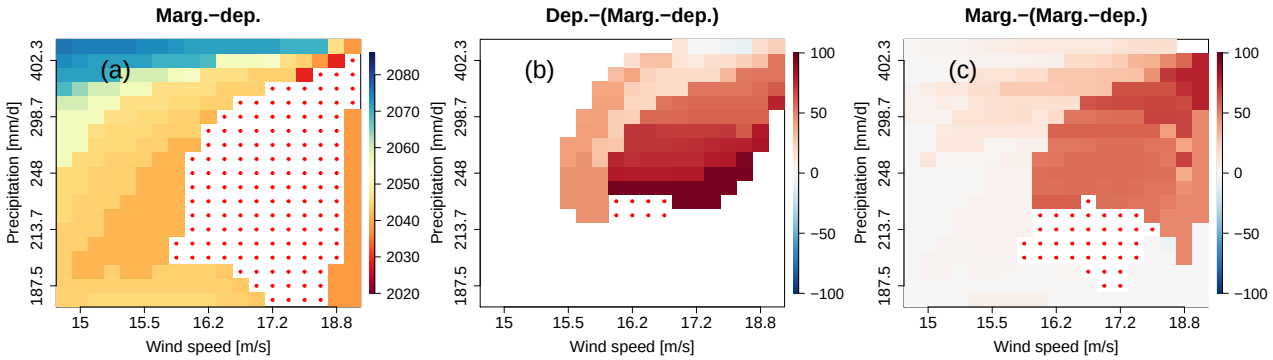

Figure S2. CNRM-CM6 differences of time of emergence at 68 % confidence level between time of emergence obtained by considering both marginal and dependence properties changes and (b) dependence properties changes only, and (c) marginal properties only. CNRM-CM6 (a) time of emergence at 68 % confidence level obtained by considering both marginal and dependence properties changes are also displayed. Color points indicate values lying outside the plotted ranges. White indicates that no time of emergence is detected.

