Contribution of the MTG/IRS radiance assimilation in the characterisation of the atmospheric chemical composition

Francesca Vittorioso

THÈSE

En vue de l'obtention du
DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par l'Université Toulouse 3 - Paul Sabatier

Présentée et soutenue par

Francesca VITTORIOSO

Le 4 avril 2023

Apport de l'assimilation des luminances de MTG/IRS pour caractériser la composition chimique de l'atmosphère

---

Ecole doctorale : **SDU2E - Sciences de l'Univers, de l'Environnement et de l'Espace**

Spécialité : **Océan, Atmosphère, Climat**

Unité de recherche :
CNRM - Centre National de Recherches Météorologiques

Thèse dirigée par
Nadia FOURRIÉ, Vincent GUIDARD et Marine CLAEYMAN

Jury

Mme Solène TURQUETY, Rapporteure
M. Vincent-Henri PEUCH, Rapporteur
M. Federico FIERLI, Rapporteur
M. Gilles FORET, Examinateur
Mme Nadia FOURRIÉ, Directrice de thèse
M. Vincent GUIDARD, Co-directeur de thèse
Mme Marine CLAEYMAN, Co-directrice de thèse
Mme Valerie THOURET, Présidente

«Considerate la vostra semenza:
fatti non foste a viver come bruti,
ma per seguir virtute e conoscenza[1]*.»*

Dante Alighieri, *INFERNO, Canto XXVI*

[1] *Call to mind from whence ye sprang: ye were not form'd to live the life of brutes, but virtue to pursue and knowledge high.*

Abstract

In the coming years, EUMETSAT's Meteosat Third Generation - S (MTG-S) satellites will be launched with an instrument of valuable features on board. The MTG - Infrared Sounder (IRS) will represent a major innovation for the monitoring of the chemical state of the atmosphere, since, at present, observations of these parameters mainly come from *in situ* measurements (geographically uneven) and from instruments on board of polar-orbiting satellites (highly dependent on the scanning line of the satellite itself, which is limited, over a specific geographical area, to very few times per day). MTG-IRS will present many potentialities in the area of detecting different atmospheric species and will have the advantage of being based on a geostationary platform and to acquire data with a high temporal frequency (every 30 minutes over Europe), which makes easier to track, among the others, the transport of the species of interest.

This is the context for the present research work, which aims to evaluate the impact of the assimilation of IRS radiances within a chemical transport model (CTM), such as the Modèle de Chimie Atmosphérique de Grande Echelle (MOCAGE), i.e. the CTM operational in Météo-France.

While MTG-IRS is still being built and validated by Thales Alenia Space, which is in charge of its development, the future is already being explored. An IRS instrument with very similar characteristics to the first one, but with a better spectral sampling (namely $0.325\,\mathrm{cm}^{-1}$ vs the $0.605\,\mathrm{cm}^{-1}$ of the first version) is being studied. This will be referred to as IRS*2 in the course of the manuscript. The second objective of this project is, therefore, to assess the different contributions that the two instrument versions could possibly make in providing atmospheric composition information when assimilated into a CTM such as MOCAGE.

In order to make this evaluation, IRS observations (for both its studied versions) had to be simulated as accurately as possible since IRS is not yet in orbit, and its version with a higher spectral sampling is still being evaluated. To do so, the Observing System Simulation Experiment (OSSE) method was exploited. Of the species to which IRS will be sensitive, the one treated along this study was ozone. However, the groundwork has been laid in order to be able to extend the research to other species in any work that will follow this thesis.

The results obtained from the present study indicate that the assimilation of synthetic radiances of IRS, in both its versions, always has a positive impact on the ozone analysis from the model MOCAGE.

Contribution of IRS*2 versus IRS depends on the atmospheric levels. In the stratosphere IRS assimilation seems to do better than that of IRS*2. In the troposphere, on the other hand, assimilating IRS*2 radiances into MOCAGE provides results that are closer to the reference reality compared to those obtained using IRS data.

Résumé

Dans les années à venir, les satellites Meteosat Troisième Génération - S (MTG-S) d'EUMETSAT seront lancés avec à leur bord un instrument aux caractéristiques précieuses. MTG - Infrared Sounder (IRS) représentera une innovation majeure pour la surveillance de l'état chimique de l'atmosphère, car, à l'heure actuelle, les observations de ces paramètres proviennent principalement de mesures in situ (géographiquement hétérogènes) et d'instruments embarqués à bord de satellites en orbite polaire (fortement dépendants du passage du satellite, qui est limité, sur une zone géographique spécifique, à très peu de fois par jour). MTG-IRS présentera de nombreuses potentialités dans le domaine de la détection de différentes espèces atmosphériques et aura l'avantage d'être basé sur une plateforme géostationnaire et d'acquérir des données avec une fréquence temporelle élevée (toutes les 30 minutes sur l'Europe), ce qui facilite le suivi, entre autres, du transport des espèces d'intérêt.

C'est dans ce contexte que s'inscrit le présent travail de recherche, qui vise à évaluer l'impact de l'assimilation des radiances IRS au sein d'un modèle de chimie-transport, tel que le Modèle de Chimie Atmosphérique de Grande Echelle (MOCAGE), c'est-à-dire le modèle opérationnel à Météo-France.

Alors que MTG-IRS est encore en cours de construction et de validation par Thales Alenia Space, qui est en charge de son développement, l'avenir est déjà exploré. Un instrument IRS aux caractéristiques très similaires au premier, mais avec un échantillonnage spectral plus élevé (à savoir $0,325\,\mathrm{cm}^{-1}$ *contre le* $0,605\,\mathrm{cm}^{-1}$ *de la première version) est déjà étudié. Par souci de simplicité, celui-ci sera appelé IRS*2 dans le manuscrit. Le deuxième objectif de ce projet est donc d'évaluer les différentes contributions que les deux versions de l'instrument pourraient éventuellement apporter en fournissant des informations sur la composition de l'atmosphère lorsqu'elles sont assimilées dans un MCT tel que MOCAGE.*

Pour réaliser cette évaluation, les observations d'IRS (pour les deux versions étudiées) ont dû être simulées aussi précisément que possible, puisque IRS n'est pas encore en orbite et que sa version avec un échantillonnage spectral plus élevé est encore en cours d'évaluation. Pour ce faire, la méthode OSSE (Observing System Simulation Experiment) a été exploitée. Parmi les espèces auxquelles IRS sera sensible, celle traitée dans le cadre de cette étude est l'ozone. Cependant, les bases ont été posées afin de pouvoir étendre la recherche à d'autres espèces dans les travaux qui suivront cette thèse.

Les résultats obtenus dans le cadre de l'étude indiquent que l'assimilation des radiances synthétiques d'IRS, dans ses deux versions, a toujours un impact positif sur l'analyse de l'ozone à partir du modèle MOCAGE.

*La contribution d'IRS*2 par rapport à IRS dépend des niveaux atmosphériques. Dans la stratosphère, l'assimilation IRS semble faire mieux que celle d'IRS*2. En revanche, dans la troposphère, l'assimilation des luminances IRS*2 dans MOCAGE fournit des résultats plus proches de la réalité de référence par rapport à ceux obtenus avec les données IRS.*

Contents

List of Figures

List of Tables

Introduction

The chemical composition of Earth's atmosphere represents a factor of paramount importance for the unfolding of life on the planet. Its effects can potentially impact a very wide range of events: from climate to weather events, from the development to the natural unfolding of life itself. Given the stratified structure of the atmosphere, different chemical processes characterise different atmospheric layers and can have different genesis and developments, and thus different consequences and effects (both positive and negative ones). Some atmospheric compounds, for instance, can shield harmful radiations coming from space. At the same time, other compounds exceeding limit thresholds near the surface can deteriorate air quality and cause damages to both flora and fauna.

Factors that determine the chemical composition of the atmosphere are multiple and will be better explained later in this manuscript. However, among them, human activities, due to technological development in recent centuries, are those that have most profoundly influenced the system's balance. Given its importance, the chemical state must be well monitored in order to be able, whenever necessary, to take the right steps to rectify unhealthy behaviours or to take measures in order to protect environment and population health.

As it is done in many other scientific fields, numerical models are continuously developed and refined in order to enable a proper monitoring and forecasting of the chemical situation in the atmosphere. Modelling the atmospheric composition itself, however, can be quite complex since it involves several physical and chemical processes. Sources of emission, large-scale and small-scale transport and removal processes must all be taken into account. Weather conditions, in turn, play an important role in dispersion or accumulation of chemical compounds and fine particles.

Many efforts are made in research to be able to have Chemistry Transport Models (CTMs) that predict this behaviour with quality and precision. To do this, the focus has to be set on both the efficiency of codes and methods, and on their integration with real observations of the atmospheric state. At present the observation system of the atmospheric chemical composition is mainly based on *in situ* measurements and observations from polar-orbiting satellites. However, *in situ* measurements are few and very localised. Measurements from polar-orbiting satellites, on the other hand, although nowadays quite valid and precise, are limited to the scanning line of the satellite footprint, which often limits the study of the large-scale transport of specific species. For this reason, having exploitable data from sounders

based on geostationary satellites, which can provide a full overview on larger geographical areas, and with a high acquisition frequency, would represent a significant improvement for the study of atmospheric chemical composition.

Very recently, satellite instruments capable of providing valuable information for the study of chemistry monitoring are appearing on board of geostationary platforms. The first ones on orbit are GEMS[2], an ultraviolet-visible grating imaging spectrometer, flying on board the Korean GEO-KOMPSAT-2B satellites, and the hyperspectral infrared sounder GIIRS[3] on board the Chinese FengYun-4 geostationary satellite series. The European agency EUropean Organization for the Exploitation of METeorological SATellites (EUMETSAT), however, has also envisaged to put an infrared sounder on board the geostationary Meteosat Third Generation (MTG)-S scheduled to be launched in 2024 as part of the MTG satellite series. This instrument, designed and built by Thales Alenia Space, is **IRS** (**Infrared Sounder**, indeed) and it will be the main character of the present work.

The primary objective of this thesis will be, in fact, to assess the contribution of the data that MTG-IRS will soon provide for the characterisation of the atmospheric chemical composition through a CTM. Such evaluation will be carried out by simulating IRS radiances first, and then performing their assimilation into Modèle de Chimie Atmosphérique de Grande Echelle (MOCAGE), i.e. the CTM operational in Météo-France. Since this is a preliminary study, the focus will be mainly put on a single species among those that the instrument will be able to detect and which, on the other hand, plays an important role in the unfolding of many atmospheric processes, namely the ozone. Nevertheless, at the same time, the foundations will also be laid for other species, such as carbon monoxide, to be dealt with in future evolution of this research.

As IRS is being built and finalised, furthermore, the future is already being designed and an IRS with an almost doubled spectral sampling is already being devised. In the present work also this second version of the instrument will be evaluated at the same time, in order to have a first quantification of its possibly different performances.

All the features of data, methods and state of the art, will be provided in detail throughout the course of the manuscript, which will be organised as follows:

- **Chapter 1** will provide a scientific overview on the Earth's atmospheric structure and chemistry. This is an important topic in order to understand the processes taking place among the chemical compounds of interest and to be able to interpret the studies undertaken from a chemical and physical point of view, as well as the results that will be obtained in the following. A special focus will be put on the ozone, which, as already mentioned, is the main species treated along this study.

- **Chapter 2** will aim to illustrate the theory of radiative transfer. To be able to explain the functioning of the satellite instrument being investigated, indeed, the theoretical

[2]Geostationary Environmental Monitoring Spectrometer (GEMS)
[3]Geostationary Interferometric Infrared Sounder (GIIRS)

concepts of the radiative processes taking place in the atmosphere are necessary. Furthermore, in addition to leading the functioning of many satellite meteorological instruments, the processes of interaction between radiation and matter are also modelled in the so-called Radiative Transfer Models (RTMs) to be exploited for the simulation of the atmospheric state. Accordingly, the RTM used through this study, i.e. Radiative Transfer for TOVS (RTTOV), will also be illustrated.

- **Chapter 3** will display the basics behind modelling of the chemical processes taking place in the atmospheric system, through the aforementioned CTMs. MOCAGE will then be illustrated, as it is the CTM exploited for carrying out the present study.

- **Chapter 4** will resume the data assimilation basics, necessary to incorporate the observations into the model. Data assimilation is a very large and complex field. Therefore, only those features directly related to the work here carried out will be reported. More in detail, the three-dimensional variational assimilation (3D-Var) method will be exploited and hence described.

- **Chapter 5** will list the observations available in the different networks active to this day, focusing especially on those used either in the preparatory studies or in the main core of the research at issue. Both *in situ* and remote sensed data will be illustrated. A major focus will, then, be placed on the description of IRS, in both its evaluated versions, and whose advantages are intended to be tested prior to this becoming operational. Studies to assess the sensitivity of the instrument to some specific chemical species have been carried out and will also be detailed.

- Since IRS is not yet operational at the time this work is carried out, and its versions with an improved resolution is still being designed, radiances have to be simulated with a high degree of accuracy. To do so, an Observing System Simulation Experiment (OSSE) has to be performed. In **Chapter 6**, the theoretical basis behind an OSSE definition will be introduced in first place. After that, the choices made in order to characterise the experiments specific of the present research will be explained. The effort to create a reference reality for these experiments and to set-up the specifications for the assimilation of radiances into a CTM, required side studies that have been object of a scientific paper. Within this latter, the development of the IASI[4] radiance assimilation into MOCAGE has been studied by examining different configurations. This allowed to compare the reference reality framework to real data and to define the criteria of radiance assimilation into MOCAGE, which have been exploited also for IRS. For this reason the mentioned paper will be here incorporated.

- **Chapter 7** will detail the results obtained. A first part of the Chapter will assess the impact of IRS in the characterisation of atmospheric composition for the regional domain

[4]Infrared Atmospheric Sounding Interferometer (IASI)

of MOGAGE. In the second part, on the other hand, the differences in the contribution of the two versions of IRS here evaluated will be presented.

Finally, the work will be closed by drawing appropriate conclusions following the examination of the results obtained. These will be complemented by the broad perspectives proposed for a possible follow-up of the project.

Introduction

La composition chimique de l'atmosphère terrestre représente un facteur d'une importance capitale pour le déroulement de la vie sur la planète. Ses effets peuvent avoir un potentiel impact sur un très large éventail d'événements : du climat aux événements météorologiques, du développement au déroulement naturel de la vie elle-même. Étant donné la structure stratifiée de l'atmosphère, différents processus chimiques caractérisent les différentes couches atmosphériques et peuvent avoir une genèse et un développement différents, et donc des conséquences et des effets différents (tant positifs que négatifs). Certains composés atmosphériques, par exemple, peuvent faire écran aux rayonnements nocifs provenant de l'espace. En même temps, d'autres composés dépassant les seuils limites près de la surface peuvent détériorer la qualité de l'air et causer des dommages à la flore et à la faune.

Les facteurs qui déterminent la composition chimique de l'atmosphère sont multiples et seront mieux expliqués dans la suite de ce manuscrit. Cependant, parmi eux, les activités humaines, du fait du développement technologique des derniers siècles, sont celles qui ont le plus profondément influencé l'équilibre du système. Compte tenu de son importance, l'état chimique doit être bien surveillé afin de pouvoir, le cas échéant, prendre les mesures adéquates pour rectifier les comportements nocifs ou prendre des mesures pour protéger l'environnement et la santé de la population.

Comme c'est le cas dans de nombreux autres domaines scientifiques, des modèles numériques sont développés et continuellement améliorés afin de permettre une surveillance et une prévision appropriées de la situation chimique dans l'atmosphère. La modélisation de la composition atmosphérique elle-même peut toutefois s'avérer assez complexe, car elle fait intervenir plusieurs processus physiques et chimiques. Les sources d'émission, les processus de transport et d'élimination à petite et grande échelle doivent tous être pris en compte. Les conditions météorologiques, quant à elles, jouent un rôle important dans la dispersion ou l'accumulation des composés chimiques et des particules fines.

De nombreux efforts de recherche sont déployés pour pouvoir disposer de modèles de chimie-transport, Chemistry Transport Models (CTMs) en anglais, qui permettent de prévoir ce comportement avec qualité et précision. Pour ce faire, l'accent doit être mis à la fois sur l'efficacité des codes et des méthodes, et sur l'intégration d'information d'observations réelle de l'état atmosphérique. À l'heure actuelle, le système d'observation de la composition chimique de l'atmosphère est principalement basé sur des mesures in situ et des observations provenant de satellites en orbite polaire. Cependant, les mesures in situ sont peu nombreuses et très localisées. Quant aux mesures des satellites en orbite polaire, bien qu'aujourd'hui plutôt utiles et précises, elles sont limitées au passage du satellite, ce qui restreint souvent l'étude du transport à grande échelle d'espèces spécifiques. Pour cette raison, disposer de données exploitables à partir de sondeurs basés sur des satellites géostationnaires, qui peuvent fournir une vue d'ensemble sur de plus grandes zones géographiques, et avec une fréquence d'acquisition élevée, représenterait une amélioration significative pour l'étude de la composition chimique

de l'atmosphère.

Très récemment, des instruments satellitaires capables de fournir des informations précieuses pour l'étude de la surveillance de la chimie font leur apparition à bord de plateformes géostationnaires. Les premiers à être mis en orbite sont GEMS[5]*, un spectromètre imageur à réseau ultraviolet-visible, embarqué à bord des satellites coréens GEO-KOMPSAT-2B, et le sondeur infrarouge hyperspectral GIIRS*[6] *à bord de la série de satellites géostationnaires chinoise FengYun-4. Mais l'agence européenne EUMETSAT a également prévu d'embarquer un sondeur infrarouge à bord du satellite MTG-S qui sera lancé en 2024, dans le cadre de la série de satellites MTG. Cet instrument, conçu et réalisé par Thales Alenia Space, est* ***IRS*** *(****Infrared Sounder****, en effet) et il sera au centre du présent travail.*

L'objectif principal de cette thèse sera, en fait, d'évaluer la contribution des données que MTG-IRS fournira bientôt pour la caractérisation de la composition chimique de l'atmosphère à travers un CTM. Cette évaluation sera réalisée en simulant d'abord les luminances d'IRS, puis en effectuant leur assimilation dans le Modèle de Chimie Atmosphérique de Grande Echelle (MOCAGE), c'est-à-dire le CTM opérationnel à Météo-France. Comme il s'agit d'une étude préliminaire, l'accent sera mis principalement sur une seule espèce parmi celles que l'instrument pourra détecter et qui, d'autre part, joue un rôle important dans le déroulement de nombreux processus atmosphériques, à savoir l'ozone. Par ailleurs, les bases seront également jetées pour d'autres espèces, telles que le monoxyde de carbone, qui seront traitées dans le cadre de l'évolution future de cette recherche.

Alors qu'IRS est en cours de construction et de finalisation, l'avenir se prépare déjà et un IRS avec un échantillonnage spectral presque doublé est déjà en cours d'investigation. Dans le présent travail, cette deuxième version de l'instrument sera évaluée en même temps, afin d'obtenir une première quantification de ses éventuelles différentes performances.

Toutes les caractéristiques des données, des méthodes et de l'état de l'art, seront fournies en détail tout au long du manuscrit, qui sera organisé comme suit :

- *Le* ***Chapitre 1*** *fournira un aperçu scientifique sur la structure et la chimie de l'atmosphère terrestre. Il s'agit d'un sujet important pour comprendre les processus qui se déroulent parmi les composés chimiques d'intérêt et pour pouvoir interpréter les études entreprises d'un point de vue chimique et physique, ainsi que les résultats qui seront obtenus par la suite. Un accent particulier sera mis sur l'ozone qui, comme déjà mentionné, est la principale espèce traitée dans le cadre de cette étude.*

- *Le* ***Chapitre 2*** *aura pour but d'illustrer la théorie du transfert radiatif. En effet, pour pouvoir expliquer le fonctionnement de l'instrument satellite étudié, les concepts théoriques des processus radiatifs se déroulant dans l'atmosphère sont nécessaires. En outre, en plus de piloter le fonctionnement de nombreux instruments météorologiques satellitaires, les processus d'interaction entre le rayonnement et la matière sont également*

[5] Geostationary Environmental Monitoring Spectrometer
[6] Geostationary Interferometric Infrared Sounder (GIIRS) en anglais

modélisés dans les modèles de transfert radiatif, Radiative Transfer Models (RTMs), exploités pour la simulation de l'état atmosphérique. En conséquence, le RTM utilisé dans le cadre de cette étude, à savoir RTTOV, sera également illustré.

- *Le **Chapitre 3** présentera les bases de la modélisation des processus chimiques ayant lieu dans le système atmosphérique, par le biais des CTMs susmentionnés. MOCAGE sera ensuite détaillé car il s'agit de CTM exploité pour réaliser la présente étude.*

- *Le **Chapitre 4** reprendra les bases de l'assimilation de données, nécessaire pour introduire l'information des observations dans le modèle. L'assimilation de données est un domaine très vaste et complexe, par conséquent, seules les caractéristiques directement liées au travail effectué ici seront rapportées. La méthode d'assimilation variationnelle tridimensionnelle (3D-Var) exploitée ici sera donc décrite en détail.*

- *Le **Chapitre 5** énumérera les observations disponibles dans les différents réseaux actifs à ce jour, en se concentrant particulièrement sur celles utilisées soit dans les études préparatoires, soit dans le coeur principal de la recherche en question. Les données in situ et les données de télédétection seront illustrées. Un accent majeur sera ensuite mis sur la description d'IRS, dans ses deux versions évaluées, et dont les avantages sont destinés à être testés avant que celui-ci ne devienne opérationnel. Des études visant à évaluer la sensibilité de l'instrument à certaines espèces chimiques spécifiques ont été réalisées et seront également détaillées.*

- *Étant donné qu'IRS n'est pas encore opérationnel au moment où ce travail est réalisé, et que sa version à résolution améliorée est encore en cours de conception, les radiances doivent être simulées avec un haut degré de précision. Pour ce faire, un cadre d'expérimentation permettant de simuler des systèmes d'observation (OSSE en anglais) doit être effectué. Dans le **Chapitre 6**, les bases théoriques de la définition d'un OSSE seront d'abord présentées. Ensuite, les choix effectués pour caractériser les expériences spécifiques de la présente recherche seront expliqués. L'effort pour créer une réalité de référence pour ces expériences et pour mettre en place les spécifications pour l'assimilation des radiances dans un CTM, a nécessité des études annexes qui ont fait l'objet d'un article scientifique. Dans le cadre de ce dernier, le développement de l'assimilation de radiances de l'Infrared Atmospheric Sounding Interferometer (IASI) dans le modèle MOCAGE a été étudié en examinant différentes configurations. Cela a permis de comparer le cadre de la réalité de référence aux données réelles et de définir les critères d'assimilation de la radiance dans MOCAGE, qui ont été exploités également pour IRS. Pour cette raison, l'article mentionné sera ici incorporé.*

- *Le **Chapitre 7** détaillera les résultats obtenus. Une première partie du chapitre évaluera l'impact d'IRS dans la caractérisation de la composition atmosphérique au-dessus du*

domaine régional de MOGAGE. Dans la seconde partie, en revanche, les différences de contribution des deux versions d'IRS ici évaluées seront présentées.

Enfin, le manuscrit se terminera en tirant les conclusions appropriées suite à l'examen des résultats obtenus. Celles-ci seront complétées par des perspectives plus larges proposées pour un éventuel poursuite de ce projet.

Chapter 1

Earth's Atmosphere

The atmosphere of a planet is a gaseous layer surrounding the planet itself. Planet Earth has a very special atmosphere with complex characteristics that, in subtle balance with each other, made life possible over our planet lifetime. It has the important function of shielding the surface from harmful radiations coming from space, it allows the surface to warm up due to the greenhouse effect, and prevents strong thermal excursions. At the same time, some substances in the atmosphere itself, especially coming from anthropogenic sources due to the development of certain modern technologies, have a negative impact on the health of Earth's flora and fauna.

In this Chapter, we will look at the physical and chemical properties of the atmosphere, focusing, in particular, on its structure and chemical components. The effects of life on Earth will be described at the same time. A special focus will be put on the ozone, which is the species the most discussed throughout this study. Finally, some basic outline on carbon monoxide will be provided too, since it will be necessary for the understanding of a part of the present work.

1.1 Structure of the Earth's Atmosphere

The atmosphere of the planet Earth presents a stratification that follows the temperature profile. According to their peculiarities, layers can be identified, as schematically outlined in the Figure 1.1.

- The **troposphere** is the lowest portion of the atmosphere where life takes place. On its own, it contains about 80% of all the **air**, i.e. the gaseous mix that composes the atmosphere itself, and most of its water vapour. It is in this layer that most meteorological phenomena take place.

  Within the troposphere, temperature decreases with altitude with a gradient of about 6.5 °C per kilometre. With height, the pressure of the atmospheric column decreases. Consequently, the same volume of air, moving upwards, is subject to a decrease in

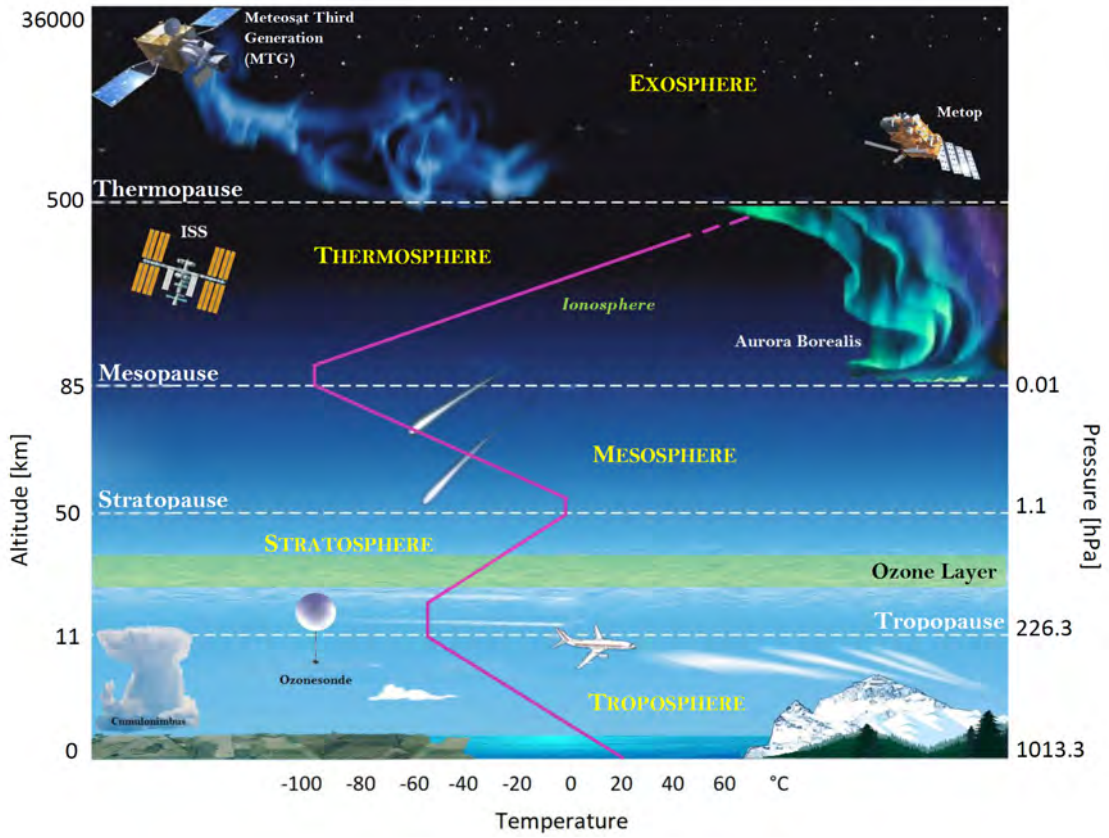


Figure 1.1: Schematic representation of atmospheric vertical structure, associated with the typical vertical temperature trend (averaged, approximate and schematic representation of the temperature profile is depicted in magenta). The altitudes and pressure levels are not in linear scale and they only provide an approximate indication of the average position of each layer.

pressure and thus in temperature (for the Ideal Gas Law[1]). The air in the upper troposphere is colder and rarefied than it is at lower levels. The temperature gradient, however, may vary depending on the time of the year, the geographical area, and the meteorological phenomena affecting the atmosphere in that particular time and place.

The tropospheric layer the closest to the surface is generally defined ***planetary boundary layer (PBL)***. By definition, the PBL is the *"part of the troposphere that is directly influenced by Earth's surface, and responds to surface forcing with a timescale of about an hour or less"* [Stull (1988)]. This layer is, indeed, strongly influenced by the surface conditions: heat transfer, humidity, pollutants emission, drag friction, terrain. Its depth can vary from a few hundreds of meters up to 2-3 km. Within it, turbulence phenomena take place causing the mixing of the chemical compounds emitted or produced at the surface.

[1] An ideal gas is an hypothetical gas whose molecules occupy negligible space, collisions are perfectly elastic and no inter-molecular forces occur. It can be characterized by the state variables *absolute pressure* (P), *volume* (V), and *absolute temperature* (T), through the equation of state: $PV = nRT$, where n is the number of moles and R the Universal Gas Constant ($R = 8.3145$ Jmol^{-1}K^{-1}).

The portion of troposphere above the PBL is commonly known as the ***free atmosphere***, where winds are not directly retarded by surface friction and most of the clouds occur.

The upper boundary of the troposphere, on the other hand, is called ***tropopause***. The height of this layer varies with latitude (from about 7 at the Poles up to about 18 km at the Equator) and the meteorological season. Also its thickness depends on the geographical band: at the Tropics, for example, it may extend over a couple of kilometres (*tropical tropopause transition layer*). In other words, it is strongly influenced by the temperature underneath and at the surface. During warmer seasons, the tropospheric boundary rises up and lows down when it gets colder. Tropopause is the coldest region in low atmosphere, reaching temperatures up to about −50 °C.

- The temperature profile, which follows adiabatic gradient, quickly reverses after the tropopause. The layer where this occurs is the **stratosphere**, that extends from the tropopause up to about 50 km.

  Stratosphere contains very little water vapour, but it is very rich in ozone, a molecule responsible for the absorption of ultraviolet (UV) radiations coming from the Sun (this process will be explained in detail in 1.3). The stratospheric ozone layer, or ***ozonosphere***, by absorbing UV light at wavelengths below about 340 nm, plays an important role in protecting life on Earth from exposure to these harmful radiations. This layer is located a few kilometres above the tropopause and, consequently, its height varies with the latitude: it is usually found between 15 and 20 km at the Poles and 25 and 30 km at the Equator. Once its maximum concentration is reached, the amount decreases exponentially with altitude.

  The temperature increasing with height inhibits the upward movements of air masses, making the structure of this atmospheric layer very stable and horizontally stratified.

  Temperature maxima occur at the stratospheric upper limit, or ***stratopause***, a layer that extends for a few kilometres.

- The stratopause marks a boundary between the stratosphere and the coldest layer of the Earth's atmosphere, the **mesosphere**. Throughout this layer, in fact, temperature decreases again up to the lowest atmospheric values of about − 140 °C, reached at its upper boundary, the ***mesopause*** .

- The **thermosphere** is the atmospheric layer lying right above the mesopause, which sees again a change in the temperature profile trend. Temperature in thermosphere is strongly influenced by the diurnal cycle and solar activity. Just as in the stratosphere, there is an absorption of UV radiation, but also of X-rays. This high-energy beam of photons coming from the Sun, by hitting the day-time side of Earth, ionizes the gaseous components of this portion of atmosphere. The ionized layer is also called **ionosphere**. The energy lost in the process goes in heat and it results in temperature increasing and

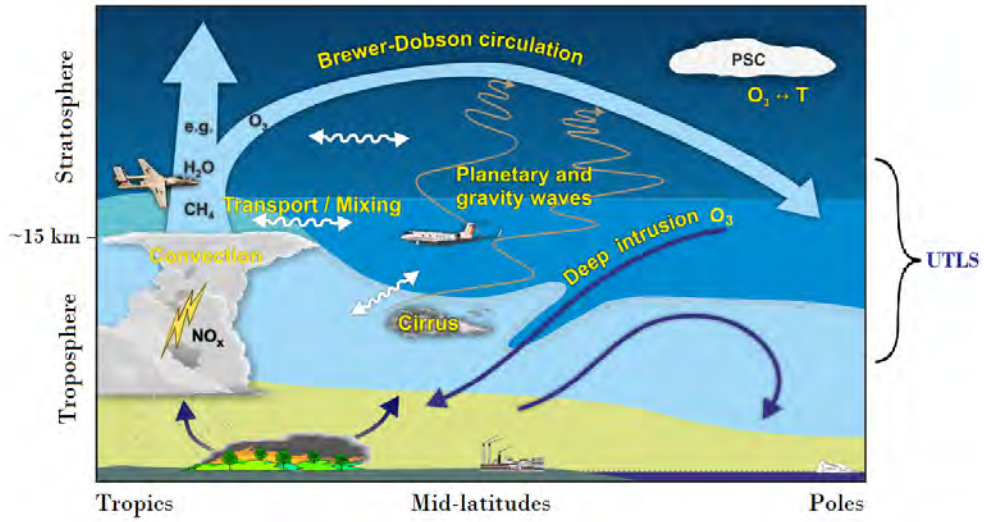


Figure 1.2: Outline of the dynamical processes and chemical mixing taking place in the Upper Troposphere - Lower Stratosphere (UTLS) [adapted from the Institute of Energy and Climate Research (IEK) website[a]].

[a] `https://www.fz-juelich.de/en/iek/iek-7/research/upper-troposphere-and-stratosphere`

reaching much higher values compared to the other atmospheric layers. Temperature in thermosphere can reach up to 2500 °C.

The thermosphere is of paramount importance in atmospheric electrical phenomena. It absorbs much of the ionising radiation coming from the outer space, protecting Earth surface by these harmful radiations. It also enables long-distance radio transmissions, by allowing radio waves to bounce back to surface. Northern lights occur in this layer. Low-orbit spatial missions take place in this layer. This is where, for example, the International Space Station (ISS) orbits.

- The **exosphere** is the outermost part of the atmosphere (about > 500 km). It contains mainly hydrogen (H) and oxygen (O) atoms that are very rarefied and therefore rarely collide with each other. From this layer, particles can manage to escape the attraction of the Earth's orbit.