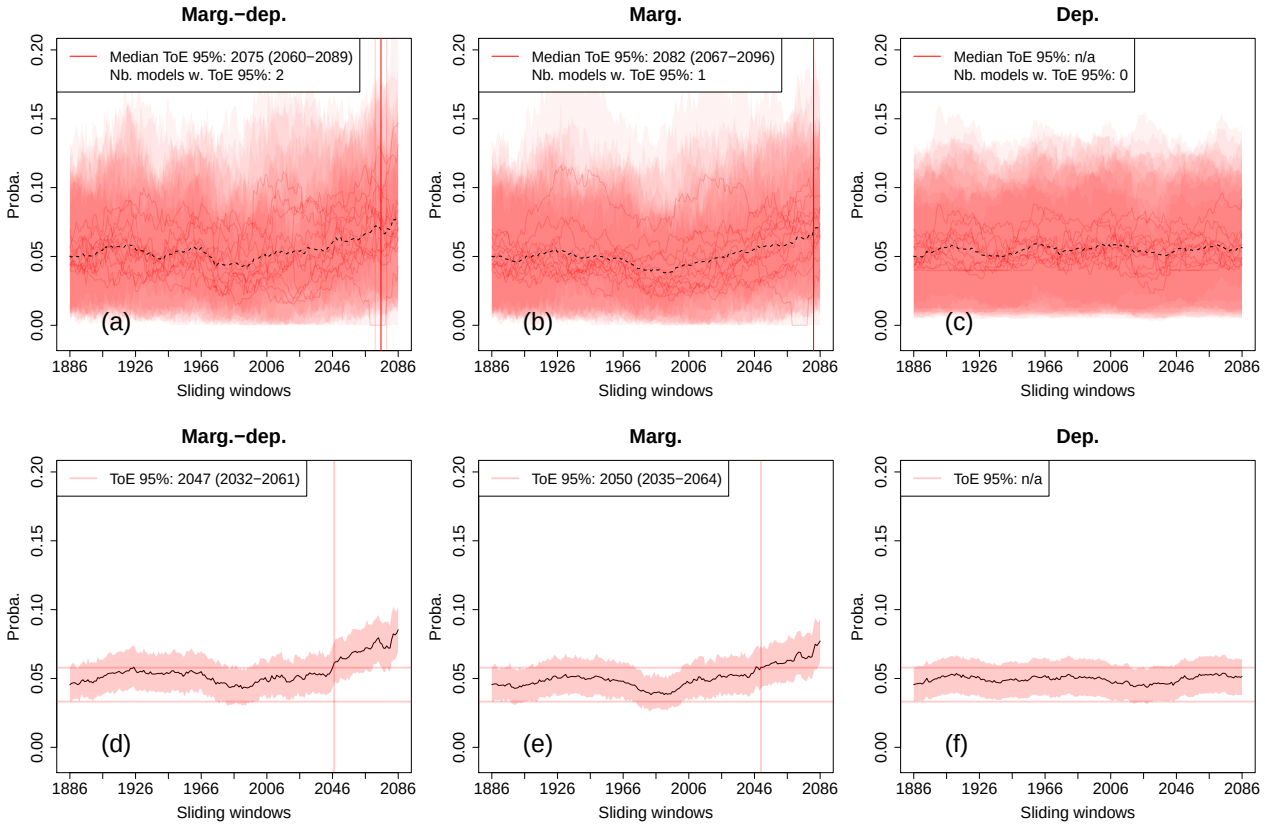

Figure S3. Same as Fig. 7 but for 95% confidence level: Probability changes and time of emergence (at 95%) of compound wind and precipitation extremes (exceeding the individual 80th percentiles of selected points of high values) for (a-c) Indiv- and (d-f) Full-Ensemble versions due to changes of (a,d) both marginal and dependence properties, (b,e) marginal properties only, and (c,f) dependence properties only. The shaded bands indicate 95% confidence intervals of the probabilities. For (a-c), individual time of emergence for the different models within the ensemble are displayed when defined (vertical light red lines), as well as the corresponding median time of emergence (vertical red lines). For information purpose, multi-model mean exceedance probability time series are also plotted (black dotted lines). Not-applicable (n/a) is indicated when no time of emergence is detected.

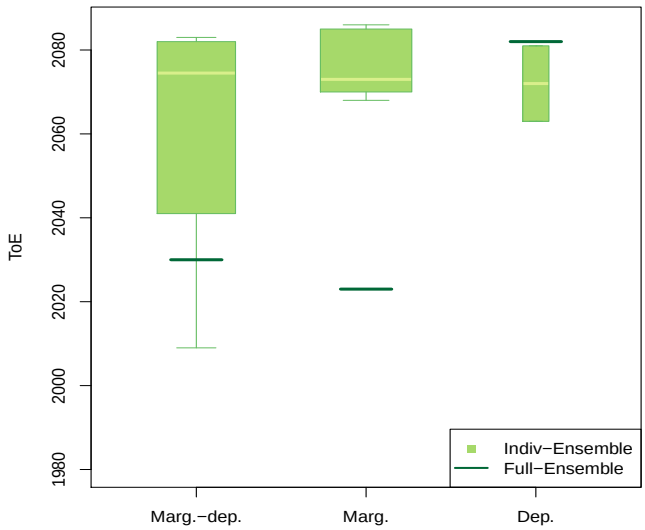

Figure S4. Boxplots of time of emergence at 68% confidence level of compound wind and precipitation extremes (exceeding the individual 80th percentiles of selected points of high values) for the Indiv-Ensemble version. Size of boxplots is proportional to the number of models presenting an emergence. For the Full-Ensemble version, values of ToE are indicated using lines.

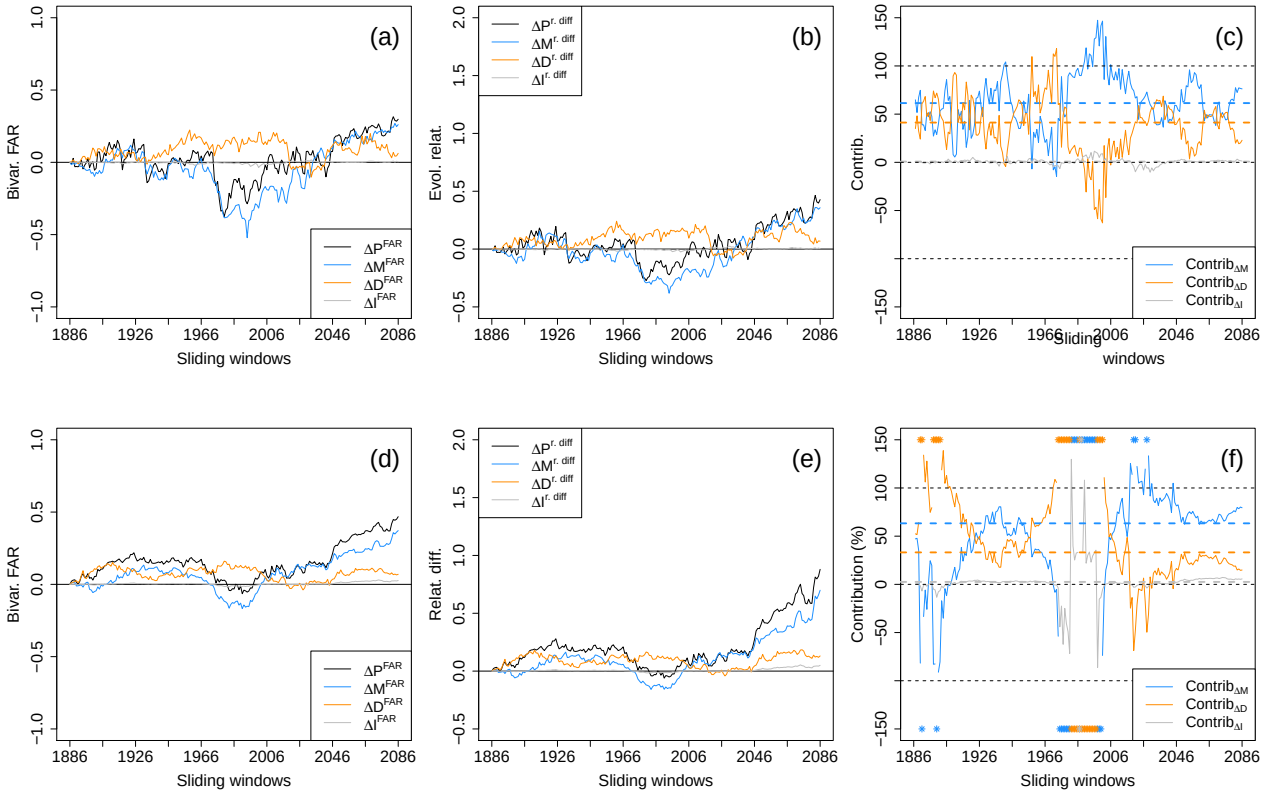

Figure S5. Evolutions of (a, d) the bivariate fraction of attributable risk (FAR), (b, e) relative difference of probabilities with respect to the reference period (1871-1900) and (c, f) contribution of the marginal, dependence and interaction terms to probability values for (a-c) Indiv- and (d-f) the Full-version. For the Indiv-Ensemble version (a-c), bivariate FAR, relative differences and contributions time series are computed by considering for each sliding window the median of the models' FAR, relative differences and contributions, respectively. Median contributions computed over all sliding windows are displayed with dotted lines. Asterisks indicate values lying outside the plotted range.

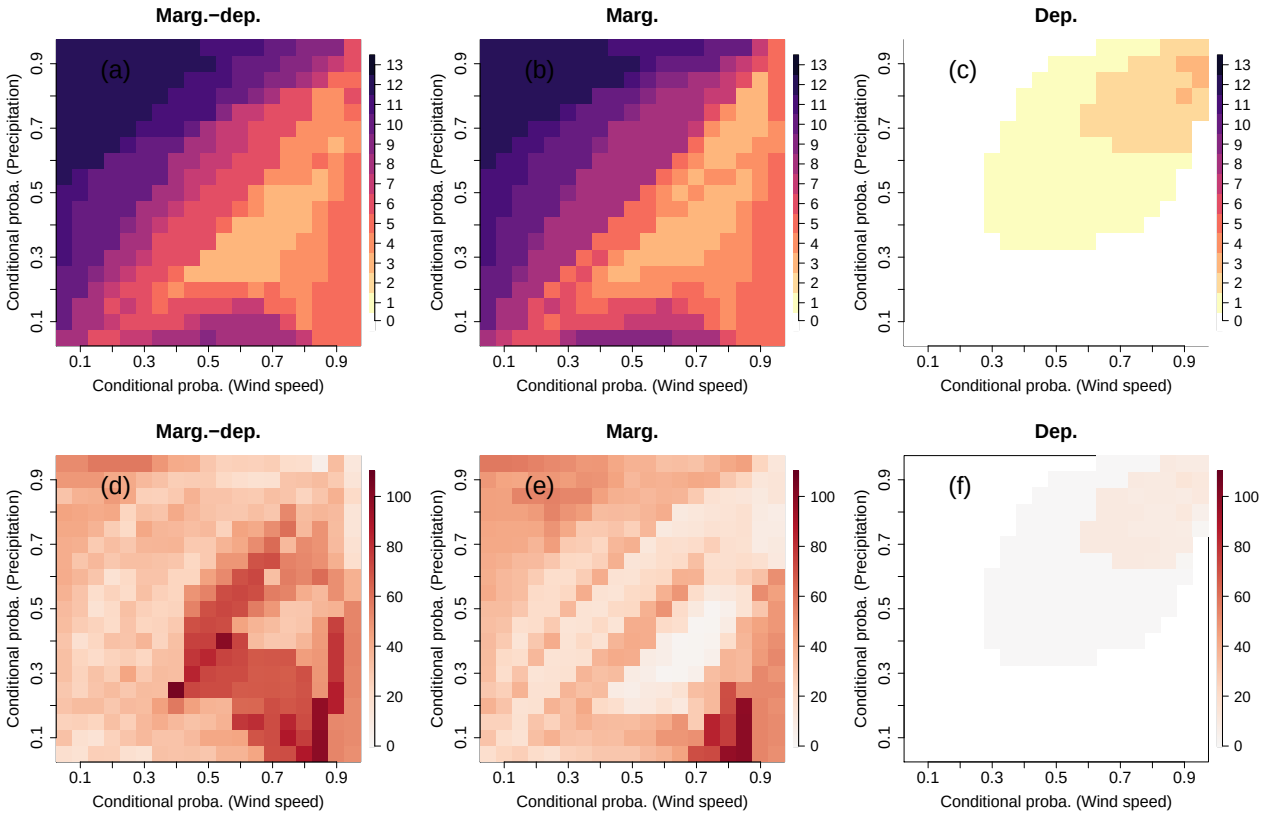

Figure S6. (a-c) Number of models within the Indiv-Ensemble framework presenting a time of emergence at 68% confidence level for compound wind and precipitation extremes. (d-f) Inter-quartile differences (Q3-Q1) of time of emergence.