Not mentioned in the classical classification of atmospheric layers is the **UTLS**[2], i.e. the coupling layer around the tropopause (broadly $\pm$ 5 km) [e.g. Gettelman et al. (2011)]. This layer is of major relevance because of the dynamical processes taking place in there and affecting the distribution of certain chemical species (see schematic representation in Figure 1.2). The stratosphere-troposphere exchanges (STEs), indeed, are important bidirectional processes influencing the chemistry of both layers [e.g. Holton et al. (1995)]. Given the barrier function of the tropopause, which normally separates the vertical transport regimes of the troposphere

[2] Upper Troposphere - Lower Stratosphere (UTLS)

and stratosphere [e.g. Hoor et al. (2002)], these processes are usually localised and require specific conditions to take place. One of the typical mixing processes is caused by massive thunderstorm columns presenting an *overshooting top*, i.e. penetrating the tropopause, and injecting water vapour into the stratosphere. Some components that contribute to the destruction of stratospheric ozone (see in more details in Subsection 1.3.1) can also be introduced during these events. Other processes, on the other hand, can lead to the intrusion of stratospheric air into the troposphere. Among others, the *tropopause folding* events, that are the most efficient phenomena of STEs at mid-latitudes, are responsible for the intrusion of dry stratospheric air into the troposphere, allowing the chemical transport of ozone and nitrogen oxides. The breaking of some planetary waves and certain components of the polarward atmospheric circulation (see *Brewer-Dobson Circulation* in the following 1.3.1), can also result in STEs.

The frequency of all of these localised mixing processes tends to be variable in nature. This leads to rapidly-changing small-scale gas features that impact the chemical composition of the atmosphere around the UTLS [e.g. Hoor et al. (2010)]. The distribution of these gases can have important impacts on the radiative balance and the climate situation. For this reason, knowledge of the phenomena taking place in such layer is of primary importance.

1.2 Atmospheric Chemical Composition

The chemical composition of the atmosphere varies according to the altitude. Temperature and pressure gradients influence the vertical distribution of each chemical compound. At the same time, the sources of emission of the gases themselves or of their *precursors*[3], together with the atmospheric dynamic action, play an important role in vertical, zonal and meridional gases's distribution.

Overall, Earth's atmosphere mainly consists of diatomic molecules of nitrogen (N_2), at almost 78% (percentage by volume of dry air), and oxygen (O_2), at $\sim 21\%$. The remaining portion of atmosphere is formed by **trace gases**, i.e. any type of gas found in a planet atmosphere in concentrations smaller than 1%. The noble gas argon (Ar) is the most widespread Earth's trace gas, followed by carbon dioxide (CO_2). Although in extremely small percentages, traces of other components are present as well and play an important role in the environment life. A schematic representation of the air composition percentages is shown in Figure 1.3, while detailed amount per volume of most of these gases are listed in Table 1.1.

Air also contains water vapour (H_2O), mainly found in troposphere and in variable quantities, often ranging in $1 - 3\%$.

[3] compounds that participates in the chemical reaction to produce another compound.

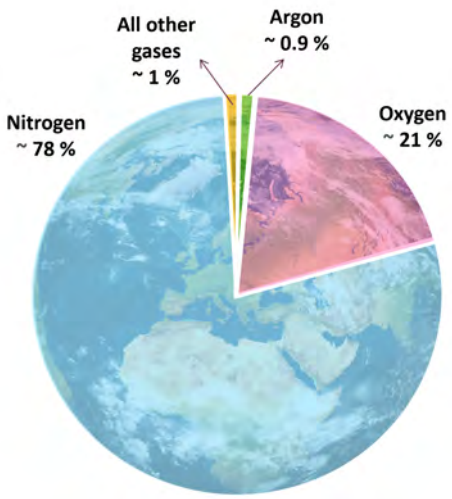


Figure 1.3: Pie chart for the amount of the main components in the Earth's atmosphere.

Gas	Chemical Formula	Percentage (%) by Volume of Dry Air
Nitrogen (molecular)	N_2	78.084
Oxygen (molecular)	O_2	20.946
Argon	Ar	0.934
Carbon Dioxide	CO_2	0.0370
Neon	Ne	$1.82 \cdot 10^{-3}$
Helium	He	$5.2 \cdot 10^{-4}$
Methane	CH_4	$\approx 1.5 \cdot 10^{-4}$
Krypton	Kr	$\approx 1.14 \cdot 10^{-4}$
Hydrogen	H_2	$\approx 5 \cdot 10^{-5}$
Nitrous Oxide	N_2O	$\approx 5 \cdot 10^{-5}$
Xenon	Xe	$\approx 8.6\,5 \cdot 10^{-6}$
Ozone	O_3	$2 \text{ to } 200 \cdot 10^{-6}$
Carbon Monoxide	CO	$0 \text{ to } 1 \cdot 10^{-5}$

Table 1.1: Main gaseous chemical components in atmosphere and their percentages per volume of dry air [e.g. Delmas et al. (2005); Tkemaladze and Makhashvili (2016)].

1.2.1 Greenhouse Gases

The most important trace gases found in Earth's atmosphere are the so-called **greenhouse gases (GhG)**. The distinguishing feature of this class of gases is their ability to absorb infrared (IR) radiation emitted by the Earth's surface and to turn it into heat that warms up the atmosphere. Actually, the solar radiation managing to travel down to the surface and to penetrate the atmosphere without being scattered, absorbed or reflected, is absorbed by surface itself, which warms up and, in turn, re-emits radiation in the IR range. A part of this radiation passes through the atmosphere. Most of it, however, is captured by the GhG that, after initially storing it in their vibrations, reconvert it rapidly into heat warming up the atmosphere we live in. This natural effect, known indeed as the **greenhouse effect**, largely influences Earth's radiative balance. Without it our planet surface temperature would not have been suitable for life to develop and evolve.

Greenhouse gases include water vapour, carbon dioxide and methane (CH_4), followed, at a lesser extent, by surface-level ozone, nitrous oxide (N_2O) and fluorinated gases, such as the the hydrofluorocarbones (HFCs).

1.2.2 Non-gaseous Components

The atmosphere does not contain only gaseous components, but also solid or liquid ones, which can be hydrometeors or **aerosols**. The latter not only have an important impact on human health, but also affect climate and meteorology. They take part in reflection and absorption of

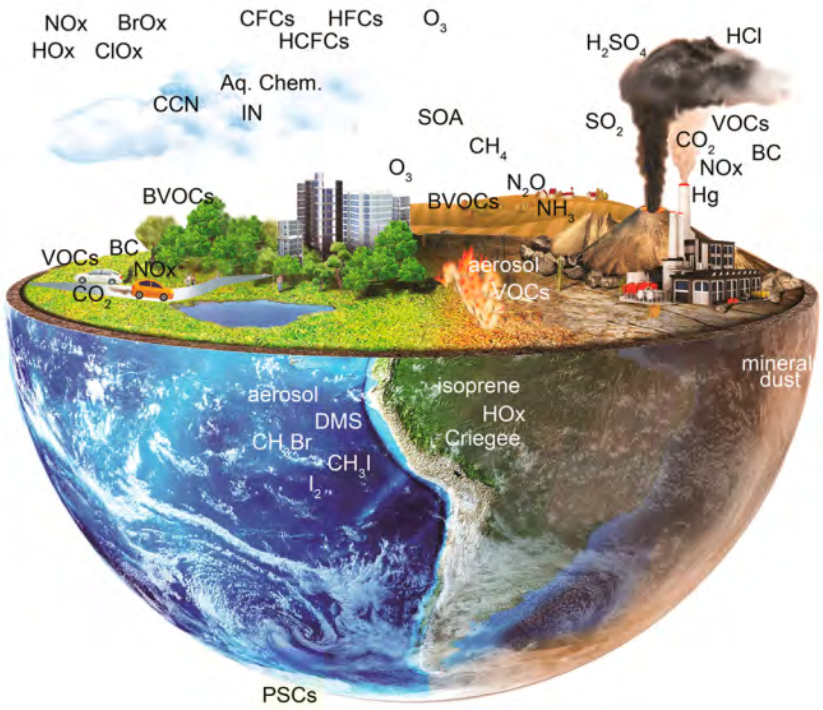


Figure 1.4: Schematic view of biogenic and anthropogenic emission sources and some of the main atmospheric species [Burkholder et al. (2017)].

the visible (VIS) solar radiation (*albedo*[4]), and can also absorb IR radiation taking part to the Earth's radiative balance. Aerosols also act as substrates for the formation of hydrometeors (*cloud condensation nuclei*) influencing the chemical composition of the precipitation itself. Aerosols can be ***primary***, which are directly emitted into the atmosphere, and ***secondary***, formed through oxidation, condensation, and multi-phase chemical processes.

1.2.3 Sources and Transport

The sources of the main atmospheric components are mostly located at the surface. A distinction can be made between natural sources, including **wind erosion**, **volcanic** and **biogenic** emissions, and human induced emissions, or **anthropogenic**. Some sources, however, are both natural and anthropogenic, e.g. wildfires. Figure 1.4 summarises the main emission sources and some key atmospheric species. Many human activities release in the air chemical compounds, which interfere with normal biological cycles. Burning of fossil fuels, deforestation, land use and land-use changes, livestock production, fertilisation, waste management and industrial processes are responsible for the injection in the air we breathe of many compounds, like CO_2, or other greenhouse gases (see in the following), and volatile

[4]ratio of the light received by a body to the light reflected by that body: a perfectly white, i.e. reflecting, body has albedo equal to 1, while a perfectly black body has albedo equal to 0, i.e. it absorbs all the radiation it receives.

organic compounds (VOCs) that deeply influence the Earth climate and the biome[5].

The biogenic sources responsible for releasing chemical components into the atmosphere are mainly vegetation, soil, wetland, oceans. The natural cycles of these elements, however, are by now also inevitably influenced by human activities. Volcanoes, although in the spectrum of biogenic sources, deserve separate mention. Sulphur dioxide (SO_2) is one of the most common gases released in volcanic eruptions, together with H_2O and CO_2. The massive injection in atmosphere of this compound has the ability to influence climate on a global scale. Sulphur dioxide is, indeed, converted to sulphuric acid (H_2SO_4), which condenses in stratosphere to form fine sulfate aerosols. These latter increase the reflection of solar radiation back to space, thereby cooling the atmospheric layer underneath. On the local scale, on the other hand, in its gaseous form and because of its oxidises forming sulphate aerosols, it directly affects human health.

The VOCs, which are a family of chemicals including a large variety of species with different physical and chemical behaviours, mainly come from anthropogenic sources (e.g. manufacture of paints, pharmaceutical productions, and refrigerants). VOCs are typically industrial and they are often cause of ground-water contaminants. However, also plants emit a range of volatile compounds such as nitrous oxide, carbon monoxide, and non-methane VOCs, the so-called biogenic volatile organic compounds (BVOCs).

After being emitted, many of these species can be transported (both in horizontally and vertically) and may be transformed by solar radiation. Depending on their lifetimes, they can have an impact on larger or smaller geographical and temporal scales.

1.2.4 Vertical Structure

In Figure 1.5 are the vertical distribution and mixing ratio of the most relevant atmospheric species. Species with a long lifespan, of the order of tens or hundreds of years, show an almost homogeneous concentration. This is the case, for example, of carbon dioxide or methane. Their concentrations sometimes have a little gradient between hemispheres due to a larger density of emission sources in the northern hemisphere. On the other hand, gases with lifetime shorter than the transport period (i.e. a few months) have a more varied distribution. This is the case for example for tropospheric ozone or carbon monoxide. Distributions are highly uneven for very short-lived gases.

Air quality, which is associated with atmospheric composition near the surface, can deeply affect population health and environment. It is influenced by local emissions, chemistry and small-scale and large-scale transport. Its modelling is complex because it involves several physical and chemical processes. Weather conditions, and especially winds, play an important role in dispersion or accumulation of chemical compounds and fine particles. The achievement of air quality standards is a major issue in modern society in order to preserve future generations. The exploitation of fossil fuels to sustain life of the world population has

[5] a "biome" is a large, naturally occurring, community of flora and fauna occupying a major habitat.

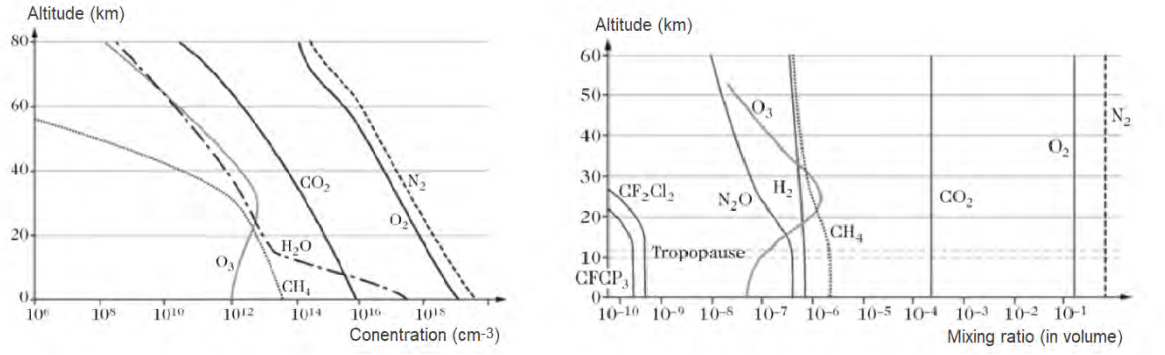


Figure 1.5: Vertical distribution of major chemical species present in Earth's atmosphere and mixing ratio [Delmas et al. (2005)].

been so high that it led to production of many waste materials and to the release in the environment of many harmful compounds that have probably already compromised the future equilibrium of Earth ecosystems.

1.3 Ozone

As already introduced in the previous section, ozone is a trace gas in planet Earth atmosphere. It is a secondary species, i.e. it is never emitted, but rather formed through chemical reactions. Ozone is present in atmosphere in rather low quantities compared to other components. Nevertheless, it plays a fundamental role in the unfolding of life on Earth because of the chemical processes it is involved in.

The largest portion of atmospheric ozone is found in stratosphere (almost 90%), where it shields UV radiation coming from the Sun, which can be very harmful for life. On the other hand, a smaller percentage of ozone is also found in troposphere. This usually comes from intrusion of stratospheric air into the troposphere or, more often, from reactions taking place in the lowest layers and involving primary compound of natural or anthropogenic sources (e.g. hydrocarbons, like CH_4, and nitrogen oxides).

In the following, we will have a careful look at the chemical processes of production and removal of both stratospheric and tropospheric ozone.

1.3.1 Stratospheric Ozone

Ozone in stratosphere is generally found a few kilometres above the tropopause and its height is, therefore, associated with the altitude of the tropopause itself. Most of the stratospheric ozone production occurs in the equatorial regions, which are exposed to a maximum of solar radiation throughout the whole year. However, the maximum concentrations in total columns[6] are not found at the Equator, but rather at the Poles. This kind of air masses transport is lead by the so-called **Brewer-Dobson Circulation (BDC)** [see e.g. Butchart (2014)]. As

[6]The *total column* (TC) is a hypothetical column extending from the surface to the top of the atmosphere.

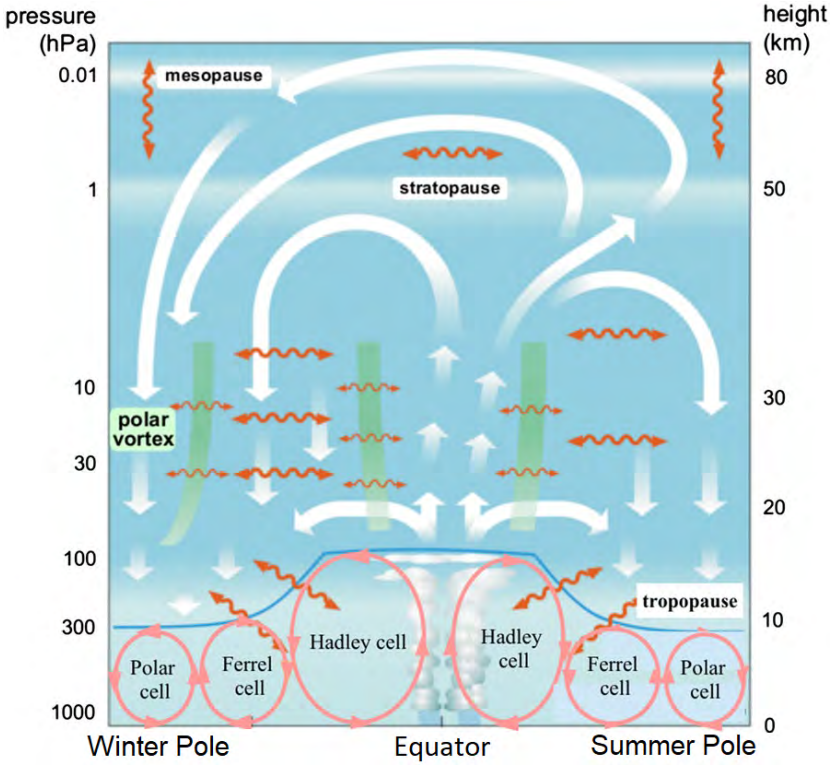


Figure 1.6: Brewer-Dobson Circulation (BDC) scheme [from Proedrou and Hocke (2016)].

schematically outlined in Figure 1.6, the BDC is driven by tropospheric air rising into stratosphere in the tropical regions due to deep convection. This causes momentum propagation in the stratosphere, which, in turn, leads to the generation of waves travelling towards the Poles. These include variations of the Rossby waves [e.g. Rossby (1939); Rossby (1949)], called ***stationary planetary waves***. This overall meridional stratospheric circulation, together with a subsidence phenomenon, causes large ozone-rich air masses to be transported towards the polar regions, progressively reducing their altitude.

The one just described, however, is just the general pattern and significant differences, in strength and behaviour, take place between the two hemispheres. The BDC, indeed, is usually weaker in the Southern Hemisphere due to the different distribution of land masses and the associated differences in the generation of the planetary waves. The reduced amplitude of such waves in Southern winter causes a polar vortex[7] that is often much more isolated and colder compared to the Northern one. As a result the horizontal mixing does not always succeed in reaching the Antarctic region, remaining rather confined in lower latitudes of the Southern Hemisphere. In the Northern Hemisphere, on the other hand, due to the higher intensity of the planetary waves, the horizontal mixing is more easily carried to reach the Pole. The BDC has experienced an acceleration in lower stratosphere, during recent decades, due to the

[7] *Polar vortex* is a large low-pressure area of cold air on top of both North and South Pole, which is held in place by a counter-clockwise flow of air (vortex indeed). It is always present near the poles, but it gains strength during winter and weakens during summer.

steady increase in GhG abundances, which are responsible of the global warming phenomenon [WMO (2022) Executive Report]. This uneven circulation is additionally accentuated by the mesospheric flows. The atmospheric gravity waves, indeed, while propagating upwards, are filtered by the zonal wind in the stratosphere, until part of it reaches the mesosphere [e.g. Kuilman and Karlsson (2018)]. Because of the decreasing density with altitude, and as a result of energy conservation, the waves grow in amplitude, until they become unstable and break. When this happens, the waves deposit their momentum into the background flow, creating a drag on the mesospheric zonal winds, which in turn establishes a pole-to-pole circulation that drives temperature far away from radiative equilibrium [e.g. Holton (1982)]. In the upper mesosphere ($\sim$ 70 to 90 km), temperature increases from the Winter Pole to the Summer Pole, with peaks that are sometimes lower than 130 K in the summer polar mesopause [Lübken et al. (1990)].

The maxima of ozone total columns are found at the end of winter and in spring over the Canadian Arctic and eastern Siberia, while minima occur during autumn in the tropical regions. Figure 1.7 from Fioletov (2008) shows the zonal monthly distribution of ozone as a function of latitudes computed using old data on total ozone columns (from 1964 to 1980).

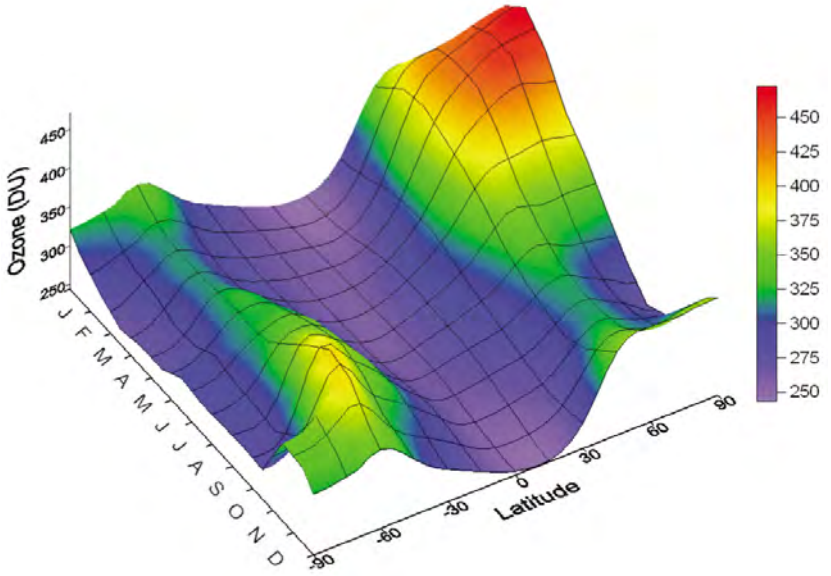


Figure 1.7: Zonal monthly ozone total column as a function of latitude and month estimated from ground-based data for the period from 1964 to 1980 [source Fioletov (2008)].

Although now different in terms of quantity, especially at the Southern-Polar region, as already explained, this chart illustrates the average distribution of ozone with the seasons, which remains qualitatively similar in its cycle even today. Tropical ozone total columns are roughly constant throughout the whole year, while at Mid-latitudes and Poles a clear annual cycle arises. The extra-tropical amount of ozone is, indeed, led by both the transport and photochemical mechanisms just described. The seasonal fluctuations of these two classes of processes create a seasonal cycle, with an increase of total ozone during winter, when ozone transported by atmospheric circulation is accumulated at high latitudes, and a decline during late-spring and summer, when transport weakens and photochemical processes of ozone loss are stronger, due to the stronger incoming solar radiation. The larger abundance of ozone found at high latitudes of Northern Hemisphere during boreal winter,

compared to the amount found in Southern Hemisphere during the austral winter, is due to the different action of the Brewer-Dobson circulation in the two hemispheres already described. On the other hand, the autumn values appear to be comparable in both hemispheres, suggesting that, during this season, the ozone abundance is determined by photochemical processes rather than by the transport.

As already mentioned above, high altitude ozone is involved in shielding harmful UV rays coming from the Sun and generally lying in the 240− 340 nm wavelength range. The mechanism under discussion is a cycle of production and destruction of ozone at high altitudes, exploiting the aforementioned solar radiation for processes of photodissociation. This cycle was described as early as 1930 by Chapman and hence it is named after him **Chapman cycle**. It begins with the photodissociation of O_2 molecules to form single oxygen atoms:

$$O_2 + h\nu \longrightarrow 2O \tag{1.1}$$

where $h\nu$ represents a photon of light with frequency ν. In this phase of the reaction, the wavelength of the radiation involved is $\lambda < 242$ nm. The oxygen atoms thus formed quickly recombine with dioxygen molecules to form ozone:

$$O + O_2 + M \longrightarrow O_3 + M + heat \tag{1.2}$$

The reaction takes place with the help of another molecule or atom, usually N_2 or O_2, and here denoted by M, acting as a catalyst for the reaction itself, so that energy and momentum are conserved. The ozone production at this stage also releases heat (about 24 kcal for each mole formed).

The dissociation of ozone, for Chapman, can take place in two ways. One involves the ozone molecule photodissociation by radiations of wavelength $240 < \lambda < 310$ nm:

$$O_3 + h\nu \longrightarrow O_2 + O \tag{1.3}$$

The second one is by collision of atomic oxygen and ozone to form molecular oxygen all over again:

$$O + O_3 \longrightarrow 2O_2 \tag{1.4}$$

Inter-conversion of O atoms and O_3 molecules is often so rapid in most atmospheric regions, that they can be approached as one single entity called ***odd oxygen***, denoted by $\mathbf{O_x}$.

The only reactions just described, however, are not sufficient to explain the actual amount of ozone detected in stratosphere. The Chapman cycle on itself should produce an ozone layer far larger than the one actually existing. Other chemical reactions, indeed, take place involving the combination of molecular oxygen and ozone, and catalysed by specific radicals. These are, in part, naturally present in the stratosphere, such as the hydroxyl radical (OH) or the nitric oxide (NO). On the other hand, other catalysts, such as the atoms of chlorine (Cl) and bromine

(Br) mainly come from human activities. The emissions of chlorofluorocarbons (CFCs) and halons starting from the 1950s, has enormously increased the Cl and Br concentrations, playing a key role in the thinning of the ozone layer in the last century. The typical reaction taking place, using as a catalyst the chlorine, is the following:

$$\begin{array}{r} \mathrm{Cl} + \mathrm{O_3} \longrightarrow \mathrm{ClO} + \mathrm{O_2} \\ \mathrm{ClO} + \mathrm{O} \longrightarrow \mathrm{Cl} + \mathrm{O_2} \\ \hline \mathrm{O} + \mathrm{O_3} \longrightarrow 2\mathrm{O_2} \end{array} \tag{1.5}$$

The catalysts are produced again at the end of each cycle. However, some of them may transform in species less reactive, reducing their power to catalyse ozone destruction.

Signs of ozone hole observations over the Antarctic region date back to the 1970s. This phenomenon, in addition to being facilitated by the Brewer-Dobson circulation being weaker in the southern hemisphere (as explained above), was triggered by the kind of catalytic reaction just introduced. Each atom of Cl and Br, indeed, can catalyse tens of thousands reactions removing ozone from the stratosphere. In the polar regions also other additional reactions occur in late-winter and early-spring. The process is triggered by reactions taking place at the surface of the polar stratospheric clouds (PSCs), which contain ice crystals that can turn non-reactive compounds into reactive ones:

$$\begin{aligned} \mathrm{ClNO_2} + \mathrm{H_2O} &\xrightarrow{\text{PSCs}} \mathrm{HNO_3} + \mathrm{HOCl} \\ \mathrm{ClNO_2} + \mathrm{HCl} &\xrightarrow{\text{PSCs}} \mathrm{HNO_3} + \mathrm{Cl_2} \end{aligned} \tag{1.6}$$

In spring, as soon as solar light becomes available, HOCl and $\mathrm{Cl_2}$ are photodissociated:

$$\begin{aligned} \mathrm{Cl_2} + h\nu &\longrightarrow 2\,\mathrm{Cl} \\ \mathrm{HOCl} + h\nu &\longrightarrow \mathrm{ClO} + \mathrm{H} \\ \mathrm{HOCl} + h\nu &\longrightarrow \mathrm{ClO} + \mathrm{OH} \end{aligned} \tag{1.7}$$

It leads to an abundance of ClO that catalyses the ozone transformation into oxygen. Typical reactions are:

$$\begin{array}{r} \mathrm{ClO} + \mathrm{ClO} \longrightarrow (\mathrm{ClO})_2 \\ (\mathrm{ClO})_2 + \mathrm{h}\nu \longrightarrow \mathrm{Cl} + \mathrm{ClO_2} \\ \mathrm{ClO_2} \longrightarrow \mathrm{Cl} + \mathrm{O_2} \\ 2(\mathrm{Cl} + \mathrm{O_3} \longrightarrow \mathrm{ClO} + \mathrm{O_2}) \\ \hline 2\mathrm{O_3} \longrightarrow 3\mathrm{O_2} \end{array} \tag{1.8}$$

$$\begin{array}{r} \mathrm{ClO} + \mathrm{BrO} \longrightarrow \mathrm{Cl} + \mathrm{Br} + \mathrm{O_2} \\ \mathrm{Cl} + \mathrm{O_3} \longrightarrow \mathrm{ClO} + \mathrm{O_2} \\ \mathrm{Br} + \mathrm{O_3} \longrightarrow \mathrm{BrO} + \mathrm{O_2} \\ \hline 2\mathrm{O_3} \longrightarrow 3\mathrm{O_2} \end{array} \tag{1.9}$$

Human activities have thus strongly perturbed the natural equilibrium of the stratospheric ozone cycle. Cl and Br are responsible of the $70 - 90\%$ ozone loss as a function of the

stratosphere height. The larger the reduction of the ozone layer, the weaker the shielding action it performs. Several steps were then taken towards the end of the last century in order to reduce this phenomenon. The implementation of the Montreal Protocol, one of the rare treaties to achieve universal ratification, created a timetable for controlling production and consumption of CFCs first, and to totally ban the release of CFCs and related molecules at a later stage. The hydrofluorocarbons (HFCs) were, therefore, introduced as non-ozone depleting alternatives.

The application of this protocol led to the stabilisation of the total atmospheric chlorine already at the beginning of the 1990s. The total amount of both chlorine and bromine is, at present, continuing to decline [WMO (2022) Executive Report]. Finally, in the last couple of years, the ozone hole is showing its first signs of reduction. Measurements show increases in upper-stratospheric ozone between 2000 and 2020, due to a combination of decreases of the ozone-depleting substances emitted and the decrease of stratospheric temperature caused by the augmentation of CO_2 concentrations. Measurement made via satellite and ozonesondes (see Chapter 5) have shown an Antarctic ozone hole smaller in recent years than it was in the last two decades. In particular, Arctic total ozone reached exceptionally low values in spring 2020 [WMO (2022) Executive Report].

1.3.2 Tropospheric Ozone

While stratospheric ozone plays an important role in shielding UV radiation and thus in supporting life at the surface, tropospheric ozone is not as beneficial. Ozone in troposphere is an important greenhouse gas and a secondary air pollutant. It has a strong oxidising capacity and it is harmful to human and ecosystem health. It is a major component of urban smog.

An important aspect that arises when monitoring air quality is the increasing tropospheric ozone concentration over the last century. This is due to the increased emissions of nitrogen oxides (NO_x) and hydrocarbons released by human activities [e.g. Finlayson-Pitts and Pitts Jr (1997)]. Chemical reactions initiated by photodissociation (with $\lambda > 290\,\text{nm}$) of these compounds, generate a large variety of species, such as nitric acid, known as ***photochemical oxidants***. Evolution of each of these pollutants is strictly interconnected with the day and night cycles, that, under favorable meteorological conditions, are both activated. This can produce a rapid accumulation of ozone and oxidants in the atmosphere run by a series of reactions having radical OH as driving force during the day and the NO_3 at night-time.

Photolysis of NO_2 (through radiations with $\lambda < 420\,\text{nm}$) produces nitrogen monoxide (NO) and atomic oxygen:

$$NO_2 + h\nu \longrightarrow NO + O \tag{1.10}$$

The atomic oxygen recombines with an oxygen molecule, a reaction catalysed by a third molecule capable of absorbing the excess vibrational energy (denoted as above with M):

$$O + O_2 + M \longrightarrow O_3 + M \tag{1.11}$$

However, the ozone thus formed reacts with the NO, to form again oxygen and NO_2.

$$NO + O_3 \longrightarrow NO_2 + O_2 \tag{1.12}$$

These reactions creates a daytime cycle (**Leighton's cycle**) that is null. In other words, there is no net production or loss of any species involved in the cycle itself. During nighttime, however, as the reaction 1.10 does not take place due to the lack of solar energy ($h\nu$), ozone production is inhibited.

At the same time, other processes take also place, making things more complex than just explained. The VOCs and nitrous oxides injection in atmosphere from vehicles, power plants, industrial processes, landfills and other biomass and fossil fuel burning, triggers a complex photochemical cycle of ozone production. Hydroxyl radical can oxidise hydrocarbons into species that constantly convert NO into NO_2. In this case, NO is not available to carry out the ozone sink process and the NO_2 formed can be split again under the influence of UV radiation, further producing ozone. First step consists in oxidation of molecule of hydrocarbons (RH) or of carbon monoxide (CO):

$$RH + OH \xrightarrow{O_2} RO_2 + H_2O \tag{1.13}$$

$$CO + OH \xrightarrow{O_2} HO_2 + CO_2 \tag{1.14}$$

Radicals RO_2 and HO_2 reacts with NO to form NO_2:

$$\begin{aligned} RO_2 + NO &\xrightarrow{O_2} R'CHO + HO_2 + NO_2 \\ HO_2 + NO &\xrightarrow{O_2} OH + NO_2 \end{aligned} \tag{1.15}$$

where R′CHO is an intermediary organic compound. The NO_2 made available, participates in the creation of ozone through reactions as in Equations 1.10 and 1.11, without, however, going through the conversion of O_3 to O_2 by NO (Equation 1.12).

Ozone is also destroyed by photolysis in presence of water to give two OH:

$$O_3 + h\nu \xrightarrow{H_2O} 2OH \tag{1.16}$$

R′CHO is also photodissociated to form CO and peroxy radicals (RO_2)

$$R'CHO + h\nu \xrightarrow{H_2O} RO_2 + CO \tag{1.17}$$

The radicals RO_2 and HO_2 are then, destroyed through the reactions:

$$\begin{aligned} HO_2 + HO_2 &\longrightarrow H_2O_2 + O_2 \\ HO_2 + RO_2 &\longrightarrow ROOH + O_2 \\ OH + NO_2 &\longrightarrow HNO_3 \end{aligned} \tag{1.18}$$

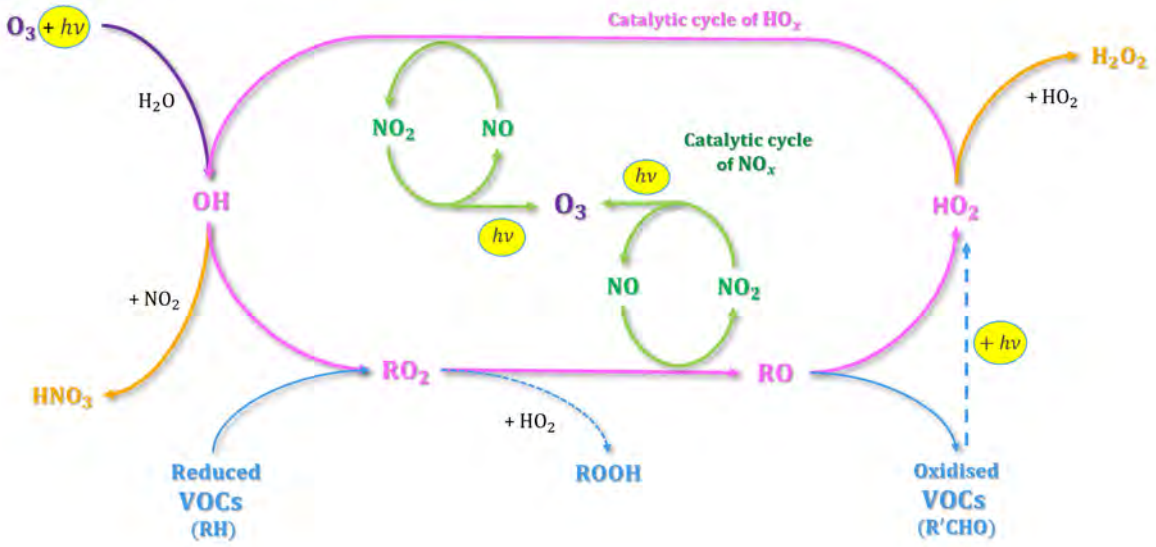


Figure 1.8: Photochemical cycle of ozone production in troposphere. Please notice that several reactions are not depicted in this diagram (e.g. those involving CO) [chart inspired by Aumont (2005) and Chailleux (2018)].