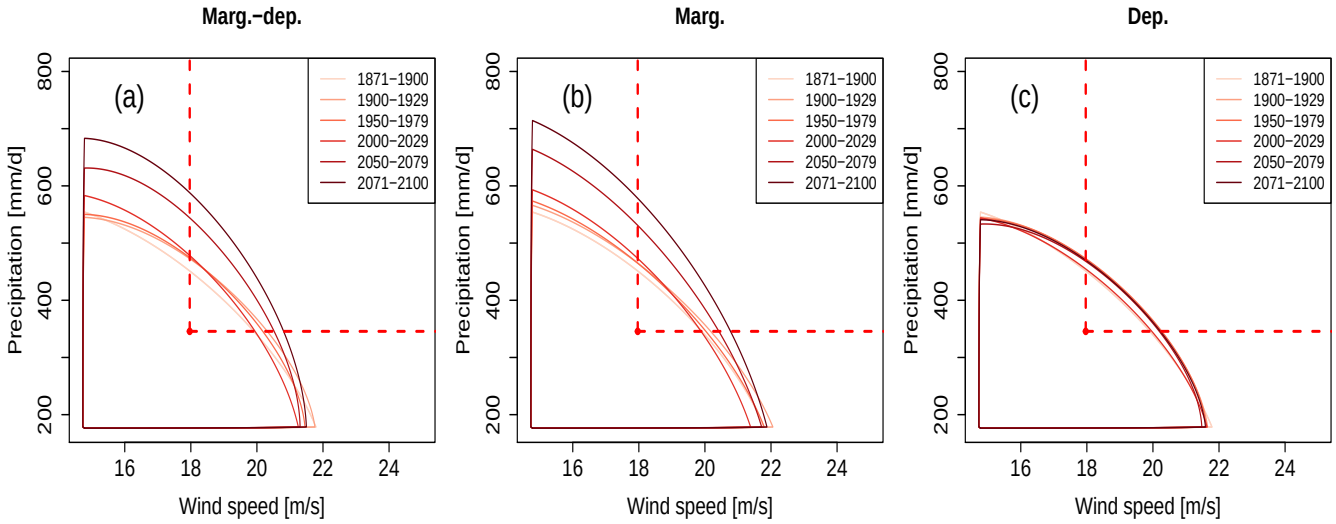

Figure S7. Same as Fig. 4 but for Full-Ensemble data: Change of compound wind and precipitation extremes distributions based on pooled data due to (a) marginal and dependence changes, (b) marginal changes while keeping dependence fixed and (c) dependence changes while keeping marginal fixed. For the bivariate distributions, contour lines encompassing 90 % of all data points are shown. A selection of six 30-years sliding windows is presented using a color gradient from light (1871-1900) to dark (2071-2100).