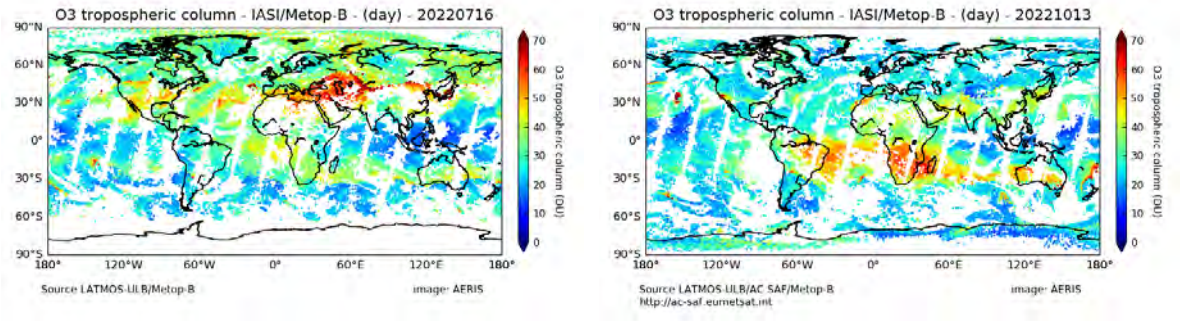
to form hydrogene peroxide (H_2O_2), hydroperoxides (ROOH) and nitric acid (HNO_3). Figure 1.8 summarises the photochemical cycle of ozone in troposphere just illustrated (for clarity some reactions, such as those involving CO, are not depicted in the chart).

Depending on the conditions that support these reactions of production and destruction, ozone can present an atmospheric lifetime of hours to weeks.

Figure1.9 shows an example of ozone values obtained from the processing of satellite data acquired by the IASI instrument (which will be introduced in detail in Chapter 5), and expressed in DU[8]. Each panel refers to a day of 2022 selected as representative of typical tropospheric ozone distributions in the two hemispheres depending on the season. Just as shown in Panel 1.9a, Northern Hemisphere usually experiences maxima in ozone concentrations during spring and summer due to greater amount of incident solar radiation and to the anthropogenic emissions of ozone precursors. These two factors result in an intense ozone photochemical production. The highest concentration of tropospheric ozone in the Southern Hemisphere (as in Panel 1.9b), on the other hand, is observed in the Southern spring as a consequence of the frequent biomass burning occurring in the inter-tropical band during the dry season. The lifetime of tropospheric ozone allows it to be transported by atmospheric dynamics. Ozone levels depend on background conditions as well as emissions of its precursors.

The main strategies implemented to prevent the formation of tropospheric ozone are primarily based on cutting the levels of atmospheric pollutants coming from man-made sources. Strict NO_x emission limits have been set for power plants and industrial combustion sources,

[8]DU (Dobson Unit) is the most common unit for measuring the total amount of ozone present in a vertical column of air above the surface of the Earth. 1 DU is the number of ozone molecules that would be required to have a layer of pure ozone of 0.01 mm thickness at a temperature of 0 °C and a pressure of 1013.25 hPa.



(a) 16th of July, 2022 **(b)** 13th of October, 2022

Figure 1.9: IASI level 2 (L2) tropospheric ozone products [from AERIS web portal `https://iasi.aeris-data.fr/`].

as well as limitations on solvents usage in factories. Cleaner-burning fuels a strong vehicle inspection programs can also help control ozone pollution. Many countries implement air quality standards to preserve population health. However, these regulations often exceed the World Health Organization (WHO)[9] guidelines established on outdoor (ambient) air pollution levels.

1.4 Carbon Monoxide

Carbon monoxide is a trace gas present in very small and variable concentration in atmosphere (Table 1.1), and, despite that, still affecting the chemical processes taking place in it. It results from the incomplete combustion of carbon-containing fuels. These could be natural gas, gasoline, or wood. As a consequence, it is often emitted by a series of combustion sources, such as motor vehicles, power plants, incinerators and wildfires. The largest sink for carbon monoxide is the reaction with OH, while, in a smaller amount, it is removed from atmosphere by deposition on the ground.

As already mentioned, its presence in atmosphere is quite variable. Global background concentration of CO ranges between approximately 0.05 and 0.12 ppm. However, in areas where severe smog episodes occur, it can reach average values of 17 ppm, with short peaks below 53 ppm. Its lifetime ranges from two to four months in the atmosphere [Wark and Warner (1981)], making it an excellent tracer of hydrocarbon sources.

An example of carbon monoxide variability over the globe is provided in Figure 1.10 derived from by IASI data and referring to an arbitrary day in the month of June 2022. The strongest concentrations are found over central Africa, most likely due to biomass burning, and over Asia and the North-East coast of USA, where urban centres, characterised by high levels of pollution, are located.

Carbon monoxide has no greenhouse effect on its own. However, its presence affects the abundance of major GhG. The CO, indeed, reacting with the OH, reduces the abundance of

[9] `https://www.who.int`

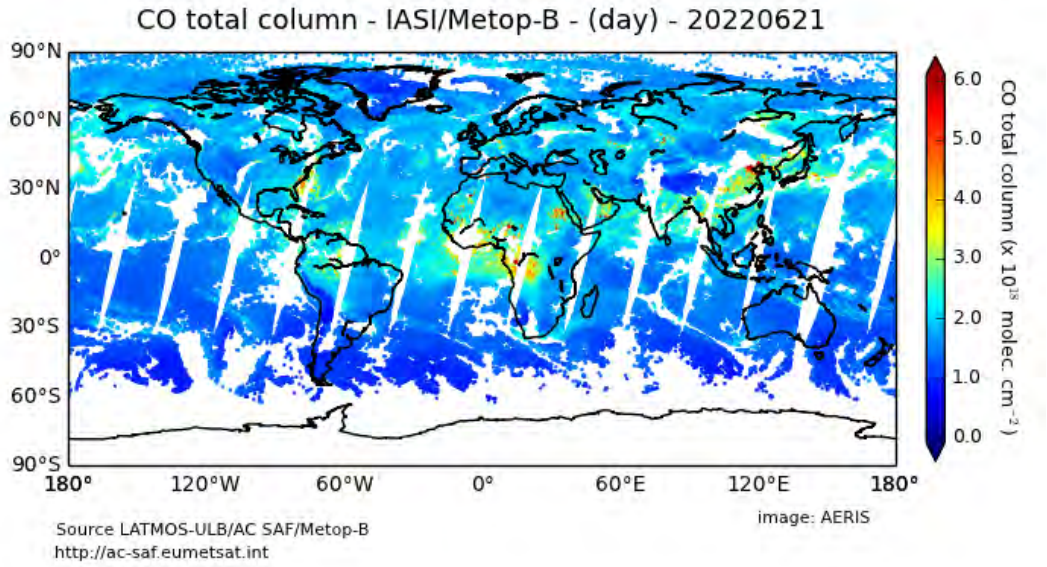


Figure 1.10: IASI level 2 (L2) carbon monoxide product for the day 21st of June, 2022 [from AERIS web portal `https://iasi.aeris-data.fr/`].

this radical. Nevertheless, OH radicals contribute to clean-up atmosphere from strong greenhouse gases, such as methane and carbon dioxide (as already mentioned). Consequently, the CO indirectly increases the global warming potential of these gases. As previously introduced in this Chapter, on the other hand, the CO can also lead to tropospheric ozone production.

1.5 Highlights of this Chapter

- The Earth's atmosphere plays a key role on the unfolding of life of the planet. It is stratified into layers, depending on the temperature profile trend, such as the *troposphere*, *stratosphere*, *mesosphere*, *thermosphere* and *exosphere*.
- The atmosphere's chemical composition varies with altitude (temperature and pressure), and depending on the anthropogenic and natural emission sources.
- **Ozone** is a trace gas in the Earth's atmosphere and it is a secondary species, i.e. it not emitted, but rather it forms through chemical reactions. Most of the atmospheric ozone (90%) is found in stratosphere, while a smaller percentage is in troposphere.
  - Because of the chemical reaction it is involved in (processes involving oxygen, ozone, and specific radicals), the ***stratospheric ozone*** has the power to shield UV harmful radiation coming from space. Its production occurs mainly in the equatorial regions and it is transported towards the poles through the Brewer-Dobson Circulation.

    Human activities, such as the release of chlorine and bromine compounds, have significantly disrupted the stratospheric ozone cycle and caused ozone depletion. The Montreal Protocol and other measures, however, have been implemented to reduce ozone-depleting substances.
  - ***Tropospheric ozone*** is a greenhouse gas and a harmful air pollutant generated by human activities.

    Increased emissions of nitrogen oxides and hydrocarbons contribute to the rise in tropospheric ozone concentrations.

    Chemical reactions driven by sunlight lead to the formation of photochemical oxidants, including ozone, in the troposphere.
- **Carbon monoxide** is a trace gas found in the atmosphere due to incomplete combustion of carbon-containing fuels, emitted by sources like vehicles, power plants, and wildfires. Although it has a weak direct greenhouse effect, it indirectly affects the abundance of other GhGs. Additionally, carbon monoxide can contribute to the production of tropospheric ozone.

Chapter 2

Radiative Transfer

In this Chapter, the fundamental physical laws governing the interaction of the electromagnetic (EM) radiation with the atmosphere and the Earth's radiative budget will be described. The fundamentals of radiative transfer are, indeed, necessary to understand the new generation of measuring instruments and the modelling and simulation of certain atmospheric fields. Finally, the fast Radiative Transfer Model (RTM) RTTOV, which is the one exploited for the present work of research, will be presented.

2.1 Electromagnetic Spectrum

The main source of energy for planet Earth is the solar radiation[1]. This travels in the form of EM waves at the speed of light $c = 2.99793 \cdot 10^8$ m/s, with a wavelength λ (expressed in m) and a frequency ν (in s^{-1}), linked by the relation:

$$c = \lambda\nu \tag{2.1}$$

Solar radiation extends over the entire electromagnetic spectrum, from very short gamma rays (γ-rays) up to very long x-rays. Figure 2.1 shows a layout of each band in the EM spectrum as a function of wavelength and frequency. The bands that are of most interest to what has been reported so far and what will be presented later in this study range from UV to the complete IR band. For this reason, more details on their further classification into sub-bands are provided in Table 2.1.

Some regions of the spectrum present wavelengths that are able to pass through the atmosphere (e.g. VIS light or some MW[2] and IR) and are referred to as ***atmospheric windows***. On the other hand, many other wavelengths interact with the atmospheric matter along their path through the gaseous layers. Thanks to some of these interactions, as already mentioned in Chapter 1, Earth's atmosphere is able to shield a range of higher energy waves

[1] the outward conduction of heat from the inside of Earth due to radiative decay is completely negligible.
[2] microwave (MW)

that can be harmful to life. Moreover, this mechanism of interaction between radiation and particles defines Earth's radiative balance and it is the basis for most of the remote sensing techniques used in meteorology.

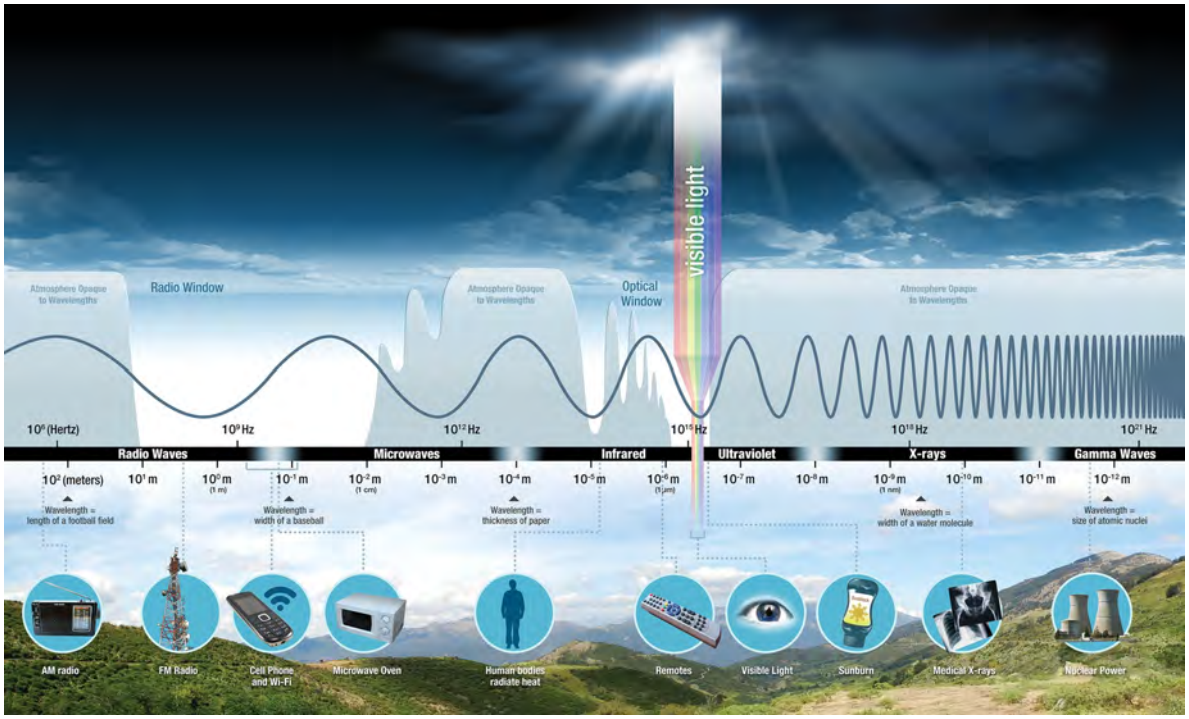


Figure 2.1: Schematic diagram illustrating the electromagnetic spectrum through frequency (in Hz) and wavelength (in m) [adapted from NASA-Science website (`https://science.nasa.gov/ems/01_intro`)].

Electromagnetic radiation band	Sub-bands	
Ultraviolet (UV) 100-400 nm	UVA 100-280 nm	
	UVB 280-315 nm	
	UVC 315-400 nm	
Visible (VIS) 400-700 nm		
Infrared (IR) 780 nm – 1mm	Near-Infrared (NIR) 780 nm – 1.4 μm	Thermal Infrared
	Short Wavelength Infrared (SWIR) 1.4 – 3 μm	
	Medium Wavelength Infrared (MWIR) 3 – 8 μm	
	Long Wavelength Infrared (LWIR) 8 – 15 μm	
	Far-Infrared (FIR) 50 μm – 1mm	

Table 2.1: Limits of the Ultraviolet (UV), Visible (VIS) and Infrared (IR) bands, plus nomenclature and boundaries of their sub-bands.

2.2 Interaction of Radiation with Matter

In order to understand how the atmosphere behaves towards EM radiation, one needs to understand the mechanisms of interaction between light and atmospheric particles, whether solid, liquid or gaseous.

Several interaction processes can take place when radiation encounters a particle, or in general a medium. The main typologies of interaction are **transmission**, **diffusion** (or **scattering**) and **absorption** (with consequent re-emission). The one that takes place depends on a few factors such as the properties of the wave (i.e. wavelength) and the medium properties of the two materials (the one in which the wave travels and the one on which the wave impacts).

In real scenarios, however, these effects often occur simultaneously. Furthermore, the conservation of energy requires an equilibrium among them. The balance of electromagnetic energy hitting a material should be equal to the energy that is transmitted, absorbed, and scattered, according to the equation:

$$\alpha_\lambda + \beta_\lambda + \tau_\lambda = 1 \tag{2.2}$$

where α_λ is the coefficient of absorption, β_λ of diffusion and τ_λ of transmission.

The main processes of interaction between radiation and matter, that are going to be illustrated in more detail just below, are schematically sketched in Figure 2.2.

2.2.1 Transmission

Transmission is the transit of electromagnetic radiation through a medium. If an object is transparent, that is because the light passes through it.

In real life transmission rarely leaves a ray passing through two different media maintaining the same speed and direction it had originally. In fact, when a ray passes from one medium to another it experiences the different properties of the media themselves, such as their density. In other words, when the ray passes from a less dense medium to a denser one, it slows down, consequently changing its original direction. When, vice versa, it passes from a denser medium to a less dense one, it speeds up. In both cases the ray will experience a change in speed and, consequently, also a change in its original direction. This phenomenon is called *refraction*.

Please note also that often materials that are transparent to a specific frequency are opaque to others.

2.2.2 Scattering

Reflection occurs when the incident wave bounces off the boundary surface between the two media (the one in which the wave travels and the one on which the wave impact) with an angle (*reflection angle*) that is always equal to the angle of incidence. Some materials are perfectly smooth so the boundary between the two media is completely flat. Therefore all incoming

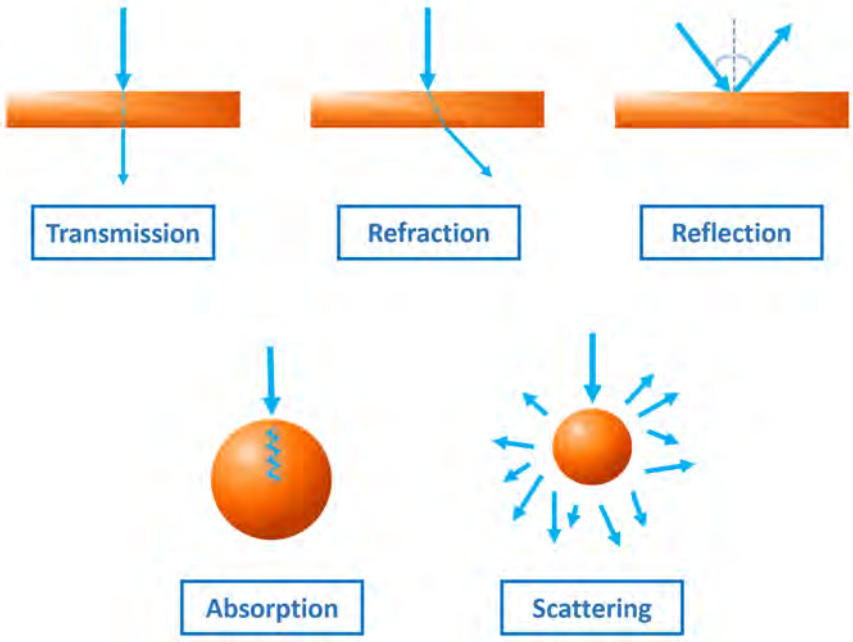


Figure 2.2: Diagram outlining the main processes of interaction between radiation and matter.

light rays will be reflected in the same direction (*specular reflection*). If, on the other hand, the boundary between the two media is not perfectly smooth, even if the incoming rays come all down in the same direction, they will be reflected according to their points of incidence. Then, this kind of reflection (*scattered reflection*) will not present regular and determined trajectories, but rather disorganised and largely random.

In the atmospheric medium, scattering occurs when particles, or large gas molecules, intercepts electromagnetic radiation causing it to be redirected in all direction without changing its original wavelength.

The kind of scattering taking place depends on several factors including the wavelength of the radiation, the abundance of particles, and the distance the radiation crosses through the atmosphere. There are three main types of scattering that can occur: Rayleigh, Mie and non-selective scattering.

Rayleigh scattering occurs when particles size is smaller compared to the wavelength of the incident radiation. This causes shorter wavelengths to be diffused much more than longer wavelengths. Rayleigh scattering is the process for which human eye perceives the diurnal sky in a blue colour. As the sunlight passes through the atmospheric layers, the shorter wavelengths of the VIS spectrum (i.e. blue) are scattered more than the other (longer) visible wavelengths. At sunrise and sunset the light has to travel farther through the atmosphere than during central hours of the day, giving the chance also to the shorter wavelengths to be scattered and to human eye to see also the warmer colors red and orange.

Mie scattering occurs when the particles are just about the same size as the wavelength

of the radiation. Dust, pollen, smoke and water vapour are common causes of Mie scattering which tends to affect longer wavelengths than those affected by Rayleigh scattering. Mie scattering occurs mostly in the lower portions of the atmosphere where larger particles are more abundant, and dominates when cloud conditions are overcast.

Non-selective scattering takes place when the particles are much larger than the wavelength of the radiation. This kind of scattering takes its name from the fact that all wavelengths are scattered about equally. Water droplets and large dust particles can cause this type of scattering. Fog and clouds appear white to our eyes because blue, green, and red light are all scattered in approximately equal quantities.

2.2.3 Absorption and Emission

Absorption is the mechanism at work when the EM radiation interacts with the atmosphere. It consists in the transformation of radiant power to another type of energy, usually heat, by interaction with matter.

To explain the phenomenon of absorption, the concept of “energy level" must be introduced. *Electron shells*, or energy levels indeed, are fixed distances from the atomic nucleus where the electrons can be found. Planck hypothesised that the energy levels of an atomic or molecular oscillator were limited to a discrete set of values meeting the relation $E_\nu = \mathrm{n}h\nu$, $\mathrm{n} = 0, 1, 2, \ldots$. This equation describes a set of energetic levels differing from each other by $h\nu$. The oscillator can represent the periodic motion of an atom or molecule and the transition from one energy level to another, called *quantum leap*, corresponds to the release or absorption of an amount of energy $h\nu$. This quantum shift corresponds to the absorption or emission of a photon of energy exactly equal to $h\nu$.

Let us consider a photon emitted by a given source that travels through space. If the photon encounters a mass, such as a particle or a molecule, then it may change phase or direction through a scattering process, or it may be absorbed. In this latter case, its energy is transferred to the substance that absorbed it. This energy may appear in the form of an increase in the internal energy of the molecule or atom, or as heat.

Since each gas in the atmosphere absorbs EM radiation in very specific regions of the spectrum (i.e. each gas absorbs and is thus sensitive to specific wavelengths), absorption is the basis of spectroscopy and remote sensing observations (see farther in the manuscript Chapter 5).

The thermal process of emission takes place, as the opposite of absorption, when electrons in atoms make a transition from a higher energy state to a lower energy state, thus releasing energy. Emission is proportional to absorption.

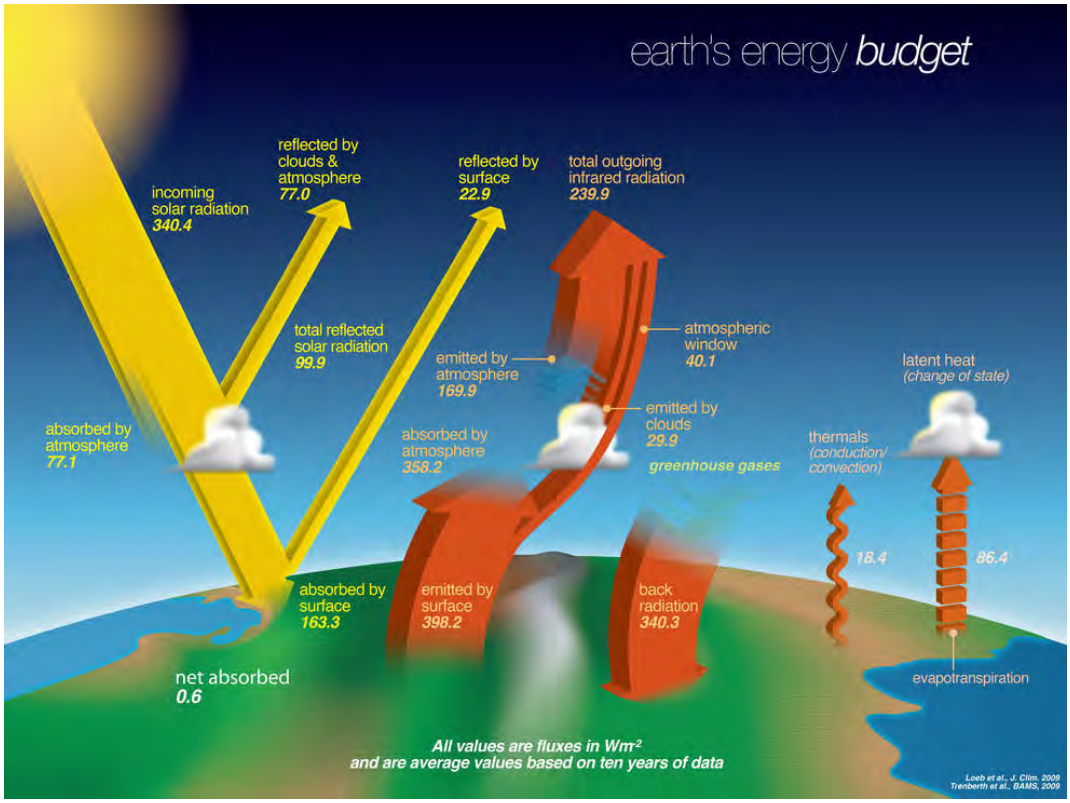


Figure 2.3: Annual global radiation balance for incident and emitted radiation in W/m^2 [credits: NASA].

2.3 Earth's Radiative Balance

For the processes of interaction between atmospheric particles and radiation just described, incoming solar radiation passing through the atmosphere is partially absorbed, partly scattered and reflected, partially transmitted by atmospheric components (gases, aerosols and clouds). The overall balance of these processes is schematically sketched in Figure 2.3. First of all, one can remark that, of the total $\sim 340\,\mathrm{W/m^2}$ of solar radiation arriving at the top of the atmosphere (TOA), 30% is reflected back by the atmosphere and surface itself (*albedo effect*). Remaining 70% is in part absorbed by gaseous components, while in a larger part it reaches the Earth's surface and is absorbed by this. The absorption occurring in atmosphere is a selective one, and it is performed by the permanent constituents of the atmosphere, such as O_2, CO_2 and water vapour, and to a lesser extent by minor pollutants. In addition to absorption, a fraction of the radiant energy passing through the atmosphere is diffused.

In order to maintain the system in a *quasi-equilibrium* state over long periods of time, the energy absorbed by the Earth must be balanced by an equal amount of energy emitted towards space as radiation from the surface and atmosphere. Any body with a temperature larger than absolute zero, in fact, emits energy in the form of radiation distributed over a wide range of wavelengths in the electromagnetic spectrum.

Most of the solar energy of interest to the climate system lies between 0.1 and 2.0 µm, thus belonging to the UV, VIS and near infrared (NIR) bands. On the other hand, most of the energy outgoing to space is located between 4.0 and 60 µm, i.e. fully in the IR. It must be said that most of the wavelengths reaching the surface belong to the VIS. The atmosphere is indeed opaque to most UV and IR rays. Ultraviolet waves should be absorbed in their near totality by the ozonosphere around 25 km altitude, thanks to the cycle of creation and destruction of ozone described in Chapter 1 (1.3). The situation is more complex when dealing with infrared radiation. Unlike UV, which originates from a single source (the Sun), IR wavelengths are also emitted by the Earth and the atmosphere itself. This causes the IR radiation to come from any direction within the atmosphere.

In order to explain the energy exchange mechanisms of absorption and re-emission taking place in the Earth system, the black-body radiation theory is used, which can easily be adapted to any type of body that absorbs or emits radiation.

2.3.1 Black-body Radiation and Planck's Law

Whenever a body has a temperature exceeding the absolute zero ($T > 0\,\mathrm{K}$), it emits EM energy. In this case, in fact, the particles of matter contain excited quantum states. The decay of these states results in the emission of energy.

In order to assess the radiative energy emitted by a generic body, one can start by considering an ideal ***black-body***, that is a perfect absorber or a body that absorbs all incident radiation. It was Planck, in the early 1900s, who established that the amount and type of energy emitted by a black-body is determined uniquely by its temperature T. The mathematical relation he derived, and named after him *Planck's Law*, allows to determine the spectral energy distribution of radiant energy, or ***spectral radiance***, emitted by a black-body:

$$B_\lambda = \frac{2hc^2}{\lambda^5} \frac{1}{e^{\frac{hc}{\lambda k_B T}} - 1} \tag{2.3}$$

or, using Eq. 2.1, as a function of frequency:

$$B_\nu = \frac{2h}{c^2} \frac{\nu^3}{e^{\frac{h\nu}{k_B T}} - 1} \tag{2.4}$$

where $h = 6.62607 \cdot 10^{-34}$ J s is the *Planck constant* and $k_B = 1.38154 \cdot 10^{-23}$ J K^{-1} the *Boltzmann constant*. The official unit for the spectral radiance is W/(m^{2}µm sr). The radiant flux emitted, reflected, transmitted or received by a surface, per unit of solid angle and projected area, instead, is called ***radiance*** and expressed in W/(m^2 sr). In meteorology the radiance is often evaluated per unit of frequency, wavenumber or wavelength. When working with IR wavelengths, in particular, the radiance is usually evaluated per unit of wavenumber (in cm^{-1}) and then expressed in W/(m^2 sr cm^{-1}).

Black-body radiation is isotropic, i.e. it is emitted uniformly in all directions. It follows

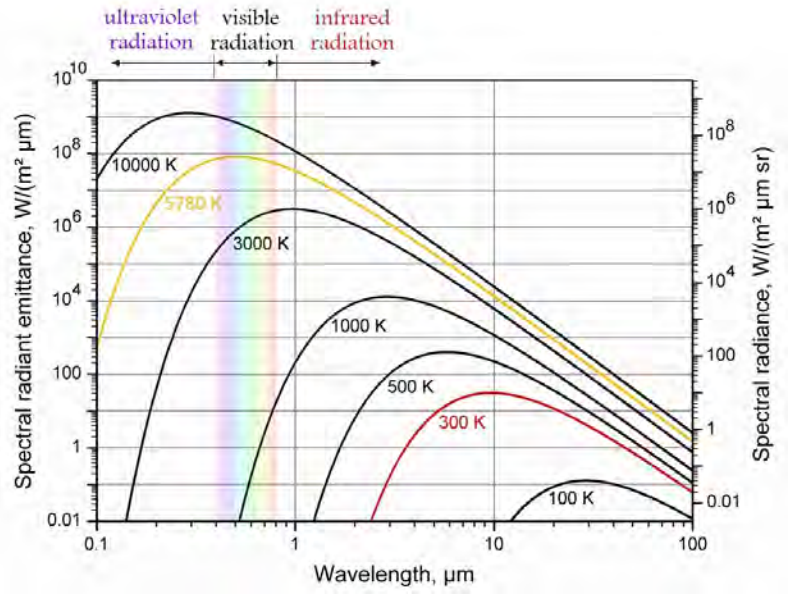


Figure 2.4: Representation of the Planck function for different black-body temperatures [adapted from EUMETRAIN website[a]]. The curves are represented in a scale of spectral radiance (right y-axis) and of spectral radiance emittance (or radiant flux per unit - left y-axis). Two functions are highlighted in colors: the orange one corresponding to the approximate maximum temperature of solar emission (∼5780 K) and in red the one associated to the maximum temperature of Earth surface and law atmosphere (∼300 K).

[a] `https://resources.eumetrain.org/data/3/30/navmenu.php?tab=5&page=1.2.0`

that the intensity is independent of direction.

Thanks to this theory, we can derive the temperature of a body simply by measuring its emitted radiation.

2.3.2 Wien's Displacement Law

The Planck's function is a very regular function with a single maximum. When evaluating its peak at different temperatures (see Figure 2.4), one can observe a shift towards shorter wavelengths as the temperature increases. *Wien's Law* establishes this relationship between the temperature of a black-body and the wavelength corresponding to the maximum energy emission (λ_{max}). According to Wien, the product between the temperature and the wavelength of maximum emission is a constant value $b = 2897.8\,\mu\text{m K}$ (*Wien's constant*):

$$\lambda_{max}\, T = b \tag{2.5}$$

Starting from this relation, a couple of considerations can be made. The radiative temperature of the Sun's surface is approximately 5780 K. Applying Equation 2.5, one can find that the maximum of B_λ occurs at 0.50 µm, i.e. in the visible, as also shown in Figure 2.4. On the other hand, the average temperature of the Earth's atmosphere is approximately 300 K. Consequently, the B_λ maximum for the emission spectrum of the Earth's atmosphere lies around 11 µm, i.e. in the thermal infrared. This explains why in the atmosphere the main

radiation that determines the temperature itself are infrared, visible and ultraviolet.

2.3.3 Stefan-Boltzmann Law

Knowing that the black-body emission is isotropic, the hemispheric flux can be computed as:

$$F_\lambda = \pi B_\lambda. \tag{2.6}$$

The total hemispheric flux out-flowing from a black-body surface, or ***emittance***, can thus be obtained as:

$$M(T) = \pi \int_0^\infty B_\lambda d\lambda \tag{2.7}$$

that, by replacing B_λ with Eq. 2.3 and using $x = h\nu/k_B T$ becomes:

$$M(T) = \frac{2\pi(k_B T)^4}{h^3 c^2} \int_0^\infty \frac{x^3 dx}{e^x - 1} = \left(\frac{2\pi^5 k_B^4}{15 h^3 c^2}\right) T^4 \equiv \sigma_B T^4 \tag{2.8}$$

where $\sigma_B = 5.6703 \cdot 10^{-8}$ W/m^{2}K^4 *Stefan-Boltzmann constant*. The total flux radiating from a black-body surface is proportional to the fourth power of its temperature.

2.3.4 Wien's and Rayleigh-Jeans's Approximations

In the high-frequency domain, i.e. short wavelengths, the exponential in Equation 2.3 is $e^{\frac{hc}{\lambda k_B T}} \gg 1$. In this case Planck's law can be described by the *Wien's approximation*:

$$B_\lambda = \frac{2hc^2}{\lambda^5} e^{\frac{hc}{\lambda k_B T}}, \qquad \lambda \ll \frac{hc}{k_B T} \tag{2.9}$$

which is not linear in temperature.

An approximation can be done also for low-frequencies, i.e. long wavelengths, or for high temperatures. In this case, the term in the exponential becomes small, and the exponential is well approximated with the Taylor polynomial's first-order term $e^{\frac{hc}{\lambda k_B T}} \approx 1 + \frac{hc}{\lambda k_B T}$. Equation 2.3 becomes then:

$$B_\lambda = \frac{2ck_B T}{\lambda^4}, \qquad \lambda T \gg \frac{hc}{k_B} \tag{2.10}$$

or *Rayleigh-Jeans's approximation*.

2.3.5 Kirchhoff's Law

The Kirchhoff's Law states that, at the thermodynamic equilibrium (TE), the radiation absorbed by a black-body is equal to the radiation re-emitted. Consequently, the spectral volumetric factor of emission is equal to the one of absorption:

$$\varepsilon_\lambda = \alpha_\lambda \tag{2.11}$$

Since a black-body is for definition a perfect absorber, $\varepsilon_\lambda = 1$. Consequently $\alpha_\lambda = 1$, while β_λ and τ_λ are equal to zero.

2.3.6 Grey-body

So far, all that have been said concerns a perfect representation of an ideal absorber, i.e. the black-body. In non-idealised conditions, however, a body is never just absorbing radiation, but also reflecting, scattering and transmitting it. We can, in this case, talk about ***grey-bodies***. They only absorb a part of the incident radiation and transmit or scatter the remaining part. At the same temperature, therefore, a grey-body emits less than a black-body under the same conditions.

The ratio of a grey-body to a black-body thermal radiation at the same temperature is called the ***emissivity*** (ε_λ) of the grey-body ($0 < \varepsilon_\lambda < 1$). A grey-body though, such as the Earth's surface, emits a spectral radiance (I) that can be considered as the radiance of a black-body weighted by the spectral surface emissivity (ε_λ):

$$I_\lambda = \varepsilon_\lambda B_\lambda \tag{2.12}$$

or, as a function of the frequency ν:

$$I_\nu = \varepsilon_\nu B_\nu \tag{2.13}$$

As already mentioned, the emissivity of a black-body is equal to 1.

The TE condition is very constraining. In atmosphere, however, it is often locally meet [local thermodynamic equilibrium (LTE)]. Kirchhoff's Law is then locally valid. Therefore, if on one hand the Earth's surface behaves as a grey-body, the atmosphere is locally comparable to a black-body.

2.3.7 Brightness Temperature

Equation 2.10 shows how, for longwave radiation, the spectral radiance of a body is directly proportional to its temperature. This provides an expression of temperature corresponding to a given radiance:

$$T = \frac{B_\lambda \lambda^4}{2ck_B} \tag{2.14}$$

valid under Rayleigh-Jeans approximation.

In general, by inverting Planck's function (2.3), one can always derive the temperature that corresponds to a given radiance (***brightness temperature***):

$$T_b = \frac{hc}{k_B} \frac{1}{\lambda \ln(\frac{2hc^2}{\lambda^5 B_\lambda} + 1)} \tag{2.15}$$

which is strongly wavelength-dependent since both absorption and scattering processes are.

2.4 Radiative Transfer

As already introduced sooner in this Chapter, the propagation of electromagnetic waves through the atmosphere is influenced by many phenomena of interaction with the atmospheric components. The radiative transfer (RT) allows to take into account the radiance variations caused by the atmospheric constituents.

Most of the remote sensing instruments work as receptors of the radiance emitted by a certain body (see better in Chapter 5). However, while the radiation is on its way to the instrument, it will interact with absorbing elements placed along the path (or in general, between the source and the observer). When reading a remotely sensed spectrum, it will not appear continuous, but marked by so-called ***absorption lines***. These usually appear as dark lines, or lines of reduced intensity, on the spectrum.

The absorption and emission capacity of a gas, in particular, are strongly dependent on the wavelength of the radiation involved. This means that certain wavelengths are particularly sensitive to the absorption of certain molecules. This feature is the basis of many detection techniques in the IR.

The main absorbers in the UV range are atomic and molecular oxygen, molecular nitrogen and ozone, while in the VIS they are water vapour, molecular oxygen, ozone and carbon dioxide. The gases absorbing in the IR are the greenhouse gases (Chapter 1). The main ozone absorption in the infrared wavelengths takes place at the centre of the atmospheric window ($\approx 9.4-\ 9.6\,\mu m$), although other minor bands of ozone absorption exist too.

Another way to visualise absorption is through ***transmittance*** (τ_ν), which, computed between two points in space z_1 and z_2, corresponds to:

$$\tau_\nu(z,\theta) = e^{-\frac{1}{cos(\theta)}\int_{z_1}^{z_2}\alpha_\nu(z)\rho(z)dz} \tag{2.16}$$

i.e. the portion of radiation at frequency ν that manages to pass through the atmospheric path between z_1 and z_2, with an angle of incidence θ, despite the attenuation caused by the absorption itself (refer to Figure 2.5). $\alpha_\nu(z)$ is again the absorption coefficient, while $\rho(z)$ the density of the medium. The denser the material, the stronger the attenuation. In other words, transmittance represents the ratio of incident radiation to the amount of radiation that passes through the medium, and, consequently, it has no unit.

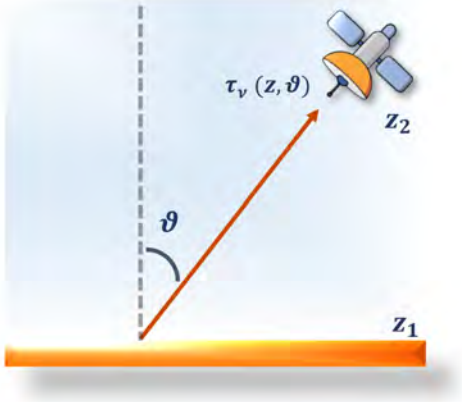


Figure 2.5: Simplified sketch of the geometry for transmittance measurement.

Figure 2.6 shows values of transmittance as a function of the wavelength, up to $15\,\mu m$, taking into account the whole thermal IR. A reading of the data shows that atmosphere is

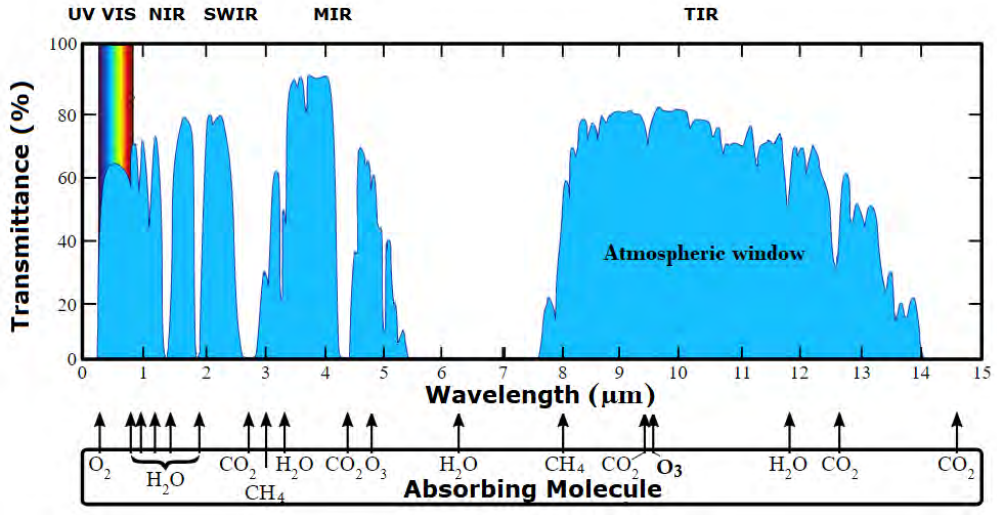


Figure 2.6: Atmospheric transmittance (%), as a function of wavelength, between 0 and 15 µm. The absorption bands associated with the main atmospheric gases are pointed [image adapted from `https://www.usna.edu/` website].

very opaque in the UV between 0 and 0.4 µm, while transparent in the VIS range between 0.4 and 0.75 µm. In the infrared range, bands that are opaque due to the absorption of molecules such as water vapour, ozone, methane and carbon dioxide, are interposed by atmospheric windows, the widest of which is between 8.5 and 12.5 µm.

2.4.1 Weighting Functions and Jacobians

A *weighting function* is the derivative of transmittance with respect to altitude:

$$K_\nu(z) = \frac{\partial \tau_\nu(z,\theta)}{\partial z} \tag{2.17}$$

It provides the relative contribution of each atmospheric layer to the radiation emitted towards space, with an angle θ. Through the weighting functions one can get an estimation of which atmospheric regions are better sensed from space at that specific wavelength.

Jacobians, on the other hand, are used to assess the sensitivity of the brightness temperature to a physico-chemical parameter of interest (i.e. temperature, water vapour, ozone, carbon monoxide, and so on). A Jacobian can thus be computed as:

$$J_\nu(X) = \frac{\partial T_b}{\partial X} \tag{2.18}$$

where X represents the generic geophysical parameter against which the T_b variation is evaluated. The gradient thus obtained will display which levels in the atmospheric column are most sensitive to the variations of the X parameter taken into account.

2.4.2 Radiative Transfer Equation

If assuming an optical path dz of a volume in atmosphere, the budget of the processes experienced by the radiation in passing through it will include the negative term of absorption plus the scattering and emission. The equation describing the radiative transfer through dz will be:

$$\frac{dI_\nu}{dz} = \underbrace{-\alpha_\nu I_\nu}_{absorption} + \underbrace{I_\nu^{scat}}_{scattering} + \underbrace{I_\nu^{emis}}_{emission} \tag{2.19}$$

where

$$I_\nu^{scat} + I_\nu^{emis} = I_\nu^{source} \tag{2.20}$$

represents the *source function* in the volume of interest. This is the **radiative transfer equation**, which is simple in its concept, but it is analytically almost impossible to solve.

In the spectral domain of the thermal infrared, one can consider the hypothesis of LTE, and thus scattering can be ignored. In this approximation:

$$I_\nu^{source} \approx I_\nu^{emis} = \alpha_\nu B_\nu(T) \tag{2.21}$$

The RT equation becomes then the **Schwarzschild equation**, which is the basis of radiometry and infrared remote sensing in clear-sky conditions (no areosols and no clouds):

$$\frac{1}{\alpha_\nu}\frac{dI_\nu}{dz} = -I_\nu + B_\nu(T_z) \tag{2.22}$$

with T_z the temperature at point z. After integration of the Eq. 2.22 between z_0 (the surface) and z_{sat} (position of the nadir-viewing sensor, i.e. $\theta = 0$):

$$I_\nu(z) = I_\nu(z_0)e^{-\zeta(z_0,z_{sat})} + \int_{z_0}^{z_{sat}} \alpha_\nu(z)B_\nu(T_z)e^{-\zeta(z,z_{sat})}dz \tag{2.23}$$

where

$$\zeta(z_0, z_{sat}) = \int_{z_0}^{z_{sat}} \alpha_\nu(z)\rho(z)dz \tag{2.24}$$

is the *optical depth*. The first term in Eq. 2.23 represents the exponential attenuation of the initial intensity due to the absorption contribution throughout the optical depth. The second term describes the thermal emission by the atmosphere, which behaves as a black-body. By replacing transmittance as in Eq. 2.16, Eq. 2.23 can also be wrote as:

$$I_\nu(z) = I_\nu(z_0)\tau_\nu(z_0, z_{sat}) + \int_{z_0}^{z_{sat}} \alpha_\nu(z)B_\nu(T_z)\tau_\nu(z, z_{sat})dz \tag{2.25}$$

Spectral radiance is attenuated by a factor τ in its path from the surface to the detector.

2.4.3 Impact of the Thermal Contrast in the Infrared

The present subsection is inspired by Claeyman (2010).

When examining the composition of a certain layer close to the surface through remote sensing techniques, one has to consider the thermal contrast existing between the surface itself and the layer all over it.

Let us assume the surface being at a temperature T_s and the atmospheric layer on the top of it at temperature T_l (scenario sketched in Figure 2.7). The radiance at the top of this layer can be obtained as:

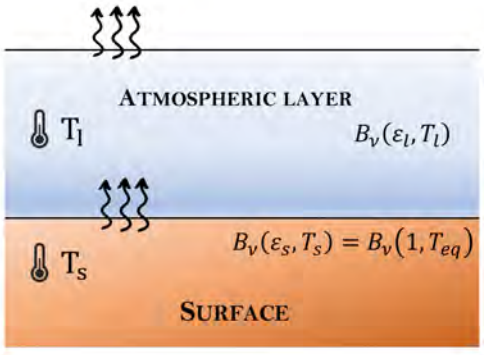


Figure 2.7: Scheme of thermal contrast at the surface.

$$I_\nu = (1 - \varepsilon_l)[\varepsilon_s B_\nu(T_s)] + \varepsilon_l B_\nu(T_l) \tag{2.26}$$

where ε_s is the surface emissivity, ε_l the emissivity of the atmospheric layer. Therefore, if T_{eq} is the temperature of a surface corresponding to a surface emissivity of $\varepsilon_s = 1$ and considering $1 - \varepsilon_l = \tau_l$ the spectral transmission of the layer, Eq. 2.26 becomes:

$$\begin{aligned} I_\nu =& \tau_l[B_\nu(T_{eq})] + (1 - \tau_l)B_\nu(T_l) \\ =& \tau_l[B_\nu(T_{eq}) - B_\nu(T_l)] + B_\nu(T_l) \end{aligned} \tag{2.27}$$

Three situations can occur:

- $T_l = T_{eq}$: temperatures being co-incident, no information can be obtained about the atmospheric layer above the surface;
- $T_l < T_{eq}$: information can be obtained about absorption of the atmospheric layer above (usually during daytime);
- $T_l > T_{eq}$: information can be obtained about emission of the atmospheric layer above (*temperature inversion*, mainly in night-time)

2.5 Radiative transfer modelling

The RT equation solution gives the spectral radiance reaching a detector in space in its path from the surface. As already detailed, this quantity can be expressed as a function of the state of the atmosphere itself. Over the past decades, models have been developed in order to simulate radiances by solving this so-called ***forward problem***, which consists in reproducing what an instrument could observe given prior knowledge of the atmospheric state. This type of modelling, can be quite complex since, as explained so far, it must take into account most

of the radiation-matter interaction processes, depending on the requirements and the degree of accuracy expected.

In the remote sensing field, however, certain atmospheric information often needs to be derived starting from the radiance acquired by a satellite-based sensor. In this case, one can talk about an ***inverse problem***. Many RTM are developed in order to carry on this task too. This kind of solution is strongly affected by the measuring instrument, and a single precise solution can not be derived from a single measurement. In order to make the computation as reliable as possible, the measure itself is usually combined with a so-called *a priori* information, such as a climatology or the output of a model.

Radiative transfer modeling for satellite data assimilation is very complex as well as very computationally expensive. Therefore, meteorological centres require also the RTMs to be fast. This is sometimes necessary in order to be able to assimilate a huge amount of data, coming from many satellite instruments, in a reasonable and efficient time-frame.

2.5.1 Types of RTMs

The RTMs currently developed and exploited by the scientific community are characterised by their different computational efficiency and the assumptions they are based-on.

A large subgroup of RTMs are the **line-by-line (LBL) models**, which take into account the absorption of each molecule present in the spectrum. The resulting transmittance is the results of each contribution. The computation is accurate but also expensive in terms of computation. A few of the famous models belonging to this category are: LBLRTM[3] [Clough et al. (1992); Clough et al. (2005)], KCARTA[4] [DeSouza-Machado et al. (1997)], MODTRAN[5] [e.g. Berk et al. (2014)].

Another family of RTMs are the **fast radiative transfer models**, assessing spectral transmittance by bands and not by monochromatic rays. These models compute a transmittance that is averaged over the band, assuming it does not vary significantly along the considered range of wavelengths. These models are both fast and efficient. Several fast models developed have been developed over the years. Among others the models AIRS-RTA[6] [Strow et al. (2003)], CRTM[7] [e.g. Ding et al. (2011)] and RTTOV, presented hereunder, can be counted.

2.5.2 RTTOV Model

The RTM used for this study is RTTOV in its most recent version at the time the work was carried out, i.e. version 12 (RTTOVv12) [Saunders et al. (2018)]. RTTOV is a very fast RTM for the simulation of radiances at the TOA (forward model). It was first developed during the

[3]Line-By-Line Radiative Transfer Model (LBLRTM)
[4]KCompressed Atmospheric Radiative Transfer Algorithm (KCARTA)
[5]MODerate resolution atmospheric TRANsmission (MODTRAN)
[6]Atmospheric Infrared Sounder radiative transfer algorithm (AIRS-RTA)
[7]Community Radiative Transfer Model (CRTM)

1990s by the European Centre for Medium-range Weather Forecasts (ECMWF), implemented in their variational system in order to enable radiance direct assimilation [Andersson et al. (1998); Saunders et al. (1999)]. The later independent development of the model took place within the Numerical Weather Prediction for Satellite Application Facilities (NWP SAF), founded by EUMETSAT. To this day it continues to be developed by Met Office, Météo-France, the German weather service DWD[8] and ECMWF and exploited by several Numerical Weather Prediction (NWP) centres around the world.

Although initially developed for TIROS Operational Vertical Sounder (TOVS) applications, RTTOV is now able to simulate radiances for passive downward-viewing satellite instruments (i.e. interferometers, spectrometers and radiometers) in a wide range of wavelengths, including VIS, IR and MW. A wide list of satellites by now supported by RTTOV [updated list available on NWP SAF website[9], list for v12 in Hocking et al. (2019)]. To each instrument corresponds specific optical depth coefficients calculated for the instrument itself, knowing the instrument spectral response function (ISFR).

RTTOV is designed to compute the forward model, but it can optionally provide also tangent linear and adjoint calculations, and Jacobian matrix. It is very useful for retrieval of satellite radiances, direct radiance assimilation in meteorological models and for the simulation of future instruments.

2.5.2.1 RTTOV Clear-Sky Equation

RTTOV, in order to simulate the TOA satellite radiances, relies on the computation of the following RT equation:

$$I_\nu(\theta_{sat},\theta_{sun}) = (1-N)I_\nu^{Clr}(\theta_{sat},\theta_{sun}) + NI_\nu^{Cld}(\theta_{sat},\theta_{sun}) \tag{2.28}$$

which contains both *cloudy* I_ν^{Cld} and *clear-sky* I_ν^{Clr} radiance contributions. θ_{sat} and θ_{sun} are the satellite and solar zenith angle respectively, while N is the product of the fractional cloud amount and the cloud emissivity assuming it behaves as a grey-body.

For the present work, RTTOV has been used in clear-sky conditions (the scattering effect of clouds and aerosols have not been taken into account):

$$\begin{aligned} I_\nu^{Clr}(\theta_{sat},\theta_{sun}) =& \tau_\nu(\theta)\varepsilon(\theta)B_\nu(T_s) + \int_\tau^1 B_\nu(T)d\tau + \\ &+[1-\varepsilon(\theta_{sat})]\tau_\nu^2(\theta_{sat})\int_\tau^1 \frac{B_\nu(T)}{\tau d\tau} L_\nu^{sol}(\theta_{sat},\theta_{sun}) \end{aligned} \tag{2.29}$$

with τ the transmittance from the surface up to space, ε is the surface emissivity, $B_\nu(T_s)$ is the Planck's function at a specific frequency and skin-temperature T_s. $L_\nu^{sol}(\theta_{sat},\theta_{sun})$ represents the direct and diffused solar radiation reflected by the surface [Saunders et al. (2018)].

[8] Deutscher Wetterdienst (DWD)

[9] `https://nwp-saf.eumetsat.int/site/software/rttov/documentation/platforms-supported/`

2.5.2.2 RTTOV Inputs

Pressure, temperature and water vapour profiles are the classical input for RTTOV. However, for IR sensors, the model is also able to simulate transmittance for atmospheric gases. Consequently, the state vector can also include concentration profiles of: O_3, CO, CH_4, CO_2, N_2O and SO_2.

The input profiles can be provided at pressure levels defined by the user. However, for the use, they will be interpolated by the model at the layering of the RTTOV coefficients, namely 54 vertical levels (from 1050 to 0.005 hPa) for multi-channel radiometers, or 101 levels (from 1100 to 0.005 hPa) for the IR hyperspectral sounders, e.g. Infrared Atmospheric Sounding Interferometer (IASI) [pressure levels listed in Hocking et al. (2019)], which need a larger vertical definition because of their larger amount of vertical degrees of freedom[10]. Once the optical depth profiles are computed, they are interpolated back to the user-defined levels for the radiative transfer computation.

Surface emissivity can be provided by the user or computed by RTTOV through physical models, such as ISEM[11] [Sherlock and Saunders (1999)] or IREMIS[12] [Saunders et al. (2017)].

Emissivity atlases are provided as part of the model package for a all the wavelengths that RTTOV can process. In the IR, the range of interest for this study, the modules are UWIREMIS[13] [Borbas and Ruston (2011)] and CAMEL[14] [Borbas et al. (2018)].

[10] NB: for IR hyperspectral sounders coefficients are also provided on 54 levels in order to reduce the computational time at the expense of accuracy if necessary.
[11] Infrared Surface Emissivity Model (ISEM)
[12] IR EMISsivity (IREMIS)
[13] University of Wisconsin IR EMISsivity (UWIREMIS)
[14] Combined ASTER MODIS Emissivity over Land (CAMEL)

2.6 Highlights of this Chapter

- The electromagnetic (EM) spectrum radiation consists into a multitude of monochromatic waves propagating in the vacuum at the speed of light. Depending on the wavelength (λ), several regions can be identified inside the EM spectrum. The one of interest for this study is the ***thermal infrared***.
- When an element of matter intercepts electromagnetic radiation, the interaction can be expressed in terms of transmission, **scattering**, **absorption** and **emission**.
- Of the total solar radiation arriving at the top of the atmosphere (TOA), 30% is reflected back by the atmosphere and surface itself. The remaining 70% is in part absorbed by gaseous components, while in a larger part it reaches the Earth's surface and is absorbed by this. The energy absorbed by the Earth must be balanced by an equal amount of energy emitted towards. Most of the energy outgoing to space is fully located in the infrared range.
- In order to explain the energy exchange mechanisms of absorption and re-emission taking place in the Earth system, the ***black-body radiation theory*** is used, which can easily be adapted to any type of body that absorbs or emits radiation.
- The **radiative transfer** considers the interactions of radiation with atmospheric constituents, leading to absorption lines in acquired spectra.
- The absorption capacity of gases depends on the wavelength, making certain wavelengths sensitive to specific molecules.
- The radiative transfer equation 2.19 describes the processes experienced by radiation as it passes through the atmosphere. The solution to the radiative transfer equation provides the spectral radiance observed by a detector in space from the Earth's surface.
- Models have been developed to simulate radiances by solving the forward problem, which replicates what an instrument would observe given knowledge of the atmospheric state.
- The **Radiative Transfer for TOVS (RTTOV)** model, developed by ECMWF and NWP SAF, is a fast radiative transfer model widely exploited for simulating TOA radiances for different satellite instruments, and is the one exploited throughout the present research work.

Chapter 3

Modelling of Atmospheric Chemistry

As already described in Chapter 1, the atmosphere is a highly complex system in which many physical and chemical processes occur simultaneously. Being able to predict the evolution of the chemical composition of the system atmosphere can be quite tricky since all of these processes need to be taken into account. This Chapter will focus on describing the basics of modelling atmospheric chemistry and the concepts underlying the creation of chemical transport models (CTMs). Special attention will then be paid to the CTM used for this research work, namely MOCAGE[1].

3.1 Modelling of Atmospheric Chemistry

Specific mathematical models for atmospheric chemistry are developed with the aim of forecasting the spatial and temporal concentration evolution of the chemical species of interest. Certain atmospheric components can, indeed, affect the air quality being detrimental to the health of the biome, or having an impact on the climate change. At the same time, high concentrations of others can be crucial to the unfolding of life on Earth (Chapter 1).

In order to make an efficient but high-quality realistic estimation, these models must take into account a series of processes and their interactions. More in detail, there are four main categories of physico-chemical processes that should be taken into account when modelling atmospheric chemistry, summarised in Figure 3.1.

- First of all, the processes injecting a certain substance into the atmosphere, or the precursors and factors necessary for its production, must be considered. These are the **sources** of emission already described in Chapter 1 and divided into *anthropogenic* and *biogenic emissions*.

[1]Modèle de Chimie Atmosphérique de Grande Echelle (MOCAGE)

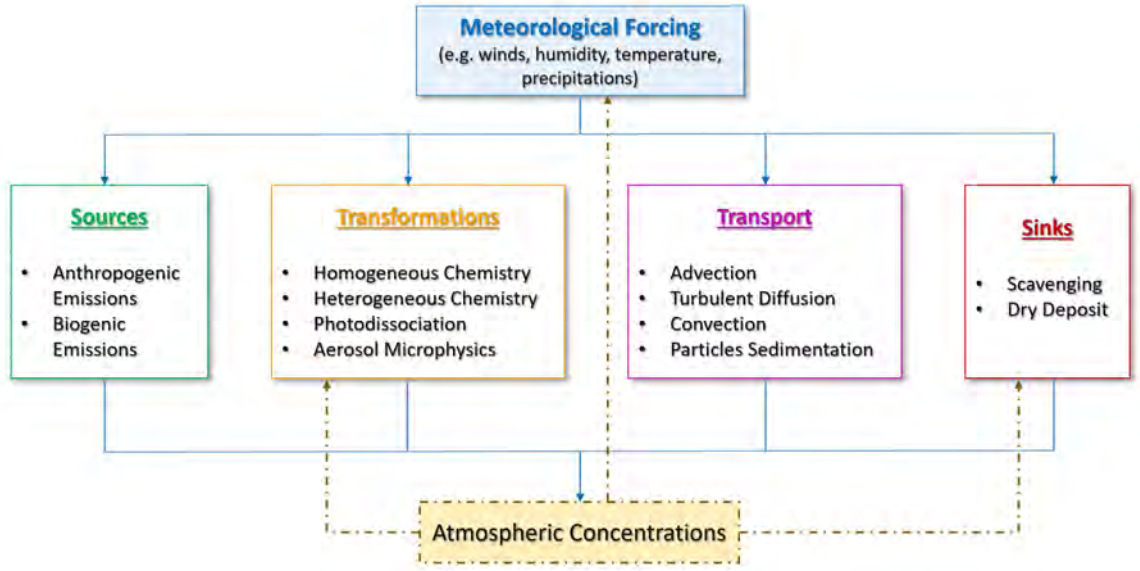


Figure 3.1: Main processes to be taken into account when creating a chemistry transport model (CTM) [inspired by Delmas et al. (2005)].

- Once they are in the atmosphere, gases and aerosols are often involved in processes of chemical **transformations**. These include chemical processes such as *photodissociation*, *homogeneous* and *heterogeneous chemistry*, and *micro-physical processes* involving the aerosols.

- As one might expect, atmospheric particles usually do not remain stationary, but rather move according to the dynamic of the atmosphere. The typical mechanisms of **transport** taking place in atmosphere are caused by *advection*, *turbulence*, *convection* and *sedimentation*.

- Of primary importance are also the processes leading to the removal of the modelled compounds. These are the so-called **sinks**, already mentioned in Chapter 1, including chemical, physical or biological processes able to remove the particle from the environment. Some of these, as chemical reactions and cycles, can be included in the category of transformations. Others, involving more physical processes, however, have also to be modelled when creating a CTM. They are the *scavenging*, i.e. the capture of atmospheric particles by mean of hydrometeors inside the cloud itself or during the precipitation (*washout*), and the *dry deposit*, i.e. the free fall to Earth of atmospheric trace gases and particulate matter.

Together, the listed processes contribute to determine the chemical concentrations in atmosphere. Each of them is influenced by the meteorological state of the atmosphere (e.g. winds, precipitation, temperature, humidity), affecting the evolution of the process itself. Therefore, including the so-called **meteorological forcing** is essential in order to properly characterise all of the other factors just described.

The concentration of chemical compounds, in turn, affect some of the aforementioned categories. Chemical transformations and processes of sinking are, of course, inter-connected with the chemical state of the atmosphere. Meteorological conditions are also influenced by the atmospheric particles concentrations. Aerosols, for instance, interact with the EM radiation, but they also have an impact on the creation of ice and liquid clouds. Gases also have an impact on the radiation. On the other hand, both the impacts on clouds and radiation have consequences on meteorological fields, such as temperature, humidity, wind and thus pressure. For this reason, all of the mutual inter-connections among processes just illustrated, as depicted in Figure 3.1, have also to be considered.

In contrast to the NWP and to the General Circulation Models (GCMs) with in-line chemistry and aerosols modelling, the CTMs do not compute the evolution of the meteorological parameters that come from external sources, and they are said to be working in ***off-line mode***, which makes the execution faster. This is the case of the thermodynamic parameters of the meteorological forcing, usually coming from other numerical models.

As already introduced, chemical transport models can have a wide range of applications. First of all, they are highly exploited for air quality forecasting. Many centres developed models able to provide daily air quality predictions or their trend over longer time-periods. They can forecast the evolution of pollutant concentrations. As a result, they prove to be a fundamental tool exploited by public authorities for decision-making in warnings and action plans to reduce the emissions. CTMs are exploited for the study of biochemical cycles, such as the injection and transport of ashes and SO_2, as a gas or transformed into sulphate aerosols, from volcanic eruptions and their interactions with the atmospheric system and the vegetation [e.g. Lamotte et al. (2021)]. They are also widely used to study the interactions between chemistry and climate change, or for ozone hole monitoring, which are very hot topics of the present days. Transport and deposition of the desert dust can also be monitored, due to the important impact dusts can have on health, climate, weather and many human activities [e.g. Martet et al. (2009); Carnevale et al. (2011)].

Many CTMs are currently developed and exploited by many centres. Among others, one can mention the MOZART[2] [e.g. Brasseur et al. (1998); Emmons et al. (2010)] developed by the NCAR[3] and working on a global scale, as well as the TM5[4] [Huijnen et al. (2010)] by KNMI[5]. The French model MOCAGE[6] works on both global and regional scale and will be widely illustrated in the following since it is the CTM used to carry out the present research. Other continental-scale CTMs to be mentioned are CHIMERE [Bessagnet et al. (2012), Menut et al. (2013)] (over Europe) and TGMs[7] (over USA). AMS-MINNI[8] [e.g. Mircea et al. (2014)]

[2]Model for OZone And Related chemical Tracers (MOZART)
[3]National Center for Atmospheric Research (NCAR)
[4]Tracer Model version 5 (TM5)
[5]Koninklijk Nederlands Meteorologisch Instituut (KNMI)
[6]Modèle de Chimie Atmosphérique de Grande Echelle) (MOCAGE)
[7]Third Generation Models (TGMs)
[8]Atmospheric Modelling System - of the Italian National Integrated Assessment Model MINNI (AMS-MINNI)

provides air quality information over Italy, while at an even smaller (urban) scale the SMOG[9] [Lu et al. (1997)] (Los-Angeles) can be mentioned. All of the models just listed, of course, are only a very small sample of those existing and under development.

Noteworthy with regard to atmospheric chemistry monitoring is the Copernicus Atmosphere Monitoring Service (CAMS), defined by the European Union and led by ECMWF, which provides both global and regional products of air quality, ozone, emissions and surface fluxes of GhG and from biomass burning, solar energy and climate forcing [Peuch et al. (2022)]. CAMS produces daily high quality global near-real-time forecasts of atmospheric composition using as initial conditions, for some of the chemical species, the assimilation of satellite retrievals. This system is constantly improved through updates to the exploited NWP model with in-line chemistry, i.e. the ECMWF's Integrated Forecast System (IFS), and including data from the newest satellite instruments [e.g. Inness et al. (2019)].

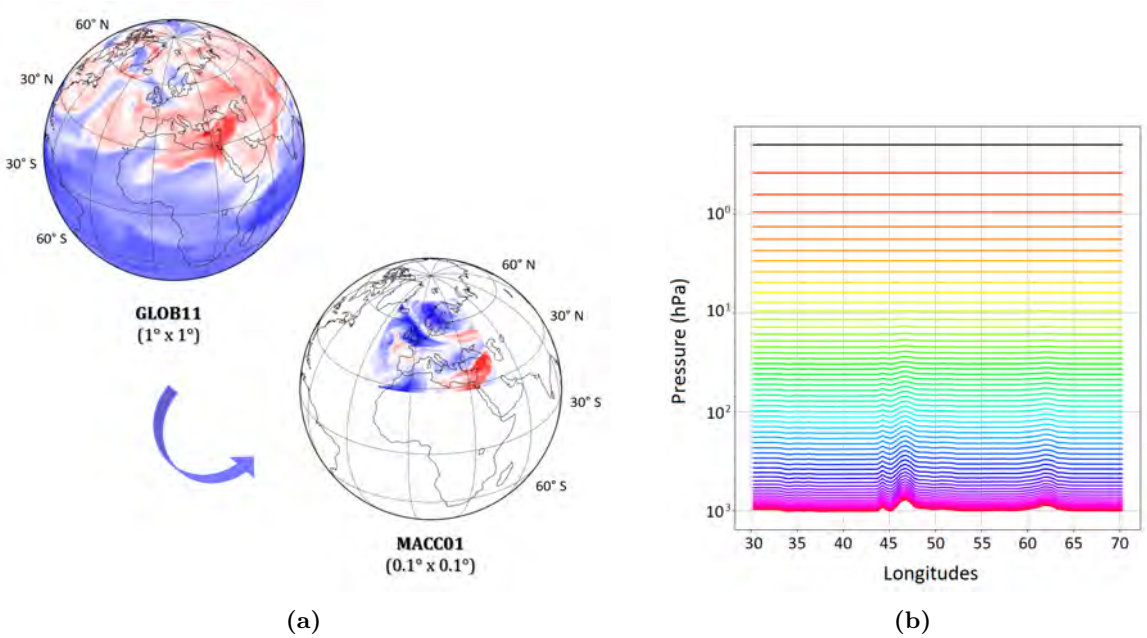
3.2 The MOCAGE model

MOCAGE is a three-dimensional chemistry transport model developed at the Centre National de Recherches Météorologiques (CNRM) since 2000 [e.g. Josse et al. (2004); Sič et al. (2015); Guth et al. (2016)]. From its creation to the present day, it has been used for a wide range of operational and research applications. It served to perform studies evaluating the impact of atmospheric chemistry on the climate change [e.g. Teyssèdre et al. (2007); Lacressonnière et al. (2012); Lamarque et al. (2013); Watson et al. (2016) and to examine the phenomenon of trace gases transport throughout the troposphere [Morgenstern et al. (2017); Orbe et al. (2018)]. Many efforts have been performed employing MOCAGE to investigate the exchanges taking place between troposphere and stratosphere using data assimilation [e.g. El Amraoui et al. (2010); Barré et al. (2014)]. It has been used as part of researches about the UTLS structure and to evaluate the impact of biomass burning plumes [e.g. Cussac et al. (2020)] or to estimate the impact of volcanic sulphur emissions on the tropospheric sulphur budget [Lamotte et al. (2021)]. Several studies on the assimilation of Aerosol Optical Depths (AODs) have been carried out within MOCAGE in order to improve the representation of aerosols within the model simulations [e.g. Sič et al. (2016); Descheemaecker et al. (2019); El Amraoui et al. (2022)]. These studies led to AODs operational assimilation in 2022. Finally, MOCAGE is a precious resource for the air quality monitoring and forecasting on the French Prev'Air platform [Rouil et al. (2009)] and over Europe within the regional ensemble of the CAMS [Marécal et al. (2015)].