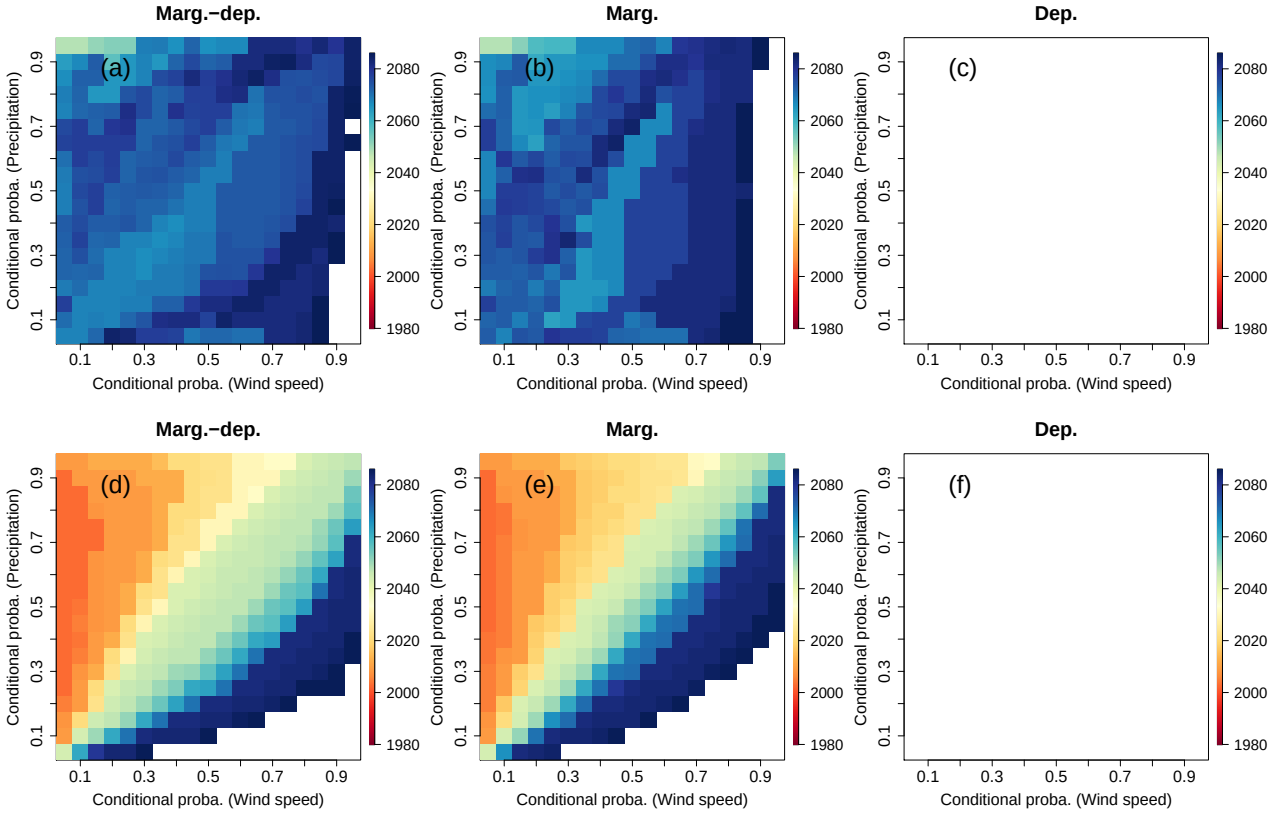

Figure S8. Same as Fig. 9 but for 95% confidence level: Time of Emergence (at 95% confidence level) matrices of compound wind and precipitation extremes due to changes of (a, d) both marginal and dependence properties, (b, e) marginal properties only, and (c, f) dependence properties only. Results are displayed for (a-c) the Indiv- and (d-f) Full-Ensemble versions for varying exceedance thresholds between the 5th and 95th percentile of compound wind and precipitation extremes data. For each subplot, white indicates that no time of emergence is detected.

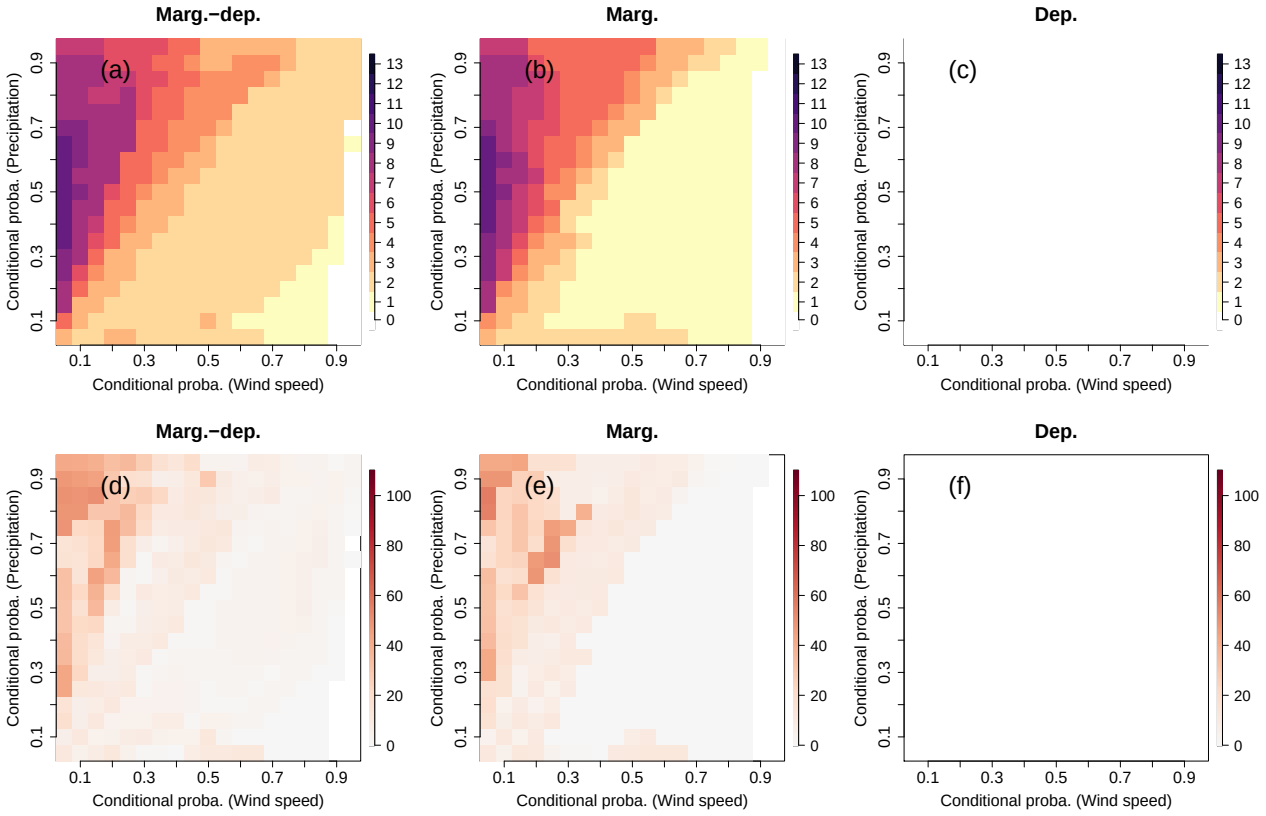

Figure S9. Same as Fig. S6 but for 95% confidence level: (a-c) Number of models within the Indiv-Ensemble framework presenting a time of emergence at 95% confidence level for compound wind and precipitation extremes. (d-f) Inter-quartile differences (Q3-Q1) of time of emergence.