3.2.1 Model Geometry Configurations

At present, MOCAGE disposes of two geographical configurations, using a two-way nested grids capacity (Figure 3.2a):

[9]Surface Meteorology and Ozone Generation (SMOG)



(a) **(b)**

Figure 3.2: **(a)** Illustration of MOCAGE global domain GLOB11, which presents a 1°lat × 1°lon horizontal resolution, and the regional domain over Europe, named MACC01, with a thinner resolution of 0.1°lat × 0.1°lon, and bounded in 28°N, 26°W and 72°N, 46°E. Colors indicate ozone field at a given level. **(b)** [adapted from El Aabaribaoune (2022)] MOCAGE vertical configuration of the 60 hybrid σ-pressure levels (each one represented with a different colour).

- **GLOB11**: a global scale with a 1° longitude x 1° latitude horizontal resolution;
- **MACC01**: a regional domain, bounded in 28°N, 26°W and 72°N, 46°E, and with a resolution of 0.1° longitude x 0.1° latitude (approximately 10 km at the latitude of 45°N).

For the vertical levels, MOCAGE uses a system of pressure on vertical **hybrid σ-pressure coordinates** [e.g. Simmons and Strüfing (1981); Eckermann (2009)]. The system represents a surface parallel to the Earth's topography, computed by scaling pressure at a level i starting from the value of surface pressure P_s. In the specific case of MOCAGE, this system is implemented by scaling the pressure $P(i)$ as follows:

$$P(i) = A(i) + B(i)P_s \tag{3.1}$$

where A_i and B_i are two scaling functions represented in Figure 3.3.

In terms of an actual depth of the layers, each one will be thicker (in metres) as the distance from the surface level will increases (i.e moving upward). Thanks to this coordinate system, each grid point is represented by a value above the surface and the topography is not a constraint. Further, since atmospheric density exponentially decreases with the altitude, the hybrid coordinates provide a thinner vertical resolution close to the surface.

The model hence has a non-uniform vertical resolution: 47 vertical altitude-pressure levels

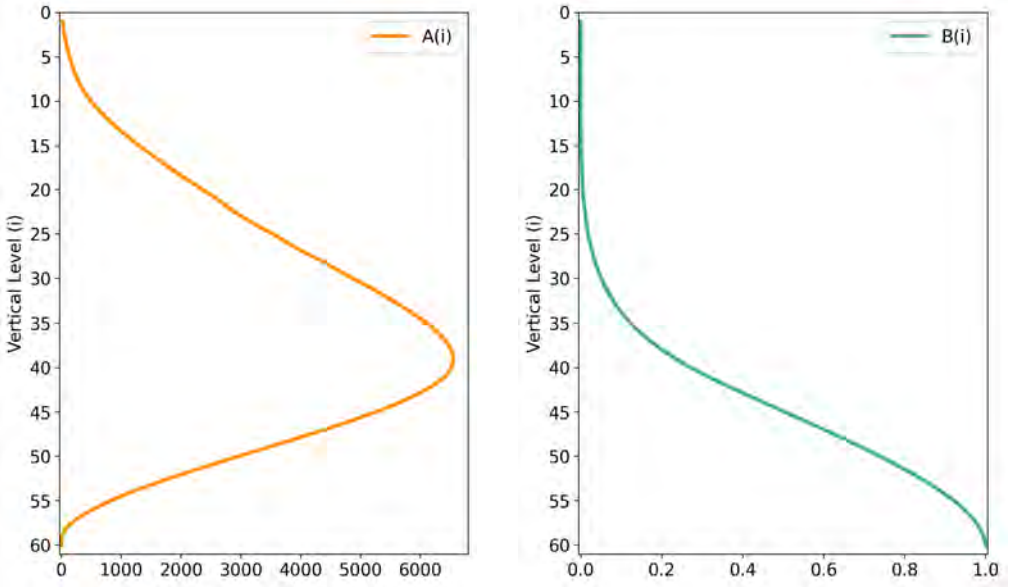


Figure 3.3: Vertical trend of A_i and B_i dimensionless coefficients used in MOCAGE to compute the hybrid-pressure vertical levels as in Equation 3.1.

from the surface up to 5 hPa. The levels are denser near the surface, with a resolution of about 40 m in the lower troposphere and 800 m in the lower stratosphere. A 60 hybrid levels version is also used in research mode and it is also the one exploited for the present work. This consists of the 47 levels computed as just described, plus 13 additional levels going up to 0.1 hPa. Resolutions in upper stratosphere is around 2 km. Please note that the MOCAGE vertical levels are numbered in descending order from the ground: the level the closest to the ground is the number 60, while the highest in the atmosphere is level 1. A schematic representation of the MOCAGE vertical configuration of the 60-hybrid pressure level is provided in panel 3.2b, while an equivalence among vertical level number and pressure can be found in Appendix A.

3.2.2 Schemes and Forcing

This model is able to simulate the chemistry in the lower stratosphere and troposphere, taking into account in detail photochemical processes and the transport of longer-lived species. It uses different chemical schemes in order to reproduce the atmospheric chemical composition: the REactive Processes Ruling the Ozone BUdget in the Stratosphere (REPROBUS) is used for the stratosphere [Lefevre et al. (1994)], while for the tropospheric representation the Regional Atmospheric Chemistry Mechanism (RACM) is exploited [Stockwell et al. (1997)]. Through the combined use of the two aforementioned schemes (called RACMOBUS), MOCAGE is able to simulate 118 gaseous species, 434 chemical reactions, primary aerosols and secondary inorganic aerosols.

MOCAGE CTM runs in an off-line mode. Depending on the application the meteorological forcing may come from a coupling with a general circulation climate model (for climate

studies), or with NWP models (e.g. for near real time applications). The core of the chemical reactions used in MOCAGE is also exploited on-line into the IFS [Huijnen et al. (2019); Williams et al. (2022)].

The transport of species is computed through a Lagrangian scheme.

The data concerning chemical emissions are provided to MOCAGE as external data-sets. For the present operational configuration, MEGAN-MACC [Sindelarova et al. (2014)] and Global Emissions Inventory Activity (GEIA) are used for biogenic emissions. MACCity [Lamarque et al. (2010); Granier et al. (2011); Diehl et al. (2012)], RCP60 [Fujino et al. (2006); Van Vuuren et al. (2011)], CAMS-REG-AP [Guevara et al. (2022); Kuenen et al. (2022)] and GEIA are those provided to cover the information about anthropogenic emissions. For the representation of the biomass burning, data from the CAMS Global Fire Assimilation System (GFAS)[10], from ECMWF, are exploited.

[10]The GFAS is a system assimilating observations of fire radiative power (FRP), issued from satellite-based sensors. These data are then exploited to produce daily estimations of wildfires and biomass burning emissions. The observations currently assimilated in the GFAS are active fire products from MODIS, on NASA's Terra and Aqua.

3.3 Highlights of this Chapter

- Atmospheric chemistry models are used to forecast the concentration evolution of chemical species in the atmosphere.
- Four main categories of physico-chemical processes are considered in modeling atmospheric chemical composition: *emission sources*, *chemical transformations*, *atmospheric transport* and *removal processes*.

  Meteorological conditions play a crucial role in these processes and need to be provided.
- *Chemistry transport models (CTMs)* are used for air quality forecasting, studying biochemical cycles, examining the interactions between chemistry and climate change, and monitoring phenomena like ozone holes and desert dust transport.
- The **Modèle de Chimie Atmosphérique de Grande Echelle (MOCAGE)** is a CTM, developed at the CNRM, that has been applied to various operational and research applications.
  - MOCAGE has *global* (GLOB11) and *regional* (MACC01) *configurations* and uses a non-uniform vertical resolution with *hybrid pressure coordinates*.
  - Various chemical schemes are employed in MOCAGE to simulate the atmospheric composition in the stratosphere and troposphere, including photochemical processes and the transport of species.
  - MOCAGE operates in an off-line mode, receiving meteorological forcing from general circulation climate models or numerical weather prediction models.
  - The transport of species is computed through a Lagrangian scheme.
  - The data concerning chemical emissions are provided to MOCAGE as external data-sets.

Chapter 4

Data Assimilation

The data assimilation (DA) methods are meant to find an optimal representation of the true state of the atmosphere, otherwise unknown, by combining the information coming from a *background model state*, i.e. the estimated atmospheric state from previous short-term forecasts, and the *observations*. This is done by seeking the best compromise between the two different data-sets, assigning each of them an appropriate weight through the associated errors.

This Chapter will set out the basic nomenclature of the main elements in DA and the techniques used in this work for the error computations. Finally, the three-dimensional variational assimilation (3D-Var) minimisation method will be illustrated.

4.1 Basic Concepts of Data Assimilation

The first step when formalising a data assimilation problem is to set up the basics of the working space. Concepts and elements here reported follow what is set out by [Bouttier and Courtier (2002)].

4.1.1 State Vector

- First of all, the quantities needed to define the atmospheric state are collected in a column matrix $\mathbf{x}$, known as the ***state vector*** .
- Of course, the actual state of the atmosphere is not known. However, the representation of reality that comes as close as possible to the truth can be designed and it is the one defined by the vector $\mathbf{x}_t$, or ***true state*** (dimension n).
- The ***background state*** is an estimation of the true state of the atmosphere made by the model, or from a climatology, before the assimilation is performed (*a priori*), and it is defined by $\mathbf{x}_b$ of dimension n. To have an idea of the magnitude of n, one can consider that, in the case of MOCAGE in its global configuration (1° of horizontal resolution) and with 60 vertical levels, for a single species, $n = 180 \times 360 \times 60$.

- As already introduced, the purpose of data assimilation is to complement the observed and modelled information in order to find a proper representation of the atmosphere. The state fulfilling these requirement is called ***analysis*** and is denoted by $\mathbf{x}_\mathrm{a}$ (dimension n).

4.1.2 Control Variable

It often happens in practice that it is not possible or affordable to solve the problem of analysing all the elements of the state vector. This may be caused by limits in computational capacity or by not being able to perform a correct analysis for all elements. In this case, the analysis is not performed in the model space, but from a space allowing to correct the background, which is known as the **control vector** and is a subset of the state vector itself to be modified during the assimilation.

4.1.3 Observations

The vector containing the observations used for a given analysis, i.e. ***observation vector***, is $\mathbf{y}$ (dimension p).

In order to be able to use the observations, it is necessary that they are in some way comparable to the state vector. If each degree of freedom of the model were directly observed, the observations could be regarded as a particular value of the state vector. In reality, though, the observations often exist in a smaller amount than the considered variables and they are irregularly positioned with respect to the model points. The only way to be able to compare observations and model state, therefore, is to dispose of a function going from the model space to the observation one, i.e. the ***observation operator*** H (from dimension n to p). This function produces the ***model equivalents of the observations*** $H(\mathbf{x})$. In practice, H consists of interpolation operators carrying out the model discretization to the observation points, and converting the model variables to observed parameters. This role is often covered by a radiative transfer model when assimilating satellite radiances.

Of central importance for data assimilation methods are therefore the differences: $\mathbf{y}-H(\mathbf{x})$. These are defined:

- ***innovations***: $\mathbf{d}_\mathrm{b}^\mathrm{o} = [\mathbf{y} - H(\mathbf{x}_\mathrm{b})]$, when computed using the background vector $\mathbf{x}_\mathrm{b}$;
- ***analysis residuals***: $\mathbf{d}_\mathrm{a}^\mathrm{o} = [\mathbf{y} - H(\mathbf{x}_\mathrm{a})]$, when using $\mathbf{x}_\mathrm{a}$.

4.1.4 Error Modelling

The attribution of appropriate errors to the terms that enter the data assimilation process is crucial to the accuracy of the analysis. This is why the error estimation becomes a fundamental step in the method.

The errors are relative to the true state of the atmosphere, which is not known. As a result, instead of computing the exact error to be attributed to a class of data, the statistics

should be estimated. Different methods have been developed in the past decades to be able to estimate either *background* or *observation errors*. Below are the definitions of these errors and the methods used in the course of this work to evaluate them.

4.1.4.1 Background Errors

The ***background error*** consists in the difference occurring between the background state and the true state:

$$\varepsilon_{\mathrm{b}} = \mathbf{x}_{\mathrm{b}} - \mathbf{x}_{\mathrm{t}} \tag{4.1}$$

The smaller the error, the closer the background state will be to the truth and the less the observation weight in the analysis.

If $\overline{\varepsilon}_{\mathrm{b}}$ represents the average[1] (or expectation) of 4.1, the **covariance matrix of the background errors** can be defined as:

$$\mathbf{B} = \overline{(\varepsilon_{\mathrm{b}} - \overline{\varepsilon}_{\mathrm{b}})(\varepsilon_{\mathrm{b}} - \overline{\varepsilon}_{\mathrm{b}})^T} \tag{4.2}$$

In a scalar system, the covariance is simply the variance:

$$\mathbf{B} = var(\varepsilon_{\mathrm{b}}) = \overline{(\varepsilon_{\mathrm{b}} - \overline{\varepsilon}_{\mathrm{b}})^2} \tag{4.3}$$

On the other hand, if the model state vector has dimension n, the covariance matrix will have dimension $n \times n$. In this case the diagonal of the matrix will contain variances, while the off-diagonal terms will be the cross-covariances between each of the model variables, between geographical grid points or between vertical levels. If, just as an example, the model variables are three and their background errors are denoted as $e_x = \varepsilon_{\mathrm{b}x} - \overline{\varepsilon}_{\mathrm{b}x},\ x = 1, 2, 3$, the $\mathbf{B}$ matrix will assume the following shape:

$$\mathbf{B} = \begin{bmatrix} var(e_1) & cov(e_1, e_2) & cov(e_1, e_3) \\ cov(e_1, e_2) & var(e_2) & cov(e_2, e_3) \\ cov(e_1, e_3) & cov(e_2, e_3) & var(e_3) \end{bmatrix} \tag{4.4}$$

The off-diagonal terms can be provide the ***error correlations***:

$$\rho(e_i, e_j) = \frac{cov(e_i, e_j)}{\sqrt{var(e_i) var(e_j)}} = \frac{cov(e_i, e_j)}{\sigma_i \sigma_j} \tag{4.5}$$

where $\sigma = \sqrt{var(e)}$ is the ***standard deviation*** of e (see Appendix B).

Another way to describe Equation 4.4 is as the matrix product between the diagonal matrix Σ, containing the background error standard deviations, and the matrix $\mathbf{C}$ containing

[1] Statistical indices mentioned in this section, e.g. average, standard deviation, variance, and so on, are detailed in Appendix B.

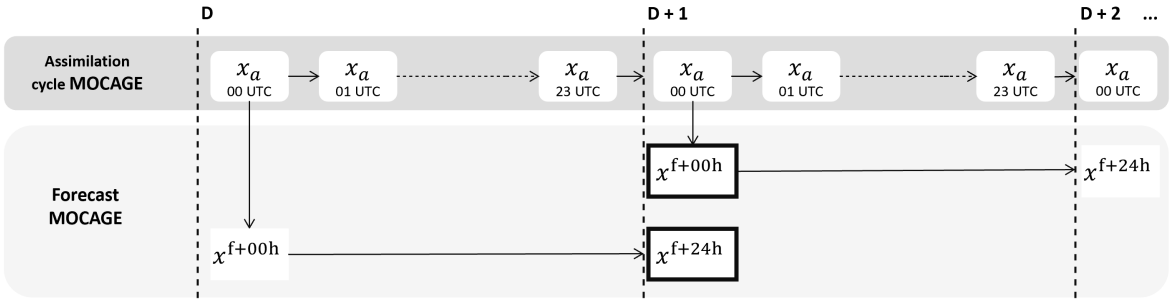


Figure 4.1: An illustrative chart of the NMC method using MOCAGE. $\mathbf{x}_a$ represents the hourly analyses obtained from MOCAGE. x^{f+00h} is the forecast at 0h initialised with the assimilation $\mathbf{x}_a$ at time 00 UTC (the two of them are superimposable). Finally, $\mathbf{x}^{f+24h}$ is the forecast at +24h. The forecasts $\mathbf{x}^{f+24h}$ for day D and $\mathbf{x}^{f+00h}$ for day D+1 are valid at the same time.

horizontal and vertical correlations (both the latter provided through Gaussian functions):

$$\mathbf{B} = \Sigma\, \mathbf{C}\, \Sigma^{T} \tag{4.6}$$

In the practice of data assimilation various methods exist for computing the errors and then building up this kind of matrix. In the course of this PhD project, essentially, only matrix Σ has been directly modified and computed through two different methods, used in different parts of the work. The description of horizontal and vertical correlations, on the other hand, was directly obtained through the process of assimilation in MOCAGE and remain the same in both cases.

More technical details and context about the usage of the two **B** matrices thus obtained will be given in the following Chapters (6 and 7). However, the basics of the two methods will be briefly outlined in this Section. The first consists of assigning the terms of the diagonal, i.e. the variances, a value obtained as a percentage of the actual concentration value of the atmospheric profile of the quantity under consideration. The other one is a method introduced by the **National Meteorological Centres (NMC)** [Parrish and Derber (1992)]. This is a technique that defines background errors using differences in NWP forecasts valid at the same time (but at different range). In this work, the NMC method is applied on MOCAGE cycles over certain time-period. As shown in Figure 4.1, each day D will provide an assimilation $\mathbf{x}_a$ for every hour of the day. At the same time, a MOCAGE forecast is run up to a 24-hour forecast range for each day D, initialised with the $\mathbf{x}_a$ assimilation of the same day at 00 UTC. The forecasts $\mathbf{x}^{f+24h}$ for day D and $\mathbf{x}^{f+00h}$ for day D+1 are valid at the same time. Thereafter the covariance of the differences between each couple are computed, averaged over time and longitudes. This produces a 2D-field that represents the standard deviations of the background errors, varying along latitudes and vertical levels.

4.1.4.2 Observation Errors

The ***observation errors*** consist of instrument noise, error in the observation operator, representativeness and pre-processing errors. They can be split into systematic and random components. Systematic errors are usually biases[2] that should be rectified before the data are assimilated. Consequently, the observation errors that mostly need to be represented are the random ones.

Back to the notation used so far, the observation errors are defined as follows:

$$\varepsilon_{\mathrm{o}} = \mathbf{y} - H(\mathbf{x}_{\mathrm{t}}) \tag{4.7}$$

of average $\overline{\varepsilon}_{\mathrm{o}}$. The **covariance matrix of the observation errors** is instead:

$$\mathbf{R} = \overline{(\varepsilon_{\mathrm{o}} - \overline{\varepsilon}_{\mathrm{o}})(\varepsilon_{\mathrm{o}} - \overline{\varepsilon}_{\mathrm{o}})^T}. \tag{4.8}$$

Everything has already been explained about the structure of the $\mathbf{B}$ matrix is also valid for the $\mathbf{R}$. However, while still a square matrix, it will have dimension $p \times p$ (if the observation vector $\mathbf{y}$ has dimension p). More specifically, p represents the number of horizontal measurement points times the number of points providing vertical information. In $\mathbf{R}$, therefore, correlations can be described concerning the vertical information, or concerning the horizontal measurement points, or both at the same time. In numerical prediction, the vertical correlations are those usually described.

Observation error correlations can sometimes assumed to be zero. In this case, the errors associated to each measure are supposed to be determined by physically independent phenomena. In the past years, however, the importance of also representing off-diagonal correlations has emerged, especially for satellite data. As a result of the above, it often proves necessary to structure $\mathbf{R}$ matrices that are not diagonal, but also contain non-zero covariance terms in order to take into account the correlations.

Throughout this work, a procedure introduced by Desroziers et al. (2005), and for this reason usually referred to as **Desroziers's diagnostics**, is the one used to estimate the structure of a full $\mathbf{R}$ matrix. This technique is very efficient and, in addition to having been used over the years for a wide range of researches [e.g. Weston et al. (2014); Vittorioso et al. (2021)], it is currently exploited in the operations by many meteorological centres. Through this method variances and covariances of observation errors can be obtained from innovations $\mathbf{d}_{\mathrm{b}}^{\mathrm{o}} = [\mathbf{y} - H(\mathbf{x}_{\mathrm{b}})]$ and residuals $\mathbf{d}_{\mathrm{a}}^{\mathrm{o}} = [\mathbf{y} - H(\mathbf{x}_{\mathrm{a}})]$ statistics. These matrix is given by the expression:

$$\mathbf{R} = E\left[\mathbf{d}_{\mathrm{a}}^{\mathrm{o}} (\mathbf{d}_{\mathrm{b}}^{\mathrm{o}})^{\mathrm{T}}\right] \tag{4.9}$$

where E is the statistical expectation operator. This operation requires to have already run an analysis in order to have a vector $\mathbf{x}_{\mathrm{a}}$ to use. To do so, a first run of assimilation is usually

[2] In statistics the *bias* is the deviation from the true value of a parameter.

run initialised by a diagonal $\mathbf{R}$ matrix, whose variances are represented by either instrumental noise or by standard deviation of the innovations.

4.1.4.3 Analysis Errors

The ***analysis errors*** are what one wants to minimize when performing data assimilation. They are defined as:

$$\varepsilon_{\mathrm{a}} = \mathbf{x}_{\mathrm{a}} - \mathbf{x}_{\mathrm{t}} \tag{4.10}$$

of average $\overline{\varepsilon}_{\mathrm{a}}$. A measure of these errors is given by the trace of the analysis-error covariance matrix $\mathbf{A}$:

$$\mathrm{Tr}(\mathbf{A}) = \overline{||\varepsilon_{\mathrm{a}} - \overline{\varepsilon}_{\mathrm{a}}||^2} \tag{4.11}$$

4.2 Three-Dimensional Variational Assimilation: 3D-Var

Several algorithms can be used in order to perform the data assimilation. The one used in this work is the so-called **3D-Var**. Through this method, the statistically optimal atmospheric state $\mathbf{x}_{\mathrm{a}}$, compromising background model state and observations, should be obtained by minimising the ***cost function***:

$$J(\mathbf{x}) = \underbrace{\frac{1}{2}\left(\mathbf{x}-\mathbf{x}_{\mathrm{b}}\right)^{\mathrm{T}}\mathbf{B}^{-1}\left(\mathbf{x}-\mathbf{x}_{\mathrm{b}}\right)}_{J_{\mathrm{b}}(\mathbf{x})} + \underbrace{\frac{1}{2}\left[\mathbf{y}-H(\mathbf{x})\right]^{\mathrm{T}}\mathbf{R}^{-1}\left[\mathbf{y}-H(\mathbf{x})\right]}_{J_o(\mathbf{x})} \tag{4.12}$$

where the term of background $J_{\mathrm{b}}(\mathbf{x})$ and observations $J_o(\mathbf{x})$ have been highlighted. As already explained, each term is weighted according to its respective error. The determination of the proper $\mathbf{B}$ and $\mathbf{R}$ is though a necessary precondition for reliability of the results. Please notice that three basics assumptions are needed for this formula to be valid:

1. background and observation errors are unbiased: $\overline{\varepsilon}_{\mathrm{b}} = 0$ and $\overline{\varepsilon}_{\mathrm{o}} = 0$;
2. observation and background errors are not correlated: $\overline{\varepsilon_{\mathrm{b}}\,\varepsilon_{\mathrm{o}}^{\mathrm{T}}} = 0$;
3. both $\mathbf{R}$ and $\mathbf{B}$ are known and positive defined.

In the hypothesis of $\mathbf{x}$ close enough to $\mathbf{x}_{\mathrm{b}}$ the observation operator can be linearised:

$$H(\mathbf{x}) = H(\mathbf{x}_{\mathrm{b}}) + \mathbf{H}(\mathbf{x}-\mathbf{x}_{\mathrm{b}}) \tag{4.13}$$

Using the definition of innovation vector already introduced, i.e. $\mathbf{d}_{\mathrm{b}}^{\mathrm{o}} = [\mathbf{y} - H(\mathbf{x}_{\mathrm{b}})]$, Equation 4.12 becomes:

$$J(\mathbf{x}) = \frac{1}{2}\left(\mathbf{x}-\mathbf{x}_{\mathrm{b}}\right)^{\mathrm{T}}\mathbf{B}^{-1}\left(\mathbf{x}-\mathbf{x}_{\mathrm{b}}\right) + \frac{1}{2}\left[\mathbf{d}_{\mathrm{b}}^{\mathrm{o}} - \mathbf{H}(\mathbf{x}-\mathbf{x}_{\mathrm{b}})\right]^{\mathrm{T}}\mathbf{R}^{-1}\left[\mathbf{d}_{\mathrm{b}}^{\mathrm{o}} - \mathbf{H}(\mathbf{x}-\mathbf{x}_{\mathrm{b}})\right] \tag{4.14}$$

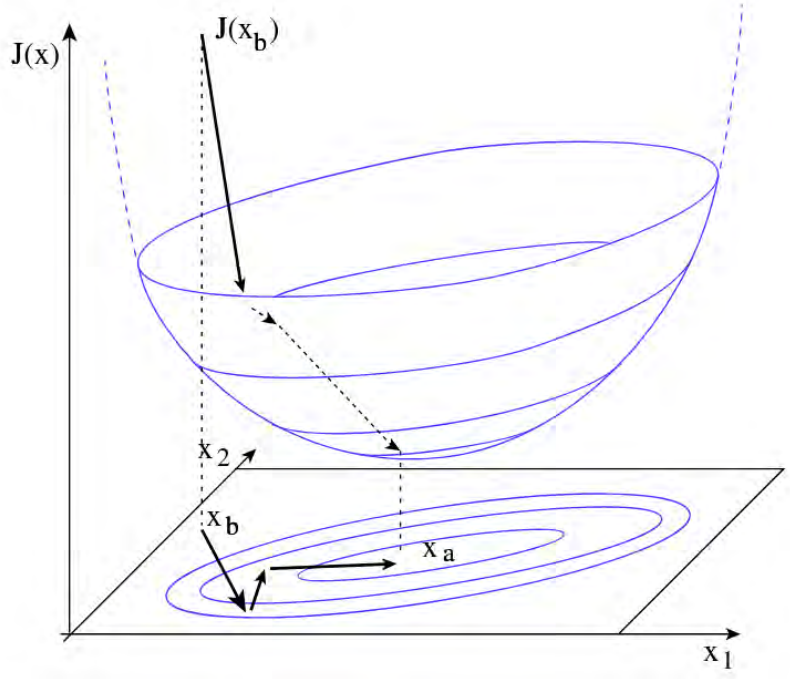


Figure 4.2: Schematic representation of the variational cost-function minimization in a two-variable model space [source Bouttier and Courtier (2002)].

The latter formulation allows the introduction of an incremental approach in which the aim becomes to find the increment $\delta\mathbf{x}$, or ***analysis increment***, that makes $\mathbf{x}_\mathrm{a} = \mathbf{x}_\mathrm{b} + \delta\mathbf{x}$ as close as possible to the true state $\mathbf{x}_\mathrm{t}$. By defining $\delta\mathbf{x} = \mathbf{x} - \mathbf{x}_\mathrm{b}$, the cost function to be minimised, then, becomes:

$$J(\delta\mathbf{x}) = \frac{1}{2}\,\delta\mathbf{x}^\mathrm{T}\,\mathbf{B}^{-1}\,\delta\mathbf{x} + \frac{1}{2}\,[\mathbf{d}_\mathrm{b}^\mathrm{o} - \mathbf{H}\,\delta\mathbf{x}]^\mathrm{T}\,\mathbf{R}^{-1}\,[\mathbf{d}_\mathrm{b}^\mathrm{o} - \mathbf{H}\,\delta\mathbf{x}] \tag{4.15}$$

Through this incremental formulation, the computational time can be significantly reduced since the problem is no longer solved using the complete vectors, but rather on the variation to be added to each background field in order to obtain the analysis.

The best estimation of the atmospheric state is, then, obtained by sicking a minimum, through a descendent algorithm, of the gradient:

$$\nabla J(\delta\mathbf{x}) = \mathbf{B}^{-1}\,\delta\mathbf{x} - \mathbf{H}^\mathrm{T}\mathbf{R}^{-1}[\mathbf{d}_\mathrm{b}^\mathrm{o} - \mathbf{H}\,\delta\mathbf{x}] \tag{4.16}$$

The solution is approximated since a limited number of iterations is performed. The limit for stopping the minimisation can be set by choosing artificially a number of iteration not to be exceeded, or by requiring the norm $||\nabla J(\mathbf{x})||$ to decrease by a specific factor during the minimisation. The geometry of the minimisation process is proposed in Figure 4.2, for a space with two model variables and a paraboloid. The arrows in the diagram represent the consecutive iterations of the minimisation process.

The initial state used to initialise the minimisation is defined ***first guess***, and it is often taken equal to the background state $\mathbf{x}_\mathrm{b}$ (but it is not compulsory). If the minimisation is well performed, it will not be significantly dependent on the choice of the first guess, but it will always be sensitive to the background.

4.3 Highlights of this Chapter

- Data assimilation methods aim to find an optimal representation of the true state of the atmosphere by combining information from a background model state and observations.
- The attribution of appropriate errors to the terms that enter the data assimilation process is crucial to the accuracy of the analysis. This is why the error estimation becomes a fundamental step in the method.
- During this project, ***covariance matrix of the background error*** (**B**) has been computed through two different approaches: the first consisted in assigning the terms of the diagonal, i.e. the variances, a value obtained as a percentage of the actual concentration value of the atmospheric profile of the quantity under consideration; the other one was a method introduced by the National Meteorological Centres (NMC) [Parrish and Derber (1992)].
- For the estimation of the ***covariance matrix of the observation error*** (**R**), a procedure introduced by Desroziers et al. (2005), and for this reason usually referred to as "Desroziers's diagnostics", is exploited.
- **Three-dimensional variational assimilation (3D-Var)** is an assimilation method used to minimise the cost function as in Equation 4.15, by weighting the background and observation terms according to their respective errors:
  - the gradient of the cost function (Eq. 4.16) is used to iteratively find the solution through a descendant algorithm;
  - the minimisation process stops after a specific number of iterations or when the norm of the gradient decreases by a specific factor;
  - the choice of the initial state, known as the first guess, typically starts with the background state.

Chapter 5

Observations

Monitoring of atmospheric chemistry, as well as meteorological parameters, is performed through different observation networks, more or less dense depending on the acquisition system. Two main classes of observations can be identified: ***in situ*** and from ***remote sensing***.

At the beginning of this Chapter *in situ* measurements will be briefly presented together with the data acquired by ozonesondes on balloons and distributed by the World Ozone and Ultraviolet Data Centre (WOUDC). These will be, indeed, used in the work for validation. Two other categories of data, although not used directly in the work, will be briefly illustrated as they are of interest for future validations. These are the data issued by surface meteorological stations and aircrafts.

In a second time, the main concepts of remote sensing necessary for the understanding of the work and the status of the available satellite observation network will be exposed. Satellite measurements from Infrared Atmospheric Sounders Infrared Atmospheric Sounding Interferometer (IASI) and Ozone Monitoring Instrument (OMI) have been used and will then be described.

Nevertheless, the main focus will be set on the Infrared Sounder (IRS) on board MTG-S, i.e. the instrument whose impact on the chemistry transport forecasted by MOCAGE is expected to be assessed through this research. Particular detail will therefore be given to the instrument, its functioning and its sensitivity to different species. The thesis, as introduced at the beginning of the manuscript, also aims to evaluate the effect that an ideal increase in resolution of the instrument in question would produce on the same assimilation frame. This hypothesis will be detailed as well.

5.1 *In Situ* Measurements

Formally, any observation taken by an instrument in direct contact with the medium it is sensing is defined *in situ*. This kind of measurements can concern **surface observations**, carried out through ground-based measurement stations, or **airborne observations**, acquired trough instruments boarded on aircrafts, balloons or drones.

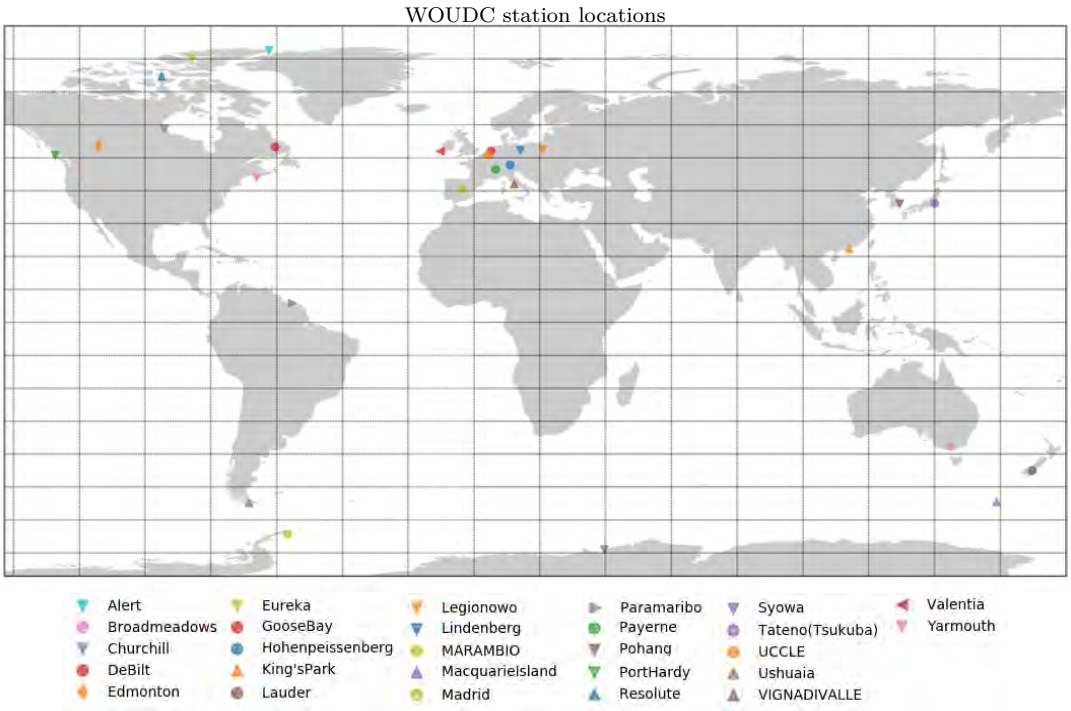


Figure 5.1: Localisation of the 27 stations provided by the World Ozone and Ultraviolet Data Center (WOUDC). Each station is depicted through a different symbol and color. Their names are given in legend.

Ground-based *in situ* observations, when dealing with atmospheric chemistry, play a crucial role in monitoring air quality and pollution events. Airborne observations, on the other hand, are central for the detection of chemical concentrations and meteorological parameters along the atmospheric column, from the surface up to 15 to 30 km height. These kind of measurements, and especially the ones obtained using instruments on board of meteorological balloons, are proven to be a good tool to explore the distribution of chemical and thermodynamic fields. They provide information necessary to better understand distribution and transport phenomena, and to analyse physico-chemical processes taking place at the UTLS and in stratosphere.

5.1.1 Ozonesondes

Meteorological services launch ozonesondes on board of meteorological balloons with an almost weekly frequency, in order to acquire *in situ* measurements of the atmospheric state. This kind of instrument can provide accurate measurements of O_3 profiles up to heights of the order of 10 hPa and with a vertical resolution of 150−200 m.

The network of meteorological ozonesondes, however, is not homogeneous across the globe. It is strongly dependent on the procedures of the meteorological sites in the measurements campaigns themselves and on the weather conditions that do or do not allow the balloons to fly efficiently. For these reasons, this network presents a geographical distribution that is

uneven and, in general, more representative in the northern mid-latitudes.

Ozonesonde data represented the most accurate measurements of tropospheric O_3 obtainable for years. Though, they served for many applications, including the validation of satellite products [e.g. Worden et al. (2008); Boxe et al. (2010); Dufour et al. (2012)] and models [e.g. Geer et al. (2006)].

Profiles of ozone concentrations acquired through ozonesondes are used in throughout the present study in order to validate the model simulations and they are those collected and distributed by the **WOUDC**[1]. An example of the localisation of the station distribution, is displayed in Figure 5.1, where the name of the stations is listed as well. Each station can provide a different amount of observations. For the period of study that has been taken into account, and that will be better detailed further in this manuscript, a total amount of 296 observed profiles is available.

5.1.2 Surface Measurements

Surface data can be exploited in order to validate the model performances at the surface.

Data have been collected from the European Environment Agency (EEA) portal[2] (European Environment Information and Observation Network, Eionet). This service gathers data from national networks of the EEA member countries, including the European Union Member States, the EEA cooperating and other contributing countries.

The database consists of air quality multi-annual time series of data and statistics computed over a number of air pollutants. Information on the monitoring networks, air quality modelling techniques, as well as definition of air quality zones, conformity and assessment regimes, air quality plans and programmes are available too.

The validated data E1a have been used for this study. These are exported from EEA database, which stores both primary validated assessment data (the E1a, indeed) and primary up-to-date assessment data (E2a). Both are reported by countries and successfully tested by automated quality control.

5.1.3 Aircraft Data

Other useful sources of atmospheric observations, of both the meteorological and chemical states, are the aircraft. This kind of acquisition network, uses predominantly existing aircraft onboard sensors to collect, process, format and transmit meteorological data to ground stations via satellite or radio links.

In this context, the European Research Infrastructure named In-service Aircraft for a Global Observing System (IAGOS), runs global observations of atmospheric composition from commercial aircraft. This service provides near real-time data for both weather prediction and

[1] `https://woudc.org`, last access: 2 October 2022
[2] `https://discomap.eea.europa.eu/map/fme/AirQualityExport.htm`, last access: 2 October 2022

air quality forecasting[3]. The basic instrument of IAGOS, is the Package1 (P1), whose main function is to perform autonomous measurement of ozone and carbon monoxide. Although not carried out in this particular trial, such data would provide an excellent source of *in situ* observations for validation [e.g. Petzold et al. (2015)].

5.2 Remote Sensing Techniques

In situ observations are extremely helpful for atmospheric sciences. However, they don't exist everywhere. Huge gaps exist between weather stations. That is where the remote sensing comes in, as it includes all the methods of measuring a certain variable at a certain distance. Remote sensing techniques can be split into *active* and *passive* according to the source of signal they sense in order to explore an object. **Active** remote sensing techniques rely on their own source of radiation. They emit an EM radiation at a certain frequency and register the radiation coming back towards the detector after interacting with the objects that modified the characteristics of the incident radiation itself. Two classes of instruments belonging to this category are: *radars*, emitting radio or MW radiation and registering the reflection occurred after its interaction with the medium sensed, and *lidars*, emitting UV/VIS and IR radiation and then assessing the diffusion undergone due to atmospheric components.

Passive sensors, on the other hand, measure the radiation from a natural external source, e.g. Sun, stars or, in the specific case of Earth's observation, atmosphere and surface. As for the passive techniques, the radiation interacts in its path from the source to the detector, thus providing information on the physical parameters of the medium considered (interaction radiation-matter detailed in Chapter 2). Passive detectors often work on signals from several spectral bands simultaneously, resulting in *multi-spectral* or *hyper-spectral* acquired data.

A very common class of passive detectors are the ***spectrometers***, i.e. instruments able to measure the spectrum of the EM radiation, or the properties of light as a function of its wavelength. Most spectrometers use the principle of interference to decompose light radiation into its wavelengths and measure its intensity with a photodiode. The best known instrument in this field is the **Michelson interferometer**, devised in 1887 by Albert Abraham Michelson, and the scheme of which is illustrated in Figure 5.2.

As it is shown, the beam of incident radiation is first split in two by a semi-transparent mirror (i.e. *beam splitter*), so that one of the new beams strikes a fixed reflective mirror, the other a movable one. The beams then reflected are brought back together and an interference pattern is measured. The optical path difference (OPD), here named δ, produces a phase shift between the two waves, allowing to obtain an interferogram $I(\delta)$. When considering a polychromatic source, the signal obtained consists of the sum of the monochromatic contributions, i.e. the integral:

$$I(\delta) = \int_0^\infty S(k)[1 + \cos(2\pi k\delta)]dk \tag{5.1}$$

[3] `https://www.iagos.org/`

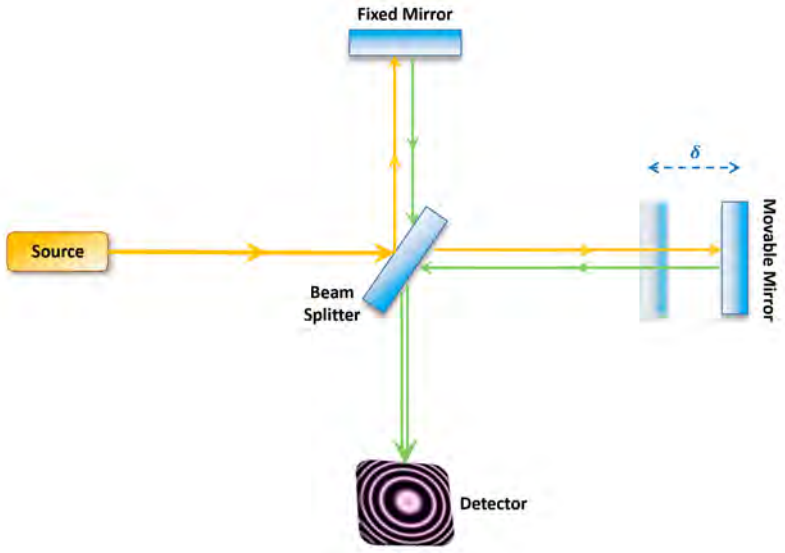


Figure 5.2: Schematic illustration of a Michelson Interferometer.

k is the *wavenumber*[4], $S(k)$ the energy characterising the radiation emitted by the source. The interferogram in Equation 5.1 consists of a constant term, plus a second one corresponding to the Fourier transform in cosine of the spectrum $S(k)$:

$$I_2(\delta) = \int_0^{\infty} S(k)[\cos(2\pi k\delta)]dk \tag{5.2}$$

The spectrum, in turn, can be obtained through the transformation:

$$S(k) = \int_{-\infty}^{+\infty} I_2(\delta)[\cos(2\pi k\delta)]d\delta \tag{5.3}$$

The instrument, therefore, registers first the interferogram as a function of δ. After that, the Fourier transform returns the spectrum as a function of wavenumber. In reality, however, the mirror displacement is limited to a maximum value δ_{max}. Consequently it is not the interferogram $I(\delta)$ that is measured, but rather its product with the rectangular function:

$$\begin{aligned} D(\delta) &= 1 \quad \text{if} \quad -\delta_{min} \leq \delta \leq \delta_{max} \\ D(\delta) &= 0 \quad \text{elsewhere} \end{aligned} \tag{5.4}$$

The source spectrum is then convoluted through cardinal sine function, that is the Fourier transform of the rectangular function in 5.4, and it is referred to as **Spectral Response Function (SRF)** of the instrument:

$$\text{sinc} = \frac{\sin(2\pi k\delta_{max})}{2\pi k\delta_{max}} \tag{5.5}$$

[4] $k = \frac{1}{\lambda}$, being λ the wavelength.

The resulting spectrum has a resolution depending on the maximum displacement of the moving mirror:

$$\Delta k = \frac{1}{2\delta_{max}} \tag{5.6}$$

In order to get rid of the secondary maxima of the *sinc* function, which degrades the signal, the so-called **apodisation** is used. This consists in performing a convolution of the SRF with an appropriate function that smooths, or completely cuts off, its sidelobes, containing indeed the secondary maxima. Many apodizing functions have been reported in the literature and are currently exploited, among others *Gauss*, *Hamming*, *Norton-Beer*, *Blackman-Harris* [see e.g. Filler (1964); Norton and Beer (1976); Norton and Beer (1977); Parker et al. (1991); Naylor and Tahic (2007); Han et al. (2015); Martin et al. (2016)]. Please notice that this apodization operation, although degrading the resolution itself, still remains beneficial.

5.3 Satellite Observations

Remote sensing in the modern sense of the term begins in the second half part of the last century, when airborne radars modified for ground meteorological use entered the forecasting practice. The greatest innovation, however, came with the launch of meteorological satellites. These could cover much more land surface than planes and also monitor areas on a regular basis.

Meteorological satellites can cover different kind of orbits. This depends on their use and the type of measurements they are designed to perform. The satellite data treated along this study are issued from instruments on board of *polar-orbiting* and *geostationary* satellites.

- **Polar-orbiting** satellites usually circle around Earth in a direction north to south passing over the poles, or close to them. Actually, an orbit, in order to still be defined as a polar one, can have a deviation from the poles up to $20 - 30°$. A full rotation takes about 90 minutes to be performed. This kind of orbit is fairly low and it takes place at a distance from the surface of approximately 200 to 1000 km. A particular subcategory of polar orbits is the **Sun-synchronous orbit (SSO)** (Figure 5.3a). Satellites in SSOs, being indeed synchronous with the Sun, are designed to always point to the same spot at the same local time. This can be very useful for a wide range of applications, since it can monitor a specific area by acquiring images of it across the instrument life-time.

- Satellites in **geostationary orbit (GEO)** occupy an orbit above the Equator, as illustrated in Figure 5.3b, which follows the Earth's rotation from west to east. It takes 23 hours 56 minutes and 4 seconds to complete an orbit since they travel at the same exact rate as Earth. For this reason, satellites in GEOs appear to be "stationary" over a fixed position. Moreover, since they need to move at the same speed as Earth, they need to be located at an altitude of 35 786 km. They are, therefore, very far from the Earth compared to satellites on other kind of orbits. Satellites in geostationary orbit

(a) **(b)**

Figure 5.3: Illustration of **(a)** polar Sun-synchronous orbit (SSO) and **(b)** geostationary orbit (GEO) [source ESA website]. For both orbits, satellite flight heights are indicated.

are able to cover a large portion of Earth at once. So just a few of them can provide a global coverage.

Currently, satellite data allow to monitor the overall chemical composition and meteorological state of the atmosphere with a very high precision and at time frames that become shorter as the technology develops. Atmospheric composition can be sensed using either the VIS-IR, the UV-VIS or the IR. Space agencies from all over the world own, produce or operate their missions in order to have the necessary data for research and operations in the field of atmospheric chemistry-physics.

Among the instruments working in the the visible-infrared range, MOPITT[5], on board the NASA[6] Terra, and TANSO[7], flying on the Japanese JAXA[8] satellites, have to be mentioned.

Among the main instruments exploiting ultraviolet-visible radiation, and in SSO, are OMI[9] better described later in this Chapter (5.3.1), OMPS[10] on board the NOAA[11] satellites and the European TROPOMI[12] on the ESA[13] Sentinel-5 Precursor. ESA, indeed, plays a primary role in planning and implementing missions for the observation of Earth from space. In Figure 5.4 is shown a summary of this and other past, present and future ESA missions, developed in collaboration with the European Union, i.e the Copernicus programme, and EUMETSAT.

Still operating in the UV/VIS, but on GEO platforms are GEMS[14], a grating imaging

[5]Measurement Of Pollution In The Troposphere (MOPITT)
[6]National Aeronautics and Space Administration (NASA)
[7]Thermal And Near infrared Sensor for carbon Observations (TANSO)
[8]Japan Aerospace Exploration Agency (JAXA)
[9]Ozone Monitoring Instrument (OMI)
[10]Ozone Mapping and Profiler Suite (OMPS)
[11]National Oceanic and Atmospheric Administration (NOAA)
[12]Tropospheric Monitoring Instrument (TROPOMI)
[13]European Space Agency (ESA)
[14]Geostationary Environmental Monitoring Spectrometer (GEMS)

spectrometer for monitoring ozone and other species, flying on board the Korean GEO-KOMPSAT-2B satellites [e.g. Baek et al. (2022)], and the NASA's TEMPO[15] [Zoogman et al. (2017)] tracing air quality over North America.

Among the hyperspectral IR instruments in SSO, one can list CrIS[16] on board the NOAA satellites [e.g. Bloom (2001); Han et al. (2013); Shephard and Cady-Pereira (2015)], TES[17] on board the NASA Aura [Beer (2006)], HIRAS[18] on board the Chinese Feng-Yun - 3, and the Russian IKFS-2[19]. Finally, definitely worth mentioning is the IASI interferometer, that will be described in more detail later in this Chapter (5.3.2) and that provides accurate measurements of the meteorological and chemical state of the Earth's atmosphere. IASI flies on board the polar-orbiting Meteorological Operational Satellite (Metop) series, developed by ESA and operated by the EUMETSAT agency.

However, satellite instruments in polar orbit are able to provide repeated measurements on the same point on Earth only a few times during the same day. Having thermal infrared measurements coming from a spectral sounder on board a GEO platform can have significant advantages in the area of atmospheric chemistry monitoring. GIIRS[20] is the first Chinese hyperspectral infrared sounder on board a geostationary satellite, namely the FengYun-4 series launched in 2016 (FY-4A) and 2021 (FY-4B) [Yang et al. (2017)]. With its hyperspectral coverage [LWIR from 680 to 1130 cm^{-1} and MWIR from 1650 to 2250 cm^{-1} with a spectral resolution of 0.625 cm^{-1}, it is able to provide important data for monitoring CO [Zeng et al. (2022)] or, for instance, the NH_3 cycle [Clarisse et al. (2021)].

The mission that marked the breakthrough in European geostationary satellite observations of the atmosphere is the Meteosat series. Since 1977 these satellites have been providing crucial data for atmospheric monitoring. The last satellite of the first generation of Meteosat (i.e. Meteosat-7) was operational until 2017. The continuity of the service was ensured by Meteosat Second Generation (MSG), established under cooperation between EUMETSAT and ESA. Meteosat -9, -10-, 11 are currently operational. In turn, further innovation will soon be brought by the third generation **MTG**. When in full operation, the MTG constellation will consist of two imaging satellites (MTG-I), working in tandem, and a sounding satellite (MTG-S), planned to be the first operational sounding satellite in a geostationary orbit. MTG-I will carry the Flexible Combined Imager (FCI) and the Lightning Imager (LI). MTG-S, on the other hand will carry the Infrared Sounder (IRS) and the Copernicus Sentinel-4 instrument.

[15] Tropospheric Emissions: Monitoring of Pollution (TEMPO)
[16] Cross-track Infrared Sounder (CrIS)
[17] Tropospheric Emission Spectrometer (TES)
[18] Hyperspectral Infrared Atmospheric Sounder (HIRAS)
[19] Infrared Fourier Spectrometer - 2 (IKFS-2)
[20] Geostationary Interferometric Infrared Sounder (GIIRS)

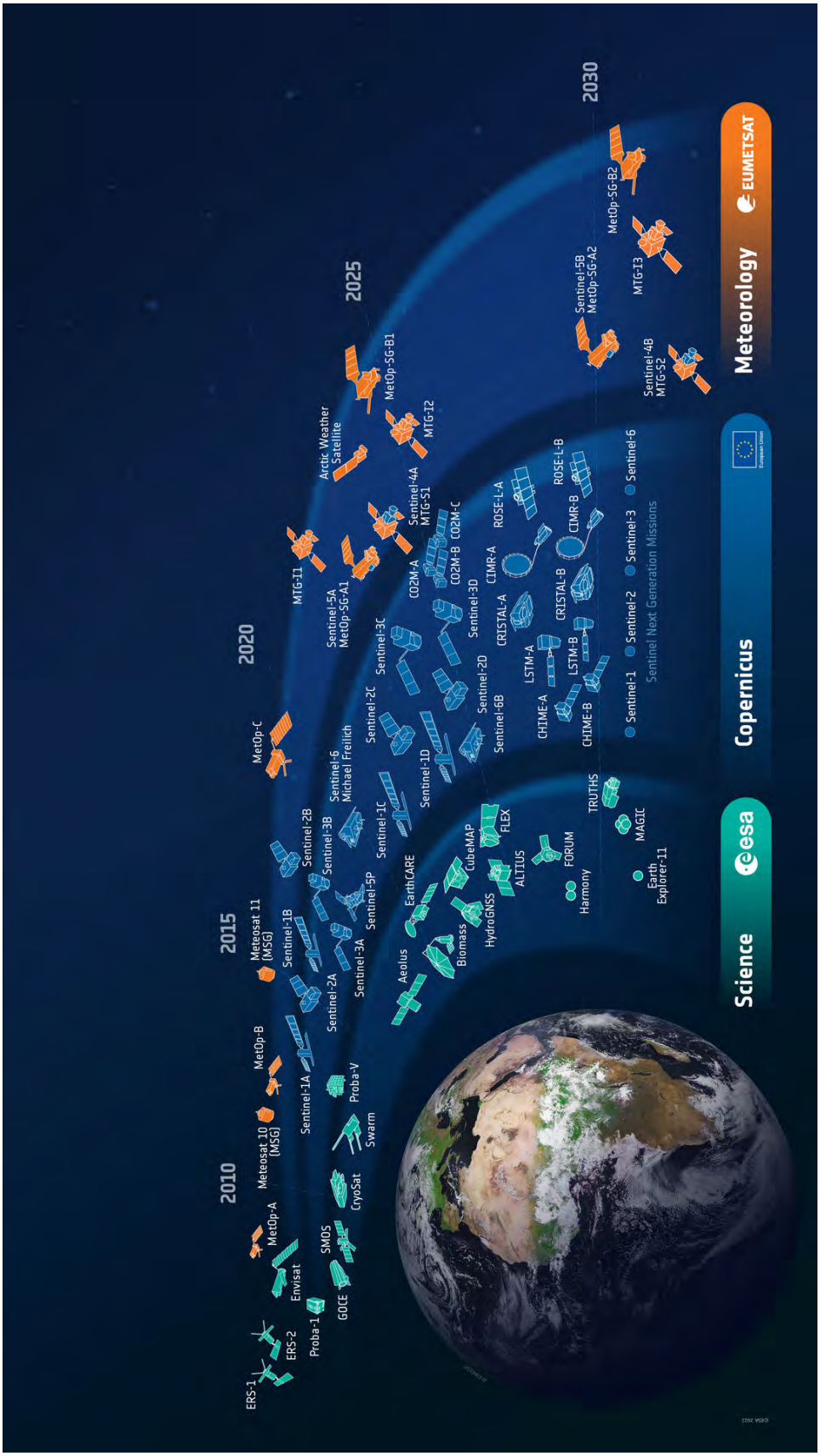


Figure 5.4: ESA-developed Earth observation missions [source `https://www.esa.int/`].

5.3.1 Ozone Monitoring Instrument (OMI)

The OMI is a UV/VIS nadir solar back-scatter spectrometer [Levelt et al. (2006)], flying since 2004 on board the NASA Aura spacecraft. It provides, from a polar Sun-synchronous pattern, a nearly global daily coverage. It is widely used to study Earth's ozone, air quality, and climate and to perform research on the composition, chemistry, and dynamics of the atmosphere. OMI is indeed able to measure trace gases, aerosols, cloud pressure and coverage with a footprint of $13 \times 24\,\mathrm{km}^2$ at nadir. In addition to monitoring the ozone layer, the data it collects are used with success in a wide range of research and operational applications, such as pollutants, air quality and climate change [e.g. Levelt et al. (2018)].

Daily ozone total columns recorded by OMI are used in this study to have a satellite-borne validation for ozone and are collected from EOSDIS NASA portal[21].

5.3.2 Infrared Atmospheric Sounding Interferometer (IASI)

The hyper-spectral IASI is a Fourier-Transform Spectrometer (FTS) and one of the main payloads of the polar-orbiting Metop series operated by EUMETSAT [Siméoni et al. (1997); Blumstein et al. (2004)].

The first IASI instrument has been launched on board the Metop-A satellite in 2006. Two supplementary instruments followed on board Metop-B (launched in 2012) and Metop-C (2018). In the time period ranging from July 2019 until the retirement of Metop-A in November 2021, the three instruments provided operational exploitable data simultaneously, with a nearly global coverage. The continuity of the IASI service will be assured by the two remaining instruments still on orbit till the launch of the IASI - New Generation (IASI-NG) planned for 2025 on board the Metop - Second Generation (Metop-SG), which will present improved characteristics compared to its predecessors [see e.g. Bermudo et al. (2014); Crevoisier et al. (2014); Vittorioso et al. (2021)].

The IASI mission was originally meant to profile atmospheric temperature and humidity for weather forecasting applications mainly, but also for measuring atmospheric compounds like ozone. It showed over time its high potential to be exploited in monitoring and forecasting atmospheric chemical composition and for climate studies too. Thanks to its fine spectral resolution, signal-to-noise ratio and wide spectrum range, it is a precious resource for detecting trace gases like ozone, methane and carbon monoxide, as well as clouds, aerosols and greenhouse gases [Phulpin et al. (2002); Clerbaux et al. (2009); Hilton et al. (2012)].

[21] `https://acdisc.gesdisc.eosdis.nasa.gov/data/Aura_OMI_Level3/OMTO3e.003/`, last access: 2 October 2022.

IASI observes the Earth with a scan geometry that is sketched in Figure 5.5. The angle of acquisition is of about 48° on both sides of the satellite track. This produces 15 views each side. Each field of view (FOV) at nadir is of 50 × 50 km, while the total swath goes up to 2 × 1100 km (considering the optical deformation at the sides of the swath). Each FOV consists of 4 circular pixels with a 12 km diameter on the ground at nadir [Clerbaux et al. (2009)].

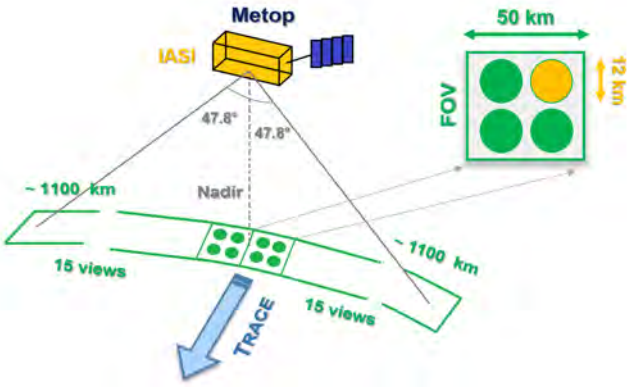


Figure 5.5: IASI scan geometry. The swath width is of about 2200 km, with a FOV of 50 km at nadir and 4 footprints of 12 km diameter each [adapted from Vittorioso et al. (2021)].

The instrument acquires spectra of atmospheric emission within 645 and 2760 cm^{-1} (3.6 - 15.5 µm), with a spectral apodized resolution of 0.5 cm^{-1} and a 0.25 cm^{-1} spectral sampling. Consequently, it measures at 8461 wavelengths (or channels). An example of spectrum that IASI acquires is provided in Figure 5.6 in brightness temperatures, where the sensitivity to the main chemical species is highlighted as well as the division in three bands: Band 1 (645 − 1210 cm^{-1}), Band 2 (1210 − 2000 cm^{-1}), Band 3 (2000 − 2760 cm^{-1}). The major sensitivities of the spectra IASI acquires are also listed in Table 5.1 together with possible applications of these properties.

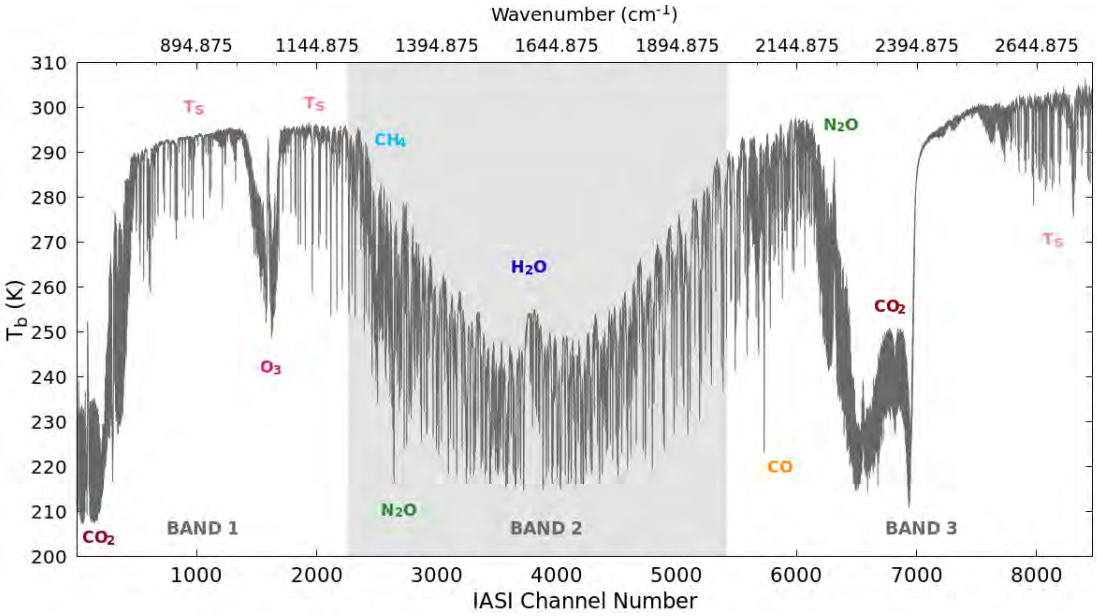


Figure 5.6: Example of a IASI spectrum in brightness temperature. This is split into three bands, which are here illustrated together with the sensitivity to different chemical species of each spectral area. Notice that T_s stands for Surface Temperature and it is used to highlighted atmospheric windows.

Name	Spectral Region (cm^{-1})	Sensitivity	Application
R1	650 – 700	CO_2 absorption	Temperature profile
R2	790 – 980	Atmospheric window	Surface temperature and clouds
R3	1000 – 1070	O_3 absorption	Ozone sounding
R4	1080 – 1150	Atmospheric window	Surface temperature and clouds
R5	1210 – 1650	H_2O absorption	Water vapour profile - N_2O and CH_4 amount
R6	2100 – 2150	CO absorption	CO concentration
R7	2150 – 2250	N_2O and CO_2 absorption	Temperature profile - N_2O concentration
R8	2350 – 2420	CO_2 absorption	Temperature profile
R9	2420 – 2700	Atmospheric window	Surface temperature and clouds
R10	2700 – 2760	CH_4 absorption	CH_4 concentration

Table 5.1: Main sensitivities and applications of the sub-bands in spectrum IASI can cover [from Cayla (2001)].

The data acquired by IASI can be classified into categories, called “Levels", according to the processing they have been through[22]:

- **Level 0 (L0)**: raw spectra (Fourier transform, radiometric calibration, band merging);
- **Level 1 (L1)**: spectra are calibrated by applying an inverse Fourier transform;
  - L1a: Further radiometric corrections are applied;
  - L1b: L1a spectra are re-sampled at a constant wavelength and interpolated into a new spectral grid;
  - L1c: apodized L1b spectra (Gaussian apodization);
- **Level 2 (L2)**: retrieval methods are applied in order to obtain products in the form of atmospheric profiles (e.g. temperature, ozone, etc.).

During preparatory studies for the present project, which will be shown in Chapter 6 (6.2.2), IASI Level 1 (L1)c data have been exploited. These were gathered on the EUMETSAT Earth Observation Portal[23] in a NetCDF format including the full spectral range. A pre-treatment of the data was performed, but it will be illustrated later in Chapter 6, when explaining how the data were utilised.

[22] `https://www.eumetsat.int/media/44030`
[23] `https://eoportal.eumetsat.int`, last access: 2 October 2022

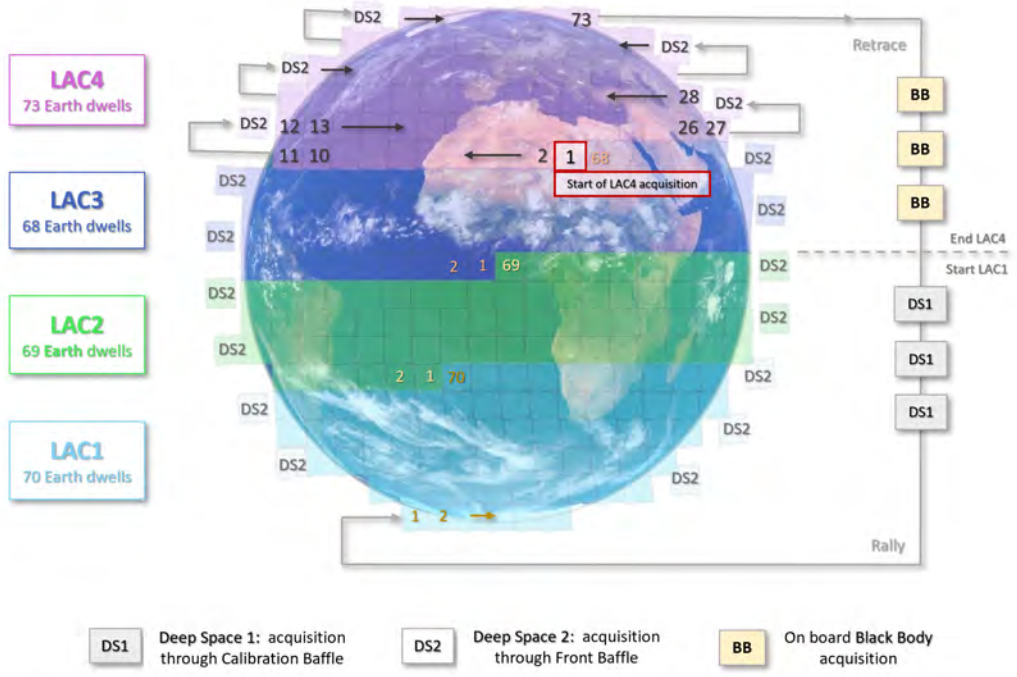


Figure 5.7: IRS Local Area Coverage (LAC) zones and dwells coverage. The geometry of acquisition is suggested starting from the first dwell in LAC1.

5.4 Infrared Sounder (IRS)

5.4.1 Instrument Characteristics

The information about IRS reported in this subsection are issued from EUMETSAT, Thales Alenia Space and Coopmann et al. (2022).

IRS is a FTS, built by Thales Alenia Space, which will be launched on board the geostationary MTG-S in 2024. After the instrument will be operational, it will provide data of the full Earth disk with a 4 km spatial sampling at nadir.

The IRS scanning sequence will firstly require the Earth disk to be split in four **Local Area Coverage (LAC)** zones, as in Figure 5.7. Each LAC will be covered in 15 minutes through the acquisition of successive stares, called **"dwells"**, in about 10 seconds each. The total amount of Earth dwells for the whole disk will be 280, while the number per LAC is listed in Figure 5.7. LAC4 will be covered every 30 minutes, while the other LACs will be acquired in between. The sequence of acquisition will start with three times (LAC3,LAC4), followed by three times (LAC2,LAC4), then again (LAC3,LAC4) for three times, and, in the end, three times (LAC1,LAC4). The cycle will, then, start over every 6 hours.

Each dwell will consist of 160×160 pixels, each one presenting a 4 km resolution at nadir (as already mentioned).