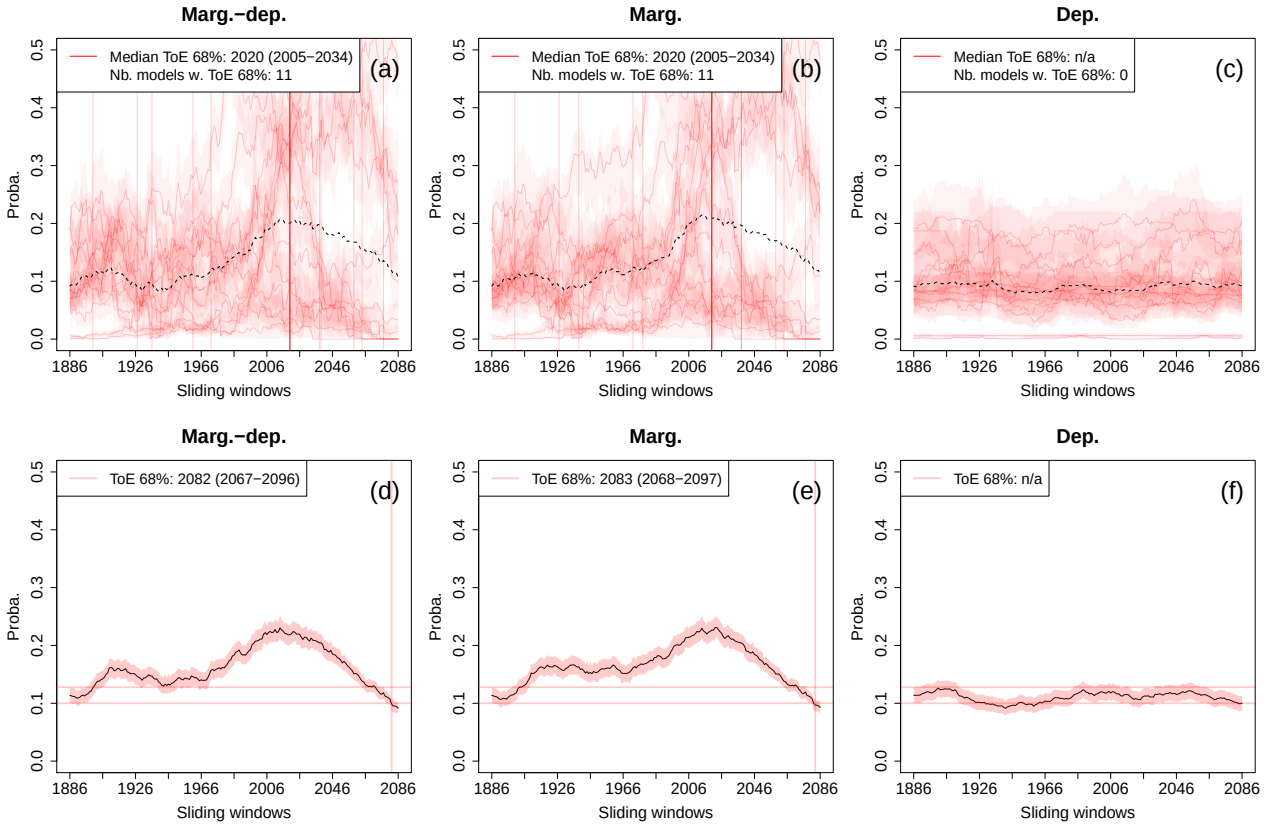

Figure S10. Same as Fig. 12 but for GDD $\geq$ 150 °C.d and minimal temperatures $\leq$ 0 °C: Probability changes and time of emergence (at 68%) of growing-period frosts (GDD $\geq$ 150 °C.d and minimal temperatures $\leq$ 0 °C) for (a-c) Indiv- and (d-f) Full-Ensemble versions due to changes of (a, d) both marginal and dependence properties, (b, e) marginal properties only, and (c, f) dependence properties only. The shaded bands indicate 68% confidence intervals of the probabilities. For (a-c), individual time of emergence for the different models within the ensemble are displayed when defined (vertical light red lines), as well as the corresponding median time of emergence (vertical red line). For information purpose, multi-model mean exceedance probability time series are also plotted (black dotted lines). Not-applicable (n/a) is indicated when no time of emergence is detected.