The sounder will cover **1960 channels** spread on two bands in the thermal infrared:

- **Band 1** in **LWIR**, bounded in the range 679.70− 1210.44 cm^{-1} and with a spectral

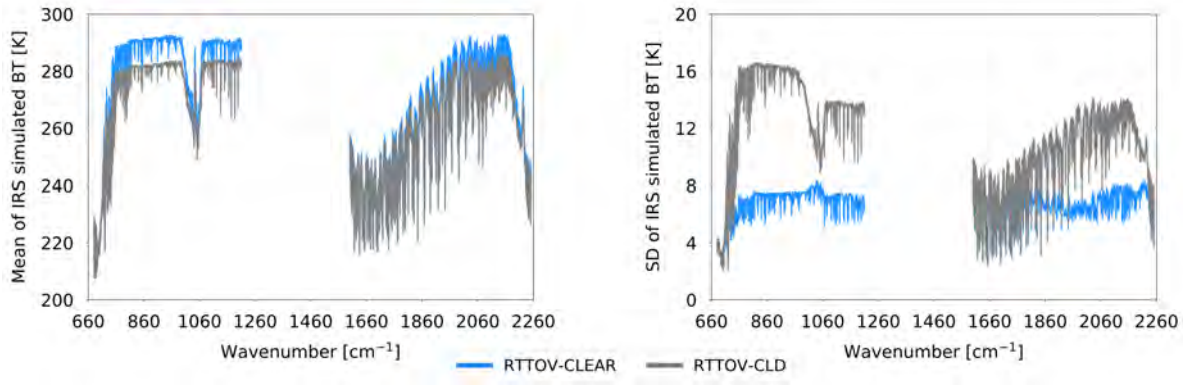


Figure 5.8: Example of spectrum that IRS will provide, simulated in brightness temperature by Coopmann et al. (2022) using RTTOV version 12 for both cloudy (gray) and clear-sky (right) conditions. Mean (left) and standard deviation (right) are computed over their 7458 case study.

sampling of $\sim 0.603\,\mathrm{cm}^{-1}$: 881 channels;

- **Band 2** in **MWIR**, in the range $1600.00-\ 2250.20\,\mathrm{cm}^{-1}$ and spectral sampling of $\sim 0.604\,\mathrm{cm}^{-1}$: 1079 channels.

Channel spacing is obtained according to Equation 5.6. The difference in the spectral sampling is due to the on-ground maximum optical depth (δ_{max}), which is slightly different for the two bands: 0.829 038 cm for Band 1 (LWIR) and 0.828 245 cm for Band 2 (MWIR). The central wavenumber is computed as follows:

$$w_n = \frac{\mathrm{N}}{2\delta_{max}} \tag{5.7}$$

with $1127 \leq \mathrm{N} \leq 2007$ for the LWIR and $2650 \leq \mathrm{N} \leq 3728$ in the MWIR. Following this set of information, which had been provided via personal communication by B. Theodore (EUMETSAT) to the RTTOV team, the RTTOV coefficients for IRS have been built and are those exploited throughout this work.

The instrument noise, provided by EUMETSAT in terms of Noise Equivalent of Differential Temperature (NEΔT) at 280 K, is illustrated in Figures 5.9a (red curve). This noise can be converted to the corresponding scene temperature by using the formula:

$$\mathrm{NE}\Delta\mathrm{T}(\mathrm{T_b}) = \mathrm{NE}\Delta\mathrm{T}(280\,\mathrm{K})\frac{(\partial B_\nu/\partial T)(280\,\mathrm{K})}{(\partial B_\nu/\partial T)(\mathrm{T_b})} \tag{5.8}$$

Values of the noise in Radiance Units are in Figure 5.9b.

An example of spectrum that IRS will be able to provide is shown in Figure 5.8. It has been computed in the study by Coopmann et al. (2022) about the application of future IRS data in the NWP field, using RTTOV version 12 in both cloudy and clear-sky conditions, over 7458 case studies. Figure shows mean and standard deviation for both IRS bands.

IRS spectra, will be distributed in the form of Principal Component (PC) scores. Its 1960

channel acquisitions will be compressed in to 300 PC scores preserving information content for near-real time applications. In the present work, however, this step will be not mimicked.

The classification of spectra into "Levels" will be slightly different compared to what is done for IASI:

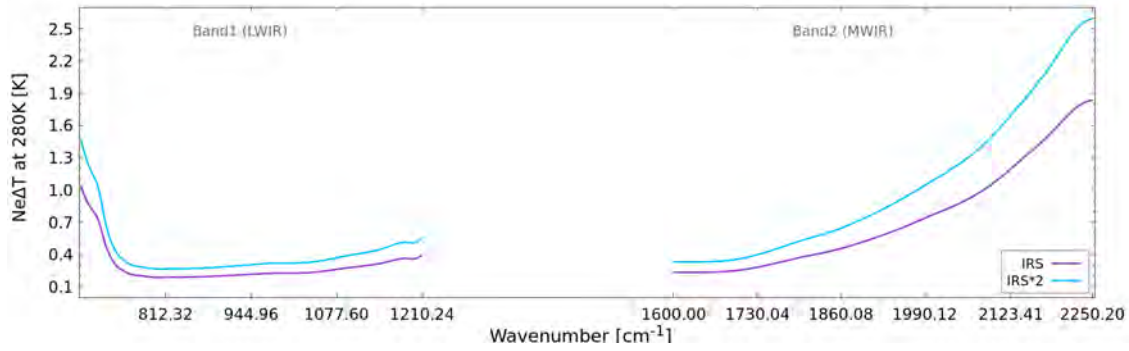
- **L0**: compressed interferogram;
- **L1**:
  - L1a: raw spectra after Fast Fourier Transform (FFT);
  - L1b: radiometric and spectral corrections are applied and spectra are geolocated on Earth;
  - L1c: apodized L1b spectra;
- **L2**: retrieval methods are applied in order to obtain products in the form of atmospheric profiles.

5.4.2 IRS with Increased Resolution

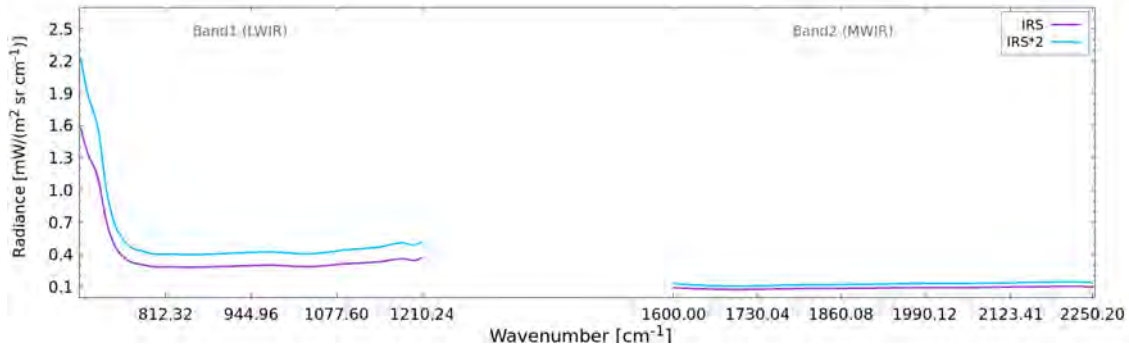
As already introduced, another version of IRS is supposed to be evaluated through the project, and this will often be referred to as **"IRS*2"** throughout the manuscript. IRS*2 is still being designed, but it would be characterised by a spectral sampling of $0.3251\,\mathrm{cm}^{-1}$. By covering the same spectral bands as IRS, this would provide a total of **3633** different **channels**, or wavelengths:

- **Band 1** in **LWIR**, bounded in the range $680.00 - 1210.24\,\mathrm{cm}^{-1}$: 1632 channels;
- **Band 2** in **MWIR**, in $1600.00 - 2250.20\,\mathrm{cm}^{-1}$: 2001 channels.

Although it would potentially offer an almost twice thinner spectral sampling compared to IRS, IRS*2 should present a stronger instrumental noise, i.e. by a factor close to $\sqrt{2} = 1.41$ (an example of the differences in the instrument noise for the two versions are shown in Figures 5.9, displayed in both NeΔT and radiance units). This is due to the way the instrument is being designed. To increase the spectral sampling, only the half of the interferogram is acquired. Since the interferogram is expected to be symmetric around the zero path difference (ZPD), only the part from ZPD to the maximum OPD is acquired. The other part (from ZPD to minimum OPD) is built using specific ground processing accounting for dedicated characterisation (Figure 5.10).



(a)



(b)

Figure 5.9: Instrument noise in NEΔT at 280 K **(a)** and radiance units **(b)** for IRS and for an IRS with an increased resolution (i.e. IRS*2), respectively in purple and cyan.

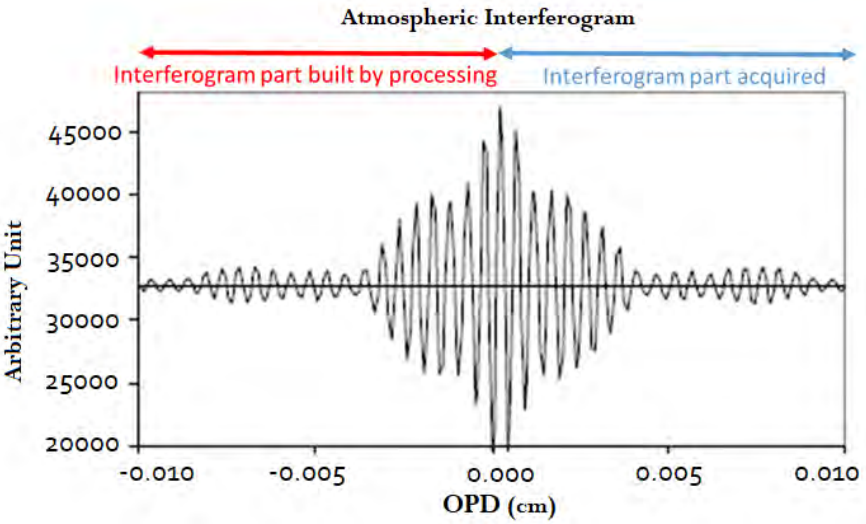


Figure 5.10: Diagram illustrating the construction of an interferogram for IRS*2. OPD stands for Optical Depth Difference.

5.4.3 Sensitivity Study

IRS characteristics can be essential in deducing information about atmospheric composition. Looking at the spectral bands that the instrument will cover, and cross-referencing with the properties of the thermal IR spectrum (see e.g. Figure 5.6 and Table 5.1), it is clear that IRS will be sensitive to different chemical species. These features, combined with the geostationary acquisition of large latitude bands within short time frames, could provide essential and detailed information for the monitoring of certain atmospheric pollutants. More into detail, the following areas can be identified:

- LWIR:
  - $680 - 770\,\mathrm{cm}^{-1}$: CO_2 absorption
  - $790 - 980\,\mathrm{cm}^{-1}$: atmospheric window
  - $1000 - 1070\,\mathrm{cm}^{-1}$: O_3 profiles
  - $1070 - 1200\,\mathrm{cm}^{-1}$: SO_2 absorption
  - $1080 - 1150\,\mathrm{cm}^{-1}$: atmospheric window
- MWIR:
  - $2100 - 2150\,\mathrm{cm}^{-1}$: CO absorption
  - $2150 - 2250\,\mathrm{cm}^{-1}$: CO_2 and N_2O absorption bands
    - CO predominates between $2150 - 2180\,\mathrm{cm}^{-1}$
    - N_2O between $2180 - 2240\,\mathrm{cm}^{-1}$
    - CO_2 between $2240 - 2250\,\mathrm{cm}^{-1}$ [Ricaud et al. (2009)].

Despite this *a priori* knowledge about the potentiality of the instrument, prior sensitivity studies should be carried out in order to establish and document the selection process of the spectral sub-sets of wavelengths to work on. As IRS is not yet operational, in fact, the research will require simulation of data using models, either RTM or CTM. Each simulation, however, presents a certain computational cost and it is time-consuming. Therefore, a band pre-selection is necessary. Moreover, the difference in response of the two version of the instrument evaluated throughout the work, needed to be tested. The sensitivity though, has been studied also as a function of the two different IRS resolutions considered throughout the study.

The so-called ***physical selection method***, suggested by Gambacorta and Barnet (2012), was applied. This consists in adding (or less commonly, subtracting) from a profile of concentration of chemical species, or thermodynamic parameters (e.g. temperature), a given perturbation, which can be a fixed value or a percentage of the profile itself. The spectral responses to these perturbations will then be evaluated in order to assess the sensitivity to the species of interest.

For this short part of the work, IRS spectral radiances will be simulated through the use of RTTOV v12, to which atmospheric profiles for the species to be assessed must be provided.

5.4.3.1 Sets of Profiles

The chemical species of primary interest for the present work, as already mentioned, is the ozone. However, with the aim of paving the way for possible future works, the carbon monoxide was also evaluated, together with temperature, humidity and surface temperature.

Profiles of these quantities have been gathered from the **diverse profile data sets on the CAMS atmospheric composition forecasting system**, provided by the NWP SAF[24]. The database, which is compiled by ECMWF using short-range forecasts from the operational version of CAMS, provides a concise but comprehensive description of key parameters in the operational system [Eresmaa and McNally (2016)] intended for any user who wants to carry out 1D studies. Profiles of mixing ratio for a concise series of reactive gases are extracted from the full list covered by CAMS. These selected species are: NO_2, SO_2, HCHO and, of course, O_3 and CO. Profiles of temperature (T) and specific humidity (q) are provided too.

The profiles are given in 60 vertical levels ranging from the surface, up to 0.1 hPa. To have the seasonal variability represented, the database contains profiles retrieved from model forecasts over a one-year period (starting on 9 November 2015). Moreover, the forecast steps included are 12, 18, 24, and 30 hours in order to cover diurnal variations too.

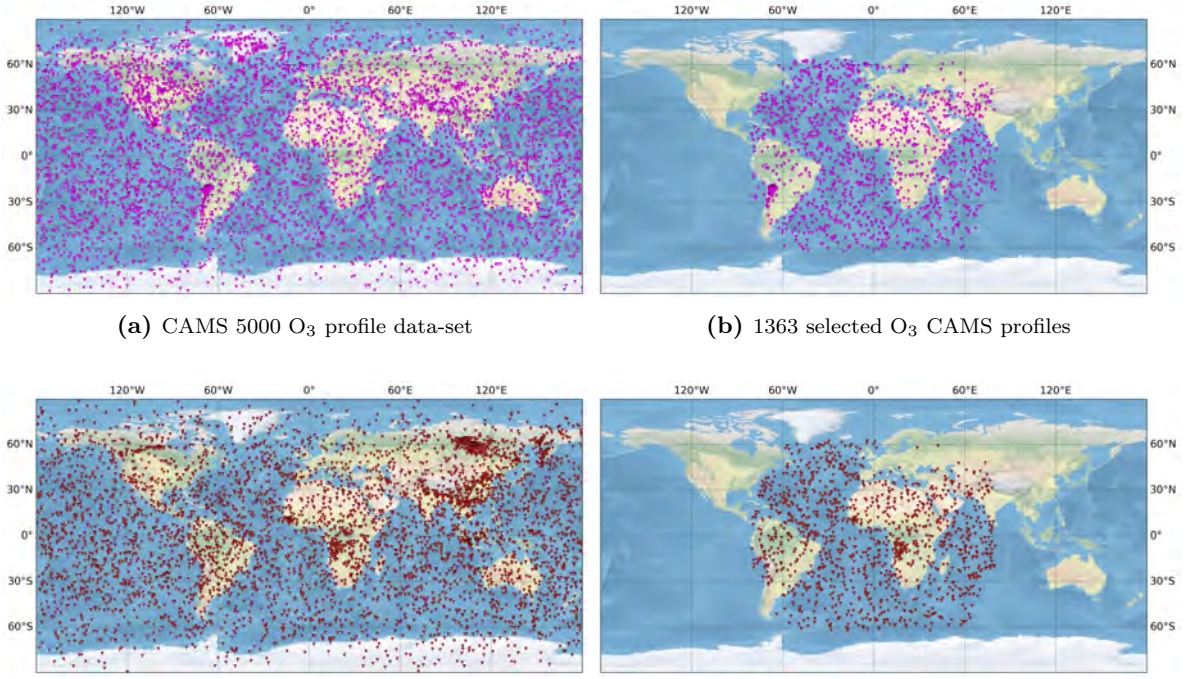
As explained by Eresmaa and McNally (2016), the whole database contains 40000 profiles, which are spread into 8 groups[25] of 5000 profiles each. The profiles composing each subset are selected separately based on the variability of one of the considered quantities (a gas or T or q). However, each subset, although selected by focusing on the variability of single parameter, also contains information on all the others, with the same geo-localization as the parameter which guided the selection. In other words, if considering, for instance, the group of 5000 profiles selected based on ozone, it will also contain information on T, q and the other four gases, all at the same points. However, when moving to the group of 5000 profiles chosen based on the carbon monoxide variability, this will also contain ozone profiles, but they will not be the same as the ozone group ones.

For the present sensitivity study, two different groups have been used, namely the O_3-based and the CO-based. This choice was made in order to ensure the best possible representation of the species of interest when assessing the sensitivity of the instrument to them. This implies, however, that the temperature and humidity profiles, and the skin temperature values considered, are not the same for the two groups.

For each of the two groups, this author performed a thinning on the number of profiles. The criteria applied in order to choose the profiles to keep were the same for both groups:

[24] `https://nwp-saf.eumetsat.int/site/software/atmospheric-profile-data/`

[25] one group for each of the 7 variables (NO_2, SO_2, HCHO, O_3, CO, T and q), plus a group of randomly distributed profiles



(a) CAMS 5000 O_3 profile data-set

(b) 1363 selected O_3 CAMS profiles

(c) CAMS 5000 CO profile data-set

(d) 1245 selected CO CAMS profiles

Figure 5.11: Location on the globe of the CAMS data-sets for the ozone **(a)** and the carbon monoxide **(c)**. The sub-sets of profiles selected for the sensitivity study are also shown in **(b)** and **(d)**.

- Spectra acquired by an instrument working on the whole Earth disk, centred around 0° of longitude, had to be simulated. Consequently, in order to avoid excessive deformation due to too large acquisition angle, only profiles between 63°N and 63°S latitudes have been retained.
- Only profiles completely over land or sea were chosen. Profiles labelled with a hybrid flag were excluded.
- Profiles had to be co-localised to IRS pixels. Only profiles presenting a distance from the centre point of an IRS pixel less then 1.5° have been retained, and then re-associated to that new couple of coordinates.

Please notice that all information about geo-points of IRS pixels or geometry, have been obtained from personal communication with Dr Olivier Coopmann, previously on a EUMETSAT fellowship.

Through this selection, 1363 profiles for the O_3 sub-set and 1245 for the CO one have been retained. The geo-localization of each profile in the subsets, before and after the thinning, are shown in panels of Figure 5.11. In Figure 5.12, on the other hand, the selected profiles are depicted as a function of the altitude, together with the corresponding temperature and specific humidity profiles. In both cases, the species of interest show a good variability in terms of intensity and diversity of profiles. This should ensure a fair reliability of the sensitivity study that followed.

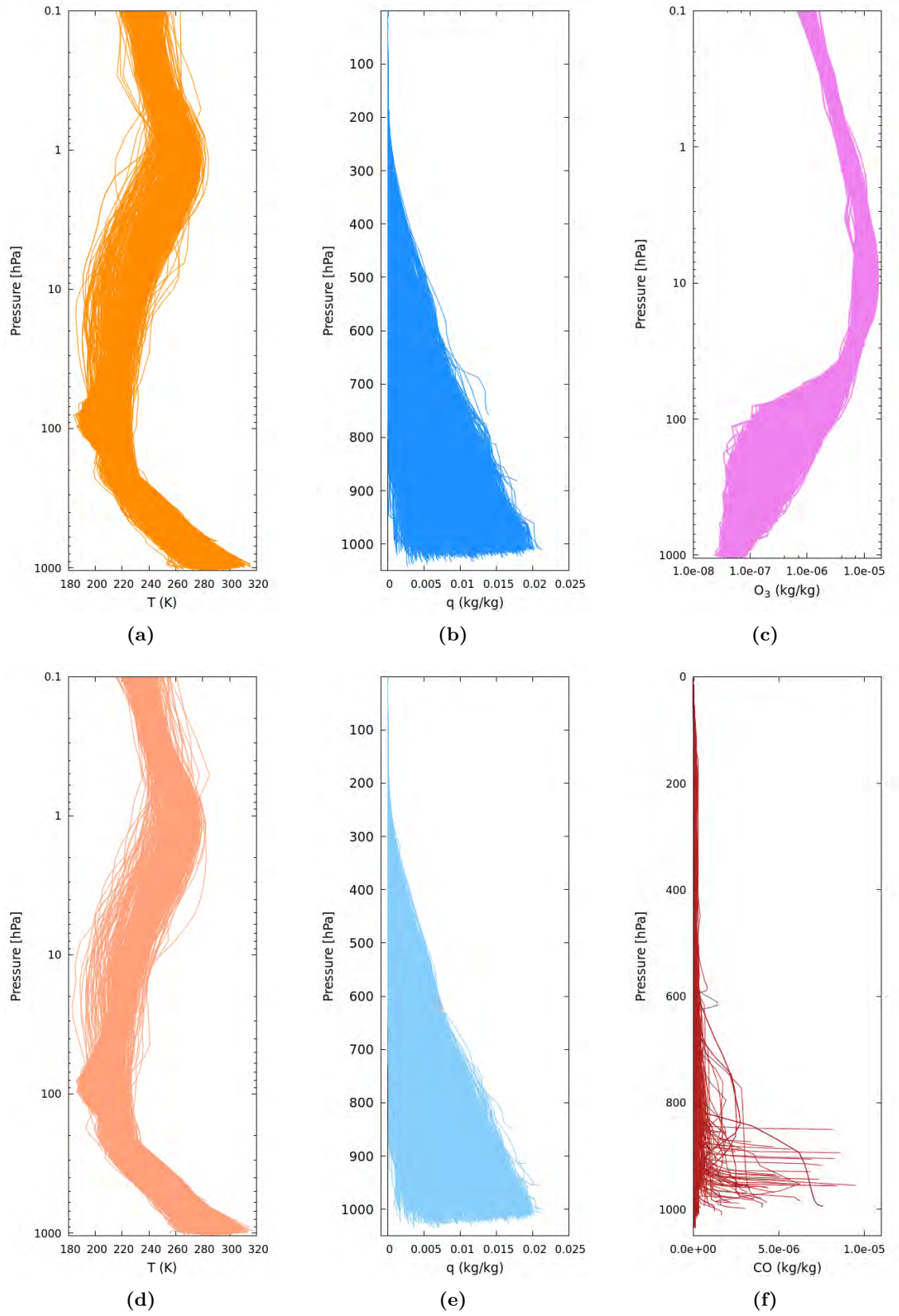


Figure 5.12: 1393 profiles of temperature **(a)**, specific humidity **(b)** and ozone (in orizontal log scale) **(c)** selected from the 5000 CAMS data-set ozone-based and 1245 profiles from the CO-based CAMS data-set for temperature **(d)**, humidity **(e)** and carbon monoxide **(f)**.

5.4.3.2 Implementation of the Physical Selection Method

The aforementioned *physical selection method* has, then, been implemented in order to assess the sensitivity of IRS, in both its versions, to ozone and carbon monoxide.

Going into more details, the procedure followed is schematically depicted in Figure 5.13 and it has been separately applied to the two groups of profiles described above.

- At a **first stage**, the profiles of temperature (T), humidity (q), ozone (or carbon monoxide) and skin temperature (T_s) values were fed into RTTOV, so that it could produce simulated spectra. To do so, RTTOV is also provided with coefficients specifically created to characterise the instrument of interest, namely IRS. The same was done in parallel for IRS*2, using its specific coefficients. The solar angles necessary for the RTM have been computed from date and time of the day, while the satellite angles were those provided by Dr Coopmann for each IRS LAC and dwell. For both instrument simulations, the outputs are ***"un-perturbed spectra"*** for each variable.

- In a **second stage**, the same process was reproduced. However, one variable at the time was formerly perturbed, while keeping the other quantities unchanged. The perturbation applied was typical of the atmospheric variability of the variable itself, i.e.:

  $\rightarrow$ $T = T + 1\,K$

  $\rightarrow$ $T_s = T_s + 1\,K$

  $\rightarrow$ $q = q + 10\%q$

  $\rightarrow$ $O_3 = O_3 + 10\%O_3$

  $\rightarrow$ $CO = CO + 1\%CO$

  These perturbations are those usually applied in literature for this kind of study [see e.g. Gambacorta and Barnet (2012); Martinet et al. (2014); Coopmann et al. (2022)]. Please notice that the perturbation applied to the CO is smaller than for the other gasses since it reflects its smaller variability in atmosphere (as seen Chapter 1). RTTOV, this time, will provide in output, for each variable perturbed in input, the corresponding ***"perturbed spectra"***.

- At the **third stage**, the final one, an average of the un-perturbed spectra was performed. The same is done for the perturbed ones (for each variable). The differences between the averaged spectra has, hence, highlighted the sensitivities to the concerned variables.

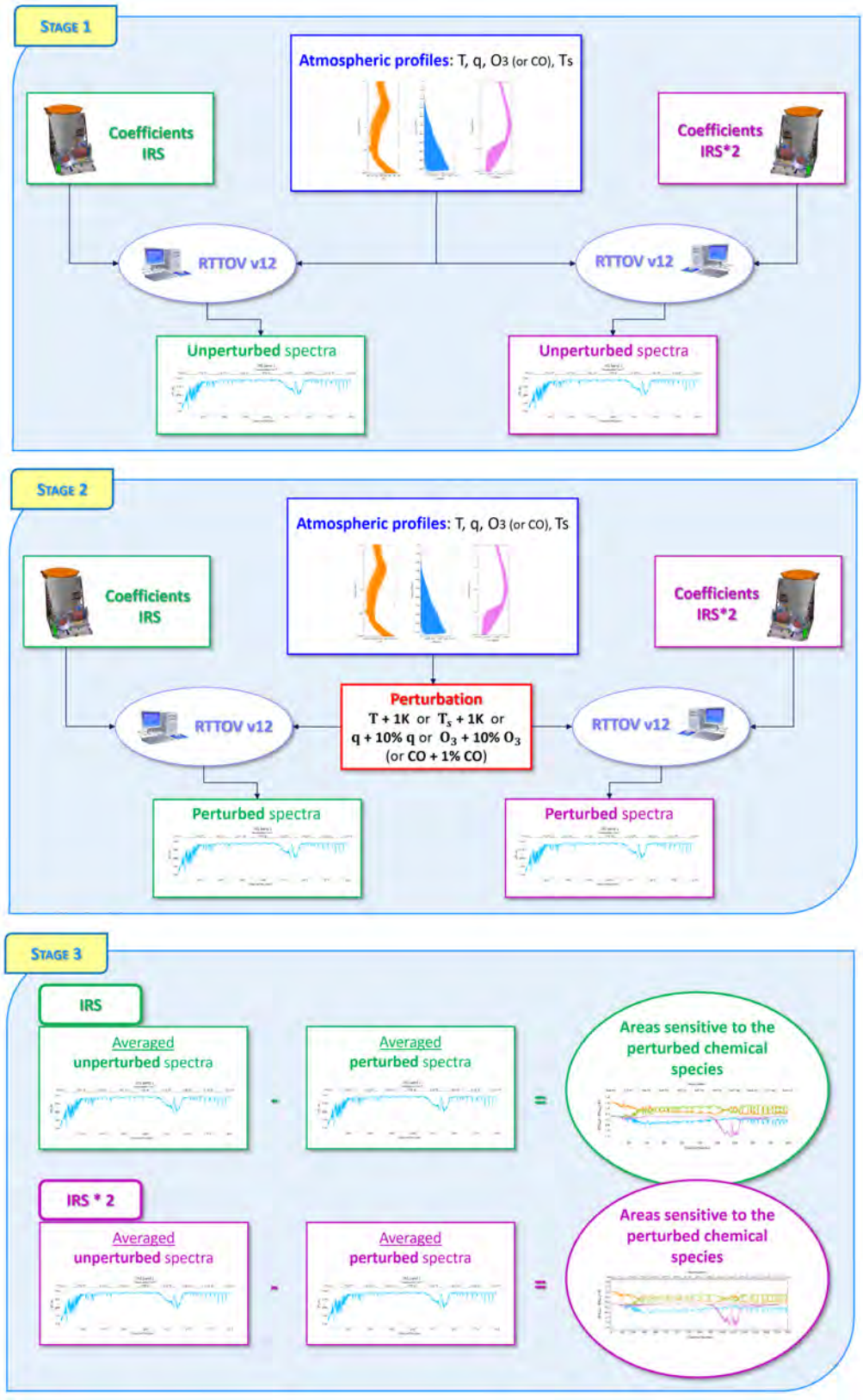


Figure 5.13: Scheme of the implementation of the sensitivity study, for both IRS and IRS*2, through the physical selection method.

5.4.3.3 Results

In Figure 5.14 the results obtained using the ozone data-set of 1363 profiles are shown. Two main areas can be observed that present sensitivity to the ozone. The first one is in the LWIR and it extends from about 930 cm^{-1} up to the end of the band. The other is found towards the end of the MWIR band and it ranges approximately between 2030 to 2150 cm^{-1}, with a very small perturbation extending to the right edge of the band. However, the former shows an intensity that goes up to values almost three times larger than the latter. The maximum variation reaches values of 1.4 K intensity in Band 1, while the response in Band 2 presents peaks with a maximum amplitude of 0.5 K.

Figure 5.15, on the other hand, reports the results obtained through the use of the 1245 profiles in the CO data-set. Looking at the CO response, which is mainly found between 2000 and 2250 cm^{-1}, the values are much lower than those obtained for the other species. This is certainly due to the applied perturbation (+1%) that is smaller than for the other gases (+10%). To better observe the response of the spectra to the CO perturbation, a zoom is also shown in Figures 5.16 for both IRS and IRS*2. The response of IRS*2 is almost twice as intense as that of IRS. By comparing with the noise characteristic of each instrument (Figure 5.9), one can deduce that the carbon monoxide detection by IRS in both its configurations is definitely possible since both present sensitivity to this gas. On the other hand, it could also be tricky and other factors should be taken into account. First of all, the band with a spectral response to the CO is juxtaposed to an O_3 sensitivity too (see also Figure 5.14). Also, the sensitivity to carbon monoxide is located at the edge of Band 2, where the noise grows in intensity. Although IRS*2 appears to have a better response to the perturbation, which may make it easier to detect this molecule using an interferometer with these characteristics, the noise is still larger than the one of IRS.

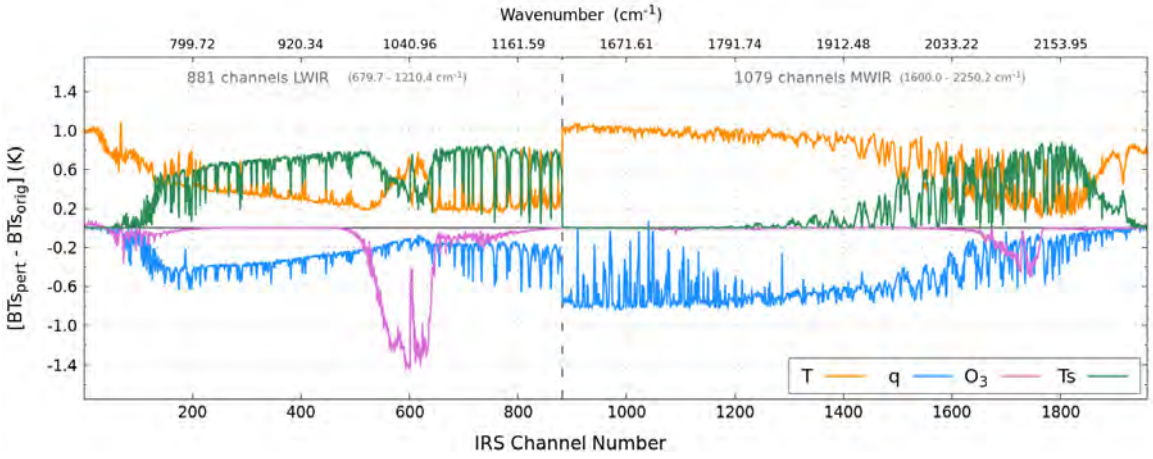
To summarise, from this analysis IRS in both evaluated resolutions is potentially sensitive to the species and variables examined. IRS*2 seems to have a better response than IRS in the case of the CO. This is not necessarily true for what is about the ozone for which, on the other hand, the responses of the two instruments are comparable (except, of course, for spectral sampling).

The areas of interest for the following steps of the PhD have also emerged in more detail. For the **ozone** a band has been chosen for both configurations of the instrument:

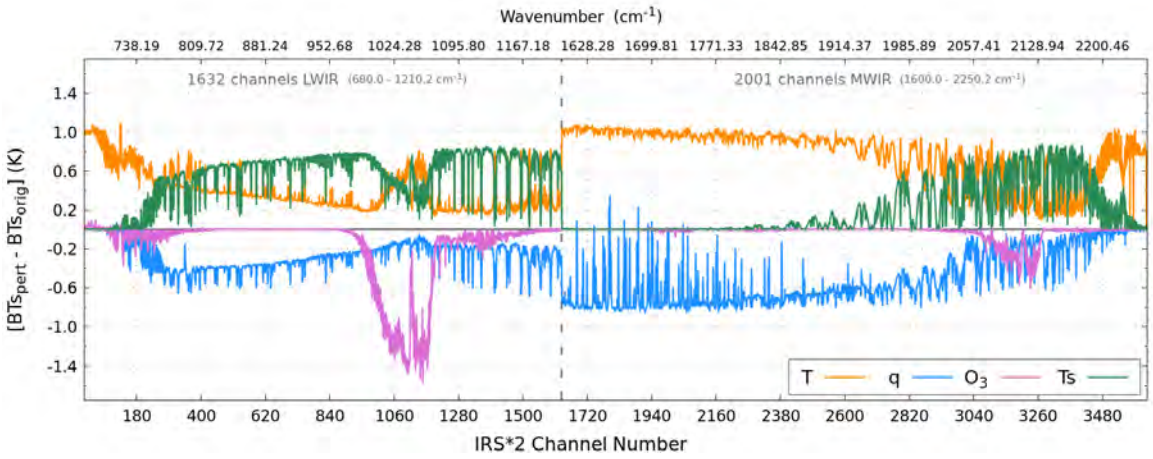
- for **IRS** this band is bounded between 982.464 and 1099.467 cm^{-1} [i.e IRS channel number 503 – 697], for a total of **195 contiguous wavelengths, or channels**;
- for **IRS*2** an equivalent band has been chosen ranging from 982.343 to 1099.379 cm^{-1} and containing **361 channels** [i.e. IRS*2 channels from 931 to 1291].