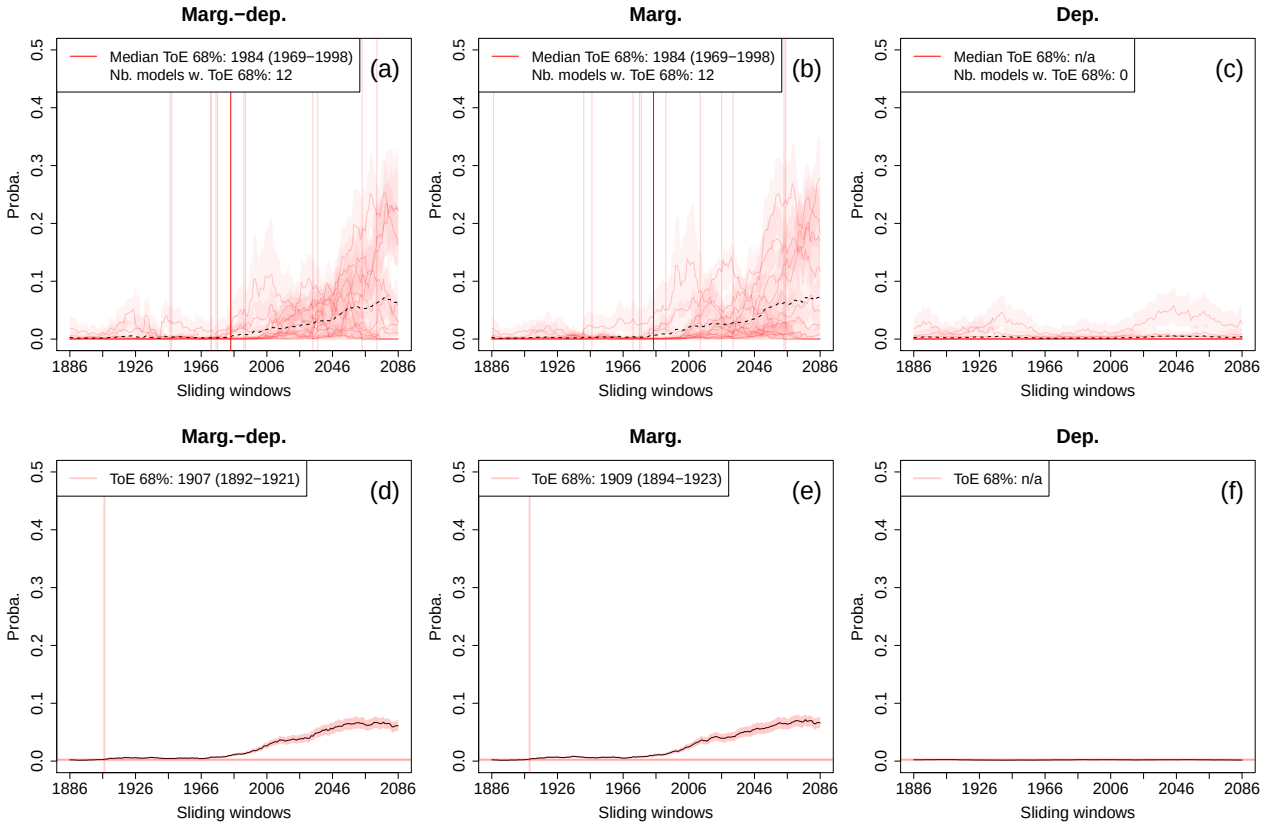

Figure S11. Same as Fig. 12 but for GDD $\geq$ 250 °C.d and minimal temperatures $\leq$ 0 °C: Probability changes and time of emergence (at 68%) of growing-period frosts (GDD $\geq$ 250 °C.d and minimal temperatures $\leq$ 0 °C) for (a-c) Indiv- and (d-f) Full-Ensemble versions due to changes of (a, d) both marginal and dependence properties, (b, e) marginal properties only, and (c, f) dependence properties only. The shaded bands indicate 68% confidence intervals of the probabilities. For (a-c), individual time of emergence for the different models within the ensemble are displayed when defined (vertical light red lines), as well as the corresponding median time of emergence (vertical red line). For information purpose, multi-model mean exceedance probability time series are also plotted (black dotted lines). Not-applicable (n/a) is indicated when no time of emergence is detected.

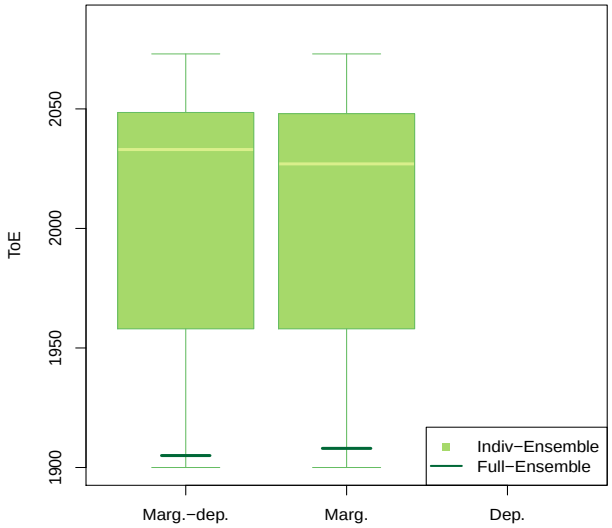

Figure S12. (a) Boxplots of time of emergence at 68% confidence level of growing-period frosts (GDD $\geq$ 200 °C.d and minimal temperatures $\leq$ 0 °C) for the Indiv-Ensemble version. Size of boxplots is proportional to the number of models presenting an emergence. For the Full-Ensemble version, values of ToE are indicated using lines.

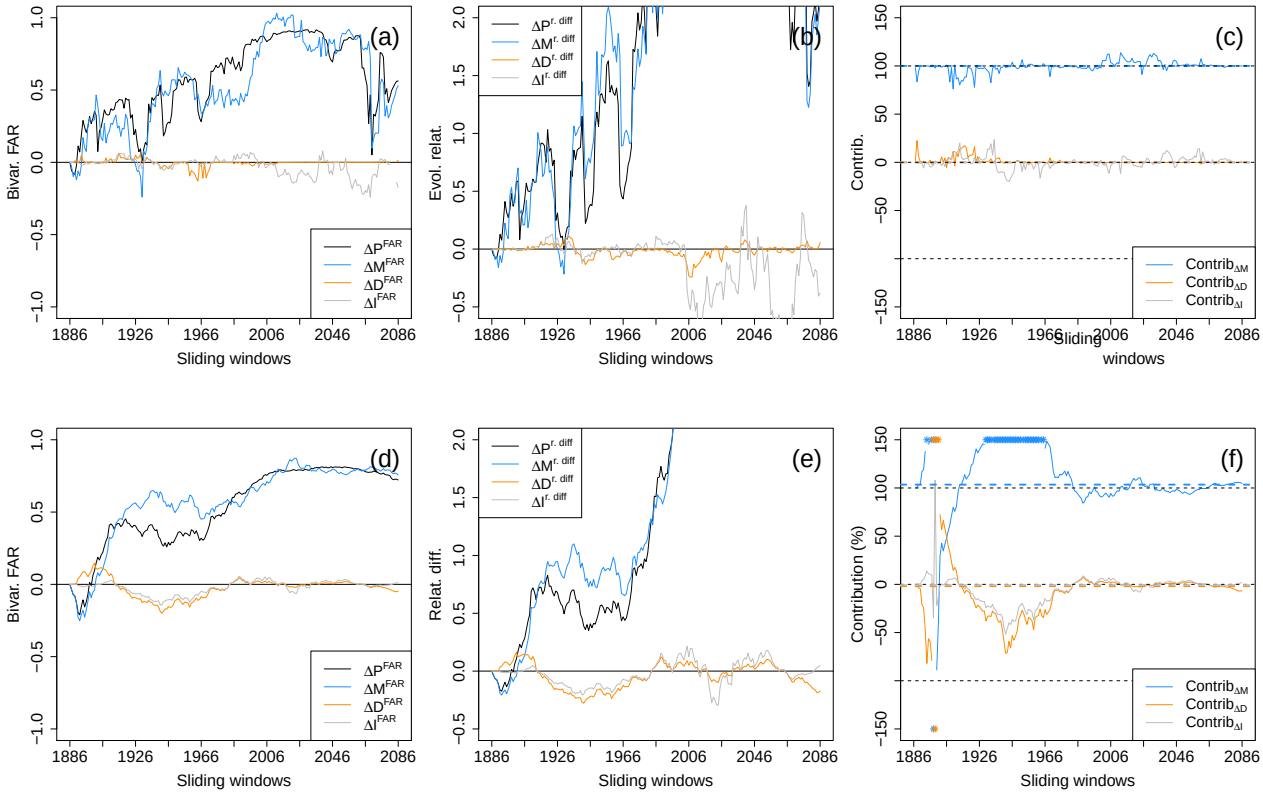

Figure S13. Same as Fig. S5 but for growing-period frost events ($GDD \geq 200 \cap T \leq 0$). Evolutions of (a, d) the bivariate fraction of attributable risk (FAR), (b, e) relative difference of probabilities with respect to the reference period (1871-1900) and (c, f) contribution of the marginal, dependence and interaction terms to probability values for (a-c) Indiv- and (d-f) the Full-version. For the Indiv-Ensemble version (a-c), bivariate FAR, relative differences and contributions time series are computed by considering for each sliding window the median of the models' FAR, relative differences and contributions, respectively. Median contributions computed over all sliding windows are displayed with dotted lines. Asterisks indicate values lying outside the plotted range.

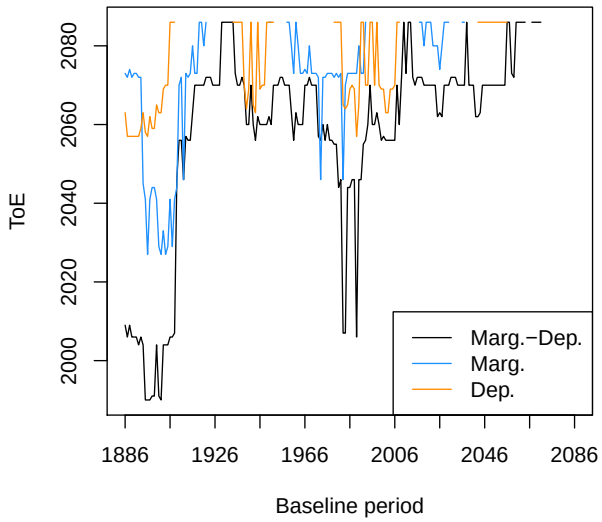

Figure S14. CNRM-CM6 time of emergence (at 68% confidence level) of compound wind and precipitation extremes probabilities ($\mathbb{P}(X > x_{80\text{lsel}} \cap Y > y_{80\text{lsel}} \mid (X,Y) \in S_{90,90}^{\text{CNRM-CM6}})$) for different baseline periods. Time of emergence are computed for probability time series when considering changes of both marginal and dependence properties ("Marg.-Dep."), marginal properties only ("Marg."), and dependence properties only ("Dep."). A blank space is left when no time of emergence is detected.