It should first be noticed that the limits of the bands are not perfectly overlapping in terms of wavelengths. This is normal since boundaries of the bands covered by the two versions of the instrument are not exactly identical and the spectral sampling is, of course, different. The

second point to observe is that not only ozone-sensitive wavelengths have been retained. In fact, wavelengths in the atmospheric window, and with a mixed sensitivity, are also found at the edges of these two bands. This is an element that will have to be taken into consideration when further evaluations or conclusions will be drawn in the following of this work.

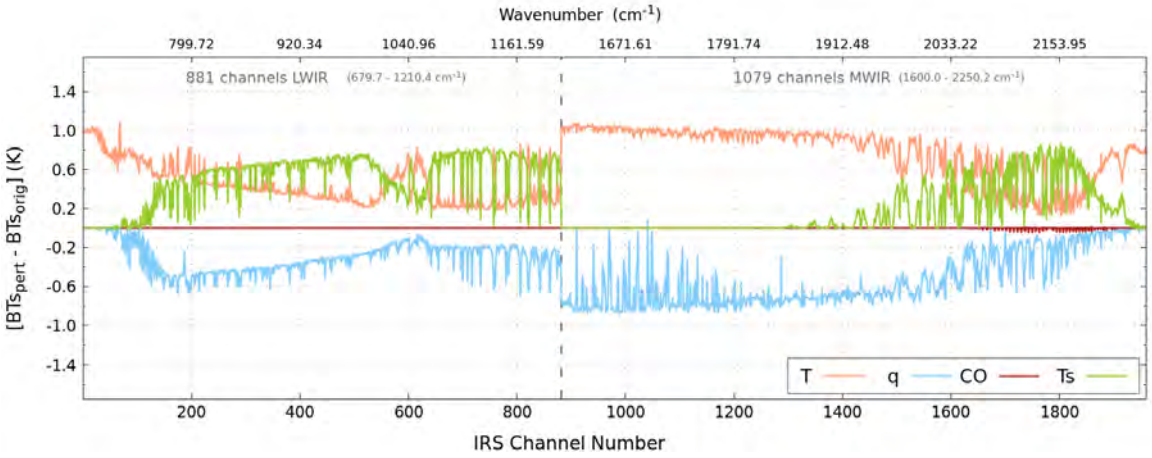


(a)

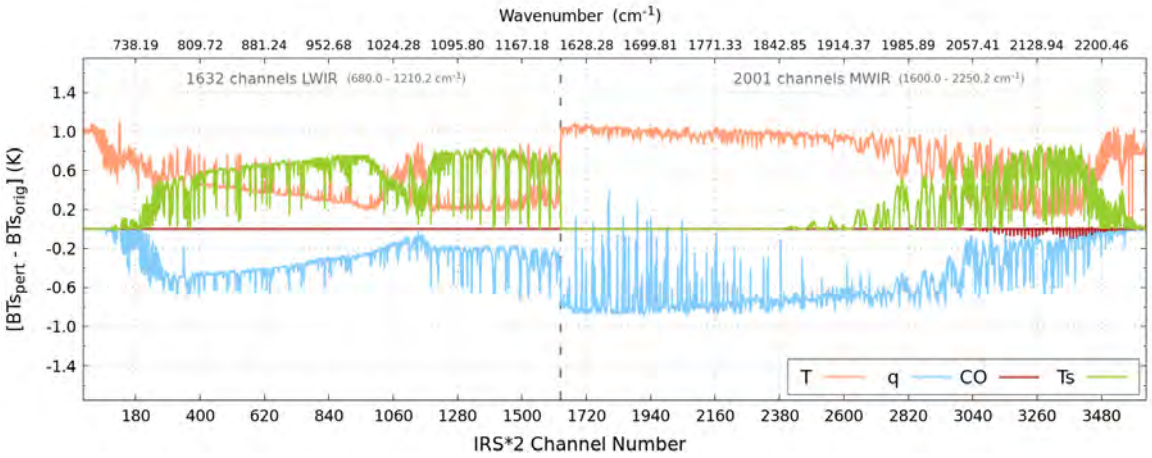


(b)

Figure 5.14: Spectral responses to the perturbation of temperature (T, orange), specific humidity (q, blue), ozone (O_3, magenta) and skin temperature (T_s, green) for IRS **(a)** and IRS with increased resolution (IRS*2) **(b)**. The curves are an average of the single responses obtained using the 1363 profiles selected from the CAMS data-set O_3-based.

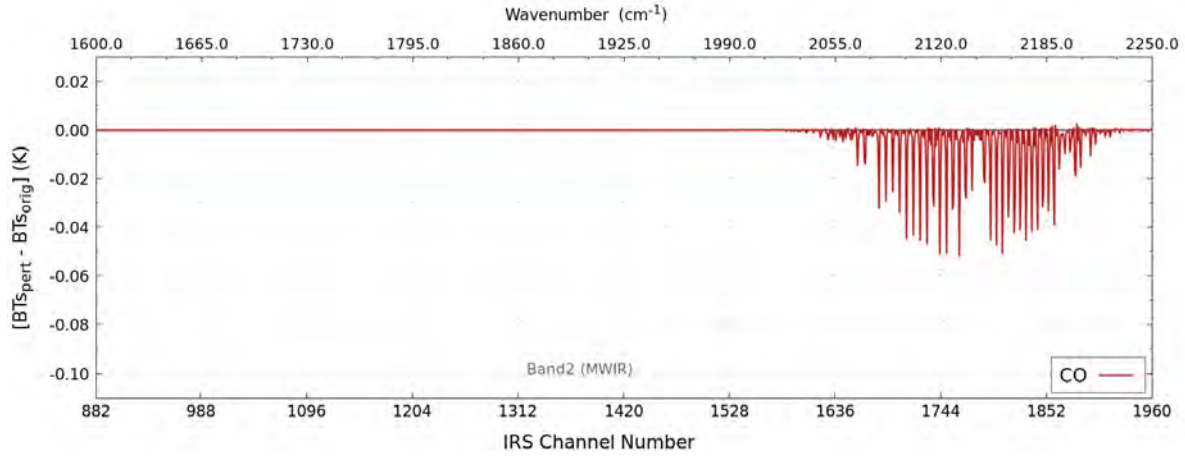


(a)

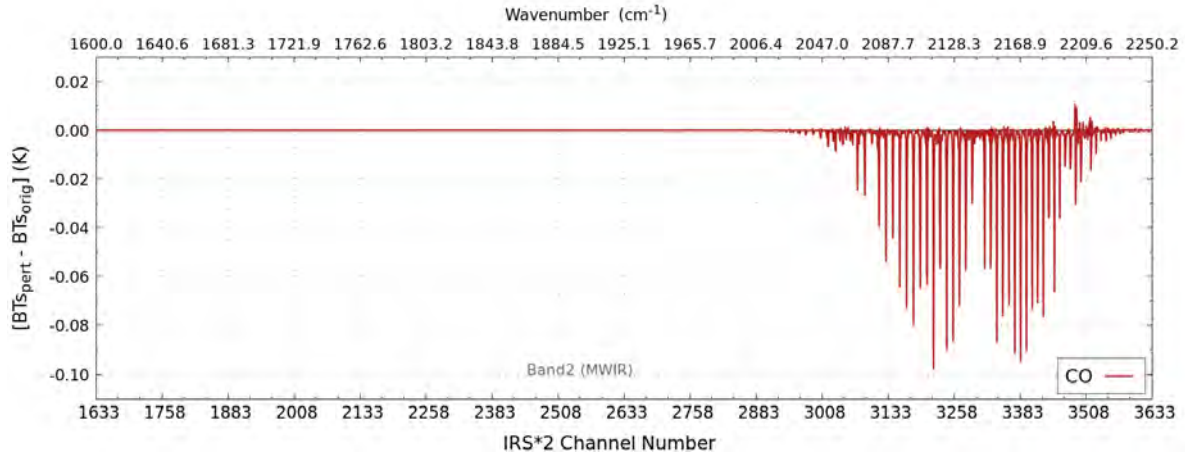


(b)

Figure 5.15: Spectral responses to the perturbation of temperature (T, salmon), specific humidity (q, light-blue), carbon monoxide (CO, bordeaux) and skin temperature (T_s, light-green) for IRS **(a)** and IRS*2 **(b)**. The curves are an average of the single responses obtained using the 1245 profiles selected from the CAMS data-set CO-based.



(a)



(b)

Figure 5.16: Magnification of what is shown in Figure 5.15 for carbon monoxide, i.e. spectral response to the perturbation of the CO profiles using instrumental characteristics of IRS **(a)** and IRS*2 **(b)**.

5.5 Highlights of this Chapter

- Monitoring of atmospheric chemistry is performed through different observations, which can be identified in two main classes: ***in situ*** and from **remote sensing**. *In situ* observations are valuable for atmospheric sciences, but they often have a limited coverage. The remote sensing covers these gaps by measuring variables from a distance.
- Remote sensing techniques can be active (using their own radiation source) or passive (measuring radiation from natural sources, such as the Sun or the Earth's atmosphere and surface).
- A very common class of passive detectors are the ***spectrometers***, that measure the spectrum of electromagnetic radiation, or the properties of light as a function of its wavelength.
- An important innovation in terms of remote sensing came with the launch of meteorological satellites, which allow to perform precise monitoring of atmospheric composition and meteorological conditions. These can cover different kind of orbits (**polar-orbiting** and **geostationary** are the most relevant for the understanding and unfolding of the present study).
- A multitude of instruments, on board of different satellite platforms, are now a day available for the monitoring of atmospheric chemistry measure different ranges of radiation. Each of them presents both strengths and limitations. Only those relevant to this study have been covered in more detail.
- **Ozone Monitoring Instrument (OMI)** is a UV/VIS spectrometer on the NASA Aura spacecraft, providing global coverage and measuring ozone, trace gases, aerosols, and cloud properties. OMI's data are exploited for studying air quality, climate, and atmospheric composition.
- The **Infrared Atmospheric Sounding Interferometer (IASI)** is a hyper-spectral Fourier-Transform Spectrometer on the polar-orbiting Metop series operated by EUMETSAT. The instruments profile atmospheric temperature, humidity, and atmospheric compounds, like ozone, with high spectral resolution and signal-to-noise ratio. IASI data can be used for monitoring and forecasting atmospheric chemical composition, climate studies, and detecting trace gas-es, clouds, aerosols, and greenhouse gases.
- The **Infrared Sounder (IRS)**, which is the instrument to be studied along the present work, is a Fourier Transform Spectrometer (FTS) built by Thales Alenia Space, which will be launched on board the geostationary MTG-S in 2024 and provide data of the full Earth disk with a 4 km spatial sampling at nadir. The sounder will cover 1960 channels spread over two bands: Band 1 in the LWIR range ($679.70 - 1210.44\,\mathrm{cm}^{-1}$) and Band 2 in the MWIR range ($1600.00 - 2250.20\,\mathrm{cm}^{-1}$).

- An alternative version of IRS, referred to as **"IRS*2"** throughout the present manuscript, is being studied with a spectral sampling of $0.3251\,cm^{-1}$ and a total of 3633 channels, and will be evaluated as well by the present PhD project.
- Due to the portions of EM spectrum this will cover, IRS (and IRS*2) will be sensitive to various chemical species, including CO_2, O_3, SO_2, CO, N_2O, and more.
- Sensitivity studies are needed to select specific subsets of wavelengths for analysis and simulations, and have been performed:
  - profiles of ozone, carbon monoxide, temperature, humidity and surface temperature have been gathered from the CAMS atmospheric composition forecasting system [Eresmaa and McNally (2016)];
  - a thinning was applied to the profiles based on acquisition criteria, resulting in 1363 ozone-based profiles and 1245 carbon monoxide-based profiles;
  - the ***physical selection method*** [Gambacorta and Barnet (2012)] was implemented to assess the sensitivity of IRS to ozone and carbon monoxide by perturbing the profiles and evaluating the spectral responses.
- From this analysis IRS in both evaluated resolutions is potentially sensitive to the species and variables examined. IRS*2 seems to have a better response than IRS in the case of the CO. This is not necessarily true for what is about the ozone for which, on the other hand, the responses of the two instruments are comparable (except for spectral sampling).
- For the ozone a band has been chosen, for the following steps of the PhD, for both configurations of the instrument:
  - for IRS this band is bounded between 982.464 and $1099.467\,cm^{-1}$ [i.e IRS channel number 503 - 697], for a total of 195 contiguous wavelengths, or channels;
  - for IRS*2 an equivalent band has been chosen ranging from 982.343 to $1099.379\,cm^{-1}$ and containing 361 channels [i.e. IRS*2 channels from 931 to 1291].

Chapter 6

Observing System Simulation Experiments

Evaluating the potential impact of a new instrument is a common issue in the field of atmospheric sciences. Preparatory studies are required for instruments that may still being designed, not yet operational or maybe just starting to work and not yet providing immediately exploitable observations. Preparing the arrival of a new set of measurements is crucial to have an idea of how it will perform in an observation system or how these data may be employed. To do so, a hypothetical observing system must be simulated as best as possible. The more accurate the simulated observations and their assimilation, the more reliable the analysis results will be.

In the context of this work, it is indeed necessary to dispose of good quality and reliable simulated spectral radiances in order to assess the impact of IRS assimilation into the CTM MOCAGE. Since IRS is not yet flying at the time the research is carried out, the strategy we adopted, and the one most commonly adopted in the atmospheric sciences in similar scenarios, is to perform an Observing System Simulation Experiment (**OSSE**) [among others Errico et al. (2007); Masutani et al. (2010); Claeyman et al. (2011); McCarty et al. (2012); Privé et al. (2013a); Privé et al. (2013b); Boukabara et al. (2016); Duruisseau et al. (2017); Descheemaecker et al. (2019); Zeng et al. (2020)]. This kind of experiments comply a series of steps and validations to ensure that the observations are reliable, truthful and to put them to test into a data-assimilation system (DAS).

An OSSE basically consists in simulating synthetic observations from an atmospheric model state representing reality, assimilating them within another model state representing the model itself, and finally evaluating their impact on analyses and forecasts (simple scheme in Figure 6.1).

- The reference reality is usually referred to as **Nature Run (NR)**. This consists of an atmospheric state that must, in fact, realistically reproduce the true state of the atmosphere. It will serve, throughout the experiment, as the state from which observations

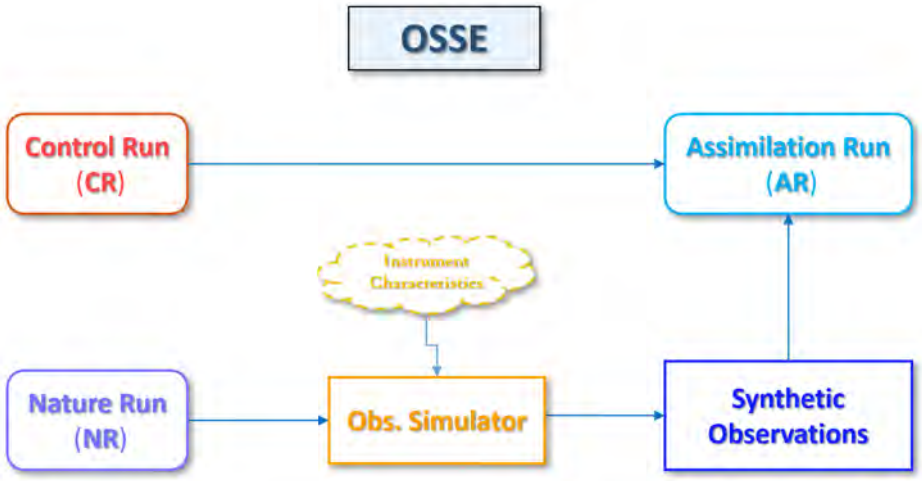


Figure 6.1: An illustrative scheme for the main stages in an OSSE.

will be simulated and as the reference against which the final assimilations will be verified. It is usually produced using a good quality model in free-run, or not providing any information coming from real observations.

- The NR reality is used to feed an ***observation simulator***, that is usually a RTM through which the sought **synthetic observations** will be produced. Into the simulator some specific instrumental proprieties have to be specified (such us optics, observational geometry, spatial and temporal resolution). The perfect observations obtained are then perturbed with an instrumental error, to be properly assessed, in order to reach their final shape.

- Another fundamental step in the development of an OSSE is the creation of the so-called **Control Run (CR)**, that is a run of a model simulating the reality. The differences between the NR and CR should be those existing between the reality itself and the output of a good quality model trying to reproduce it.

- The synthetic observations will then be assimilated into the CR. This final run, referred to as **Assimilaton Run (AR)**, is realized with the same configuration as the CR, by assimilating the synthetic observations created from the NR.

- Finally, the impact of assimilation is assessed by comparing the results to the reality of the NR, to the CR.

For an OSSE to be considered robust, it is very important that the CR is consistent with the NR, but different enough to avoid the so-called ***"identical twin" problem*** [e.g. Arnold Jr and Dey (1986); Masutani et al. (2010)]. This could occur when using an identical physical representation of process dynamics for all the steps in the DAS, i.e when using the same model. The physical representation, together with using the same inputs, leads to

underestimating model errors and producing results that are too similar between reality, i.e. NR, and assimilation. This latter will then be too close to Nature Run and, therefore, the evaluation of its impact too optimistic, i.e. biased.

Underestimated model errors also mean observations very close to the background [e.g. Atlas (1997)]. As a consequence, the impact of observations over the analysis [i.e. *Analysis minus Background* (A-B) and *Observations minus Analysis* (O-A)] will be too small compared to what happens in reality in the operational, and then false too.

All of this can be avoided by using different models to create the two scenarios, or by sufficiently differentiating the inputs and the configurations in the same model, so that the errors are properly represented and the outputs are consistent but divergent at the same time. In any case, the spatio-temporal variability in NR must be properly evaluated against the CR before assimilation is carried out [Timmermans et al. (2015)].

6.1 OSSE Frame for IRS

To the best of this author's knowledge, a few of OSSEs have already been carried out since when the instrument IRS was designed, in order to prepare its arrival and the assimilation of its data into NWP models. The first of these works concerns a preliminary study by Guedj et al. (2014) aimed to assess the benefits of high-density radiance data assimilation into a fine-scale forecast model. The model ARPEGE[1] [e.g. Courtier et al. (1991); Bouyssel et al. (2022)] was used to build the NR. The synthetic observations they simulate have been assimilated in to the model Application of Research at the Operational Mesoscale (AROME) [e.g. Seity et al. (2011); Brousseau et al. (2016)], in order to avoid the identical twin problem. However, at the time such work took place, the final specifications on which the final instrument was built were not yet available. The spectral bands simulated had slightly different margins than today, i.e. LWIR between 700 and 1210 cm^{-1} and MWIR in $1600 - 2175\,\mathrm{cm}^{-1}$, while the instrument presented a provisional spectral sampling of $0.625\,\mathrm{cm}^{-1}$.

This was followed by a study by Duruisseau et al. (2017), which was focused on the evaluating of a MW instrument, to be based on a GEO platform, the assimilation of which was integrated into a dense observation system that also included IRS. Starting from one of the configurations proposed by Guedj et al. (2014), they simulated 25 IRS water vapour channels, with an improved simulation of quality control on cloudy radiances. Their study, showed a positive impact up to a 18h forecasting range.

Another work has begun in parallel with the present doctoral project. It has been carried out by Coopmann et al. (2023) in order to evaluate the impact of IRS on the NWP, with more recent specifications for the instrument. This study has been focused on assessing the potential benefit of IRS assimilation on top of the full observing systems already used in the mesoscale NWP model AROME. Their NR and CR, produced using AROME, used

[1] Action de Recherche Petite Echelle Grande Echelle (ARPEGE)

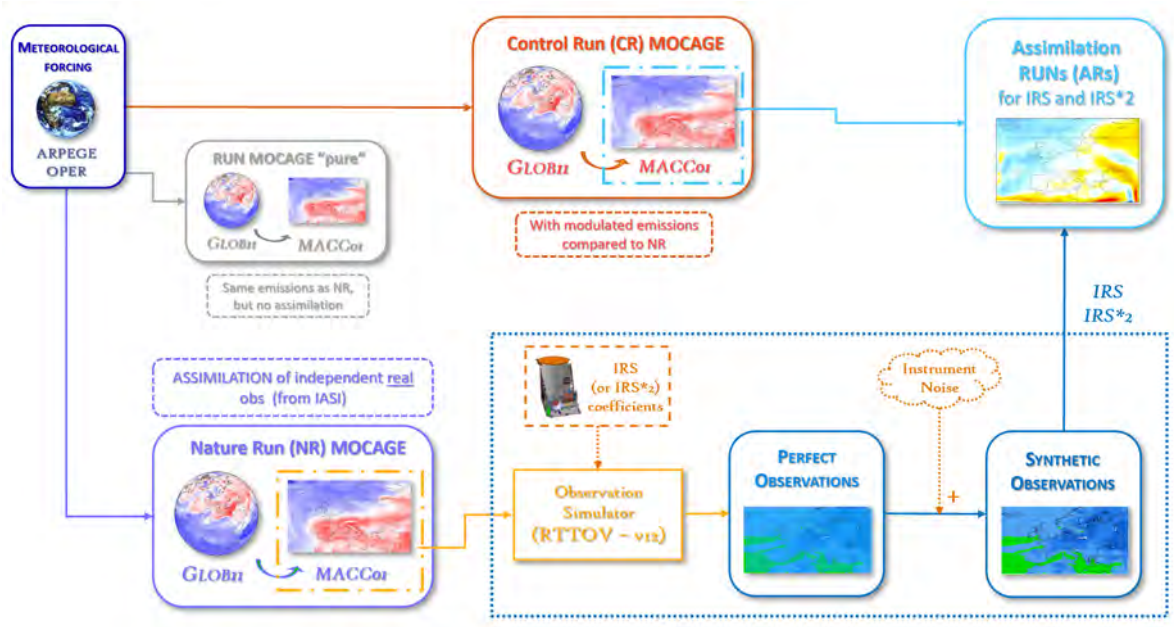


Figure 6.2: Implementation of an OSSE specific to the present work.

different horizontal resolutions and vertical range. They simulated the whole IRS spectrum, taking into account the effect of clouds and the PC compression. They selected 75 channels for the assimilation spread over the two bands. The study covered four months (January-February and July-August, 2020). The IRS radiance assimilation lead to an improvement of temperature, humidity and wind speed forecasts up to 10% over the tropospheric column, up to 48h forecast range.

Both studies, as opposed to the present one, focused on the impact of simulated radiances in a NWP model and not a CTM. In addition, the present effort also aims to evaluate a version of IRS with finer resolution. To do so, an OSSE was specifically designed with great attention in order to meet the objectives of the project and by taking into account all the available resources.

The diagram in Figure 6.2 visually summarises the strategy adopted for the creation of the OSSE specific for this study. First of all, MOCAGE was the model chosen for the creation of both CR and NR. In both cases, both global and regional domain were activated and the meteorological forcing came from ARPEGE in its version operational in Météo-France. Consequently, since the settings exposed so far are identical, a big effort had to be done in order to plan a strategy avoiding the identical twin problem. The first step was to differentiate the surface emissions to be used for the two runs. For biogenic and anthropogenic emissions in the NR framework, the configurations used in operational MOCAGE at the time this work begun, have been employed for each geographical domain. For the CR framework, on the other hand, data referring to the year 2000 were used. This provided the same kind of spatial variability for this class of emissions than the NR, but with different intensities. For the representation of the biomass burning, data from the CAMS GFAS, from ECMWF, were provided in input

	CR Framework		NR Framework	
	GLOB11	MACC01	GLOB11	MACC01
METEOROLOGICAL FORCING	• ARPEGE OPER	• ARPEGE OPER	• ARPEGE OPER	• ARPEGE OPER
BIOGENIC EMISSIONS	• CAMS-GLOB-BIO [a] (year 2000) • CAMS-GLOB-SOIL [b] for NOx (year 2000)	• CAMS-GLOB-BIO (year 2000) • CAMS-GLOB-SOIL for NOx (year 2000)	• MEGAN-MACC [g] (year 2010) • GEIA for NOx (year 1990)	• MEGAN-MACC (year 2010) • GEIA for NOx (year 1990)
ANTHROPIC EMISSIONS	• CAMS-GLOB-ANT [c] (year 2000) • GEIA [d] (for chlorine species)	• CAMS-REG-AP [f] (year 2000) • GEIA (for chlorine species)	• MACCity (year 2016) • RCP60 [h] (year 2016)	• CAMS-REG-AP (year 2017) • GEIA (chlorine species)
BIOMASS BURNING	• MACCity [e] (year 2000)	• MACCity (year 2000)	• GFAS [i] (year 2019)	• GFAS (year 2019)
DATA ASSIMILATION	• None	• None	• IASI LIC	• None

Table 6.1: Summary of the different settings chosen for the CR and the NR frameworks. The boxes in green indicate that the same parameters have been employed for both runs, given the same MOCAGE geographical domain. The boxes in red shades, on the other hand, highlight the parameters that have been differentiated.

[a] Granier et al. (2019); Sindelarova et al. (2014)
[b] Granier et al. (2019); Simpson and Darras (2021)
[c] Granier et al. (2019)
[d] Global Emissions Inventory Activity (GEIA)
[e] Lamarque et al. (2010); Granier et al. (2011); Diehl et al. (2012)
[f] Guevara et al. (2022); Kuenen et al. (2022)
[g] Sindelarova et al. (2014)
[h] Fujino et al. (2006); Van Vuuren et al. (2011)
[i] Global Fire Assimilation System (GFAS); Kaiser et al. (2012)

to MOCAGE for the NR settings, while for the CR was MACCity representative of the year 2000. For the details about the emissions provided to MOCAGE for each run and domain, see Table 6.1.

In addition to the modulation of the surface emissions, another action was taken in order to have a CR and a NR consistent with each other, but different enough to reproduce the differences that would exist between the true reality and a model that reproduces it. During one of the thesis follow-up committees, a suggestion emerged to make the NR as close to reality as possible. This strategy was slightly unconventional since it involved using the model not in free-run, as the general theory of the OSSEs states, but with assimilation of real observations. Following this suggestion, the final strategy has then been designed. It eventually involved the assimilation of radiances from the Metop-IASI spectrometer into the MOCAGE model in the GLOB11 configuration. However, the regional domain MACC01 nested in the global, and in which no assimilation was directly carried out, would have been the NR employed for the simulation of IRS observations (procedure better described in Subsection 6.2.2).

A MOCAGE run, which has been named “pure", was equally planned and performed. It exploited the same meteorological input as the other runs. No assimilation was performed in it, like the CR. At the same time, it the same surface emissions as the NR were provided. As a consequence, such a run provides a reference in order to evaluate the impact of the IASI radiance assimilation in the NR and, at the same time, the contribution of the surface emission modulation in the CR.

Next, the simulation of IRS radiances was carried out from the NR on the MACC01 domain, through the RTTOV v12 RTM. “Perfect", i.e. noise-free, observations were created in parallel for both IRS and IRS*2 in the ozone bands pre-selected through the sensitivity study described in Chapter 5 (see 5.4.3.3).

The observations thus produced over MACC01, for both instrumental configurations, were then assimilated into the MACC01 domain of the CR.

6.1.1 Period of Study

The time period running from the 1st of May to the 31st of August 2019 has been chosen to carry out the experiments. More specifically, the runs without assimilation, i.e. CR and NR, started on the day May the 1st. Assimilation into the global IASI L1c radiances began on May 15th, which left 15 days of spin-up time for the system. An additional 10 days, were left as spin-up before debuting the assimilation of IRS radiances into the CR. The ARs, thus began from May 25th. For a more clear overview of these dates, please refer to Figure 6.3.

Each evaluation on the results, however, was made starting from the 1st of June. Such time-frame was retained for a few reasons. First of all, two intense and well-documented episodes of ozone pollution over Europe occurred in Summer 2019: one in the end of June (24th-30th) and another in July (21st-28th) [Tarrasón et al. (2020); Schulz et al. (2019a,b)]. The very hot weather conditions facilitated the photochemical production of O_3 by elevated concentrations of nitrogen oxides and volatile hydrocarbons, while hindering its removal due to the absence of rain. During the June episode, high ozone values have been registered first over Germany, Netherlands, Belgium and Northern Italy, and later, on Southern Scandinavia, Poland, Austria and France. During the episode of July, among the countries that experienced high ozone concentrations were Switzerland, Belgium, Luxembourg, the Netherlands, Germany, France and Austria. Some of the highest and most persistent ozone concentrations have been recorded in Northern Italy (the Po Valley).

Furthermore, during the time-frame being considered, data from up to three IASIs are simultaneously available, depending on the day. Consequently, having data from the three instruments, operational at the same time on board Metop-A, B and C, could potentially provide an interesting data source for the assimilation of radiances in the NR preparation step.

Finally, since designing an OSSE requires a significant amount of resources and computational time, different uses have to be envisaged. Therefore, although carbon monoxide is

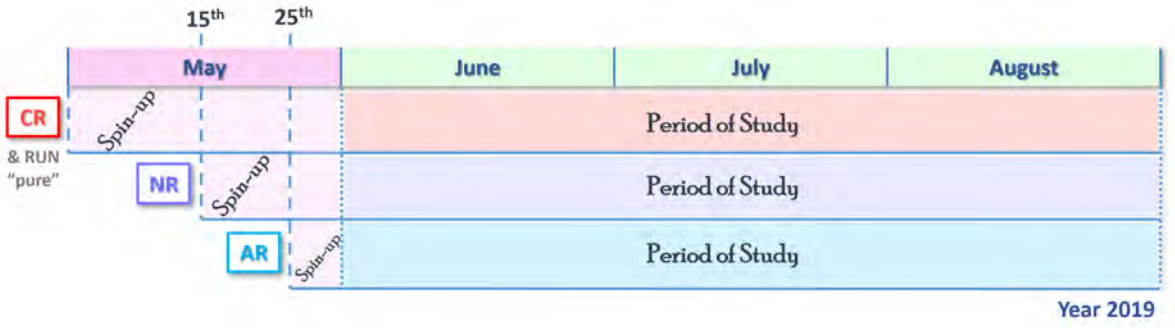


Figure 6.3: Overview of the time-frames of execution and study of each run. CR stands for Control Run, NR for Nature Run and AR for Assimilation Run. The months refer to year 2019.

not specifically treated throughout this work, future works to be done on this species are anticipated. The CO emissions, as introduced in Chapter 1, are often anthropogenic, but also often inter-connected to wildfires. These phenomena, within the regional domain chosen for the simulations, are most likely to take place during the northern-hemisphere summer period. Even if, at the best of this author knowledge, no massive fires have been documented during Summer 2019, smaller fires take place very often during Northern summer-time all over the Mediterranean area.This confirms the choice of the May-August 2019 period for the construction of the OSSE, which could also be employed for future studies based on the carbon monoxide.

Please notice that the first 16 days of the period have been used as spin-up period and were not included in the evaluations carried out all over the study. All the analysis and statistics performed have been focusing on the three-month period ranging from the 1st of June to the 31st of August.

6.1.2 Study on the Activation of Aerosols in MOCAGE

As introduced in Chapter 3, MOCAGE is able to reproduce several species and processes. However, activating all of them and running the model to its full potential would require large execution times, but also be redundant since not all of its skills are needed for the simulations required in this work. Among the others, a few considerations have been made on activating or not the aerosols. Conceptually, indeed, the activation of aerosol simulation should not have a relevant impact on the simulation of the chemical species of interest for this study. Activating the secondary aerosols, for example, could unbalance the ozone production reactions because nitrates are produced. However, this impact is supposed to be of second order.

Nevertheless, in order to test this assumption, a sanity check was carried out and two "pure" runs were executed: one with aerosol activation and one without it. In Figure 6.4 statistics on the total column of O_x produced by MOCAGE with the aerosols activation (Aer_{ON}) and without it (Aer_{OFF}) are shown. The averages (Panels 6.4b and 6.4b), which refer

to the 3-month period of study from June to August, show values very close to each other. At first glance, actually, the scenarios look identical. To be more precise in the evaluation, the relative difference between the two cases is also computed and displayed in Panel 6.4e. From the latter, one can deduce that the two frameworks differ from each other by very low percentages. The highest values of divergence, occurring over the European Continent (10 to 40°E, 45 to 60°N), never exceed the very weak percentage of 0.3%. The associated standard deviations, shown in the Panels 6.4b and 6.4d, depict scenarios that are also very similar to each other, with the same spacial variability and very similar magnitudes. The error associated to the bias between these two runs (Panel 6.4f), on the other hand, presents very low values, ranging from a minimum of 0.04 DU (in the South-Est corner of the domain) to a maximum of 0.16 DU (over the Atlantic Ocean, around 45°N). These values are infinitesimal if compared to the average values previously illustrated.

An evaluation was also carried out on the vertical atmospheric column. The O_x concentration values obtained on each of the 60 vertical levels of MOCAGE were, once again, averaged over the time period examined (1st of June - 31st of August, 2019). Afterwards, for each level an average was also computed over the geographical domain, i.e. MACC01. The results are shown in Figure 6.5a. Values referring to the case where aerosols have been activated are depicted in magenta and labelled as $\overline{\text{Aer}}_{\text{ON}}$. The standard deviations associated are added and subtracted to the average values and are shown in pink shadow $\sigma(\text{Aer}_{\text{ON}})$. The same kind of results are displayed in green for the MOCAGE run not simulating the aerosols, i.e. $\overline{\text{Aer}}_{\text{OFF}}$ and $\sigma(\text{Aer}_{\text{OFF}})$. As it is easy to notice, the curves are almost exactly overlapping. In order to be able assess a variation, the relative difference must be computed (in 6.5b). This shows differences that are dependent on the atmospheric level. Very low variations are observed between the averages, with maxima occurring at level 40 (not exceeding 0.6%). This vertical level can be expected[2] to be on average between 240 and 330 hPa, then approximately at the tropopause.

In conclusion, the results obtained through this study show that the aerosols activation in the MOCAGE setting has effect on the values of O_x that can be assumed to be negligible. Consequently, it has been decided not to activate aerosol representation for the following of the work, i.e. in the setting of the MOCAGE runs constituting the OSSE.

[2]For equivalence between vertical levels and pressure, see appendix A.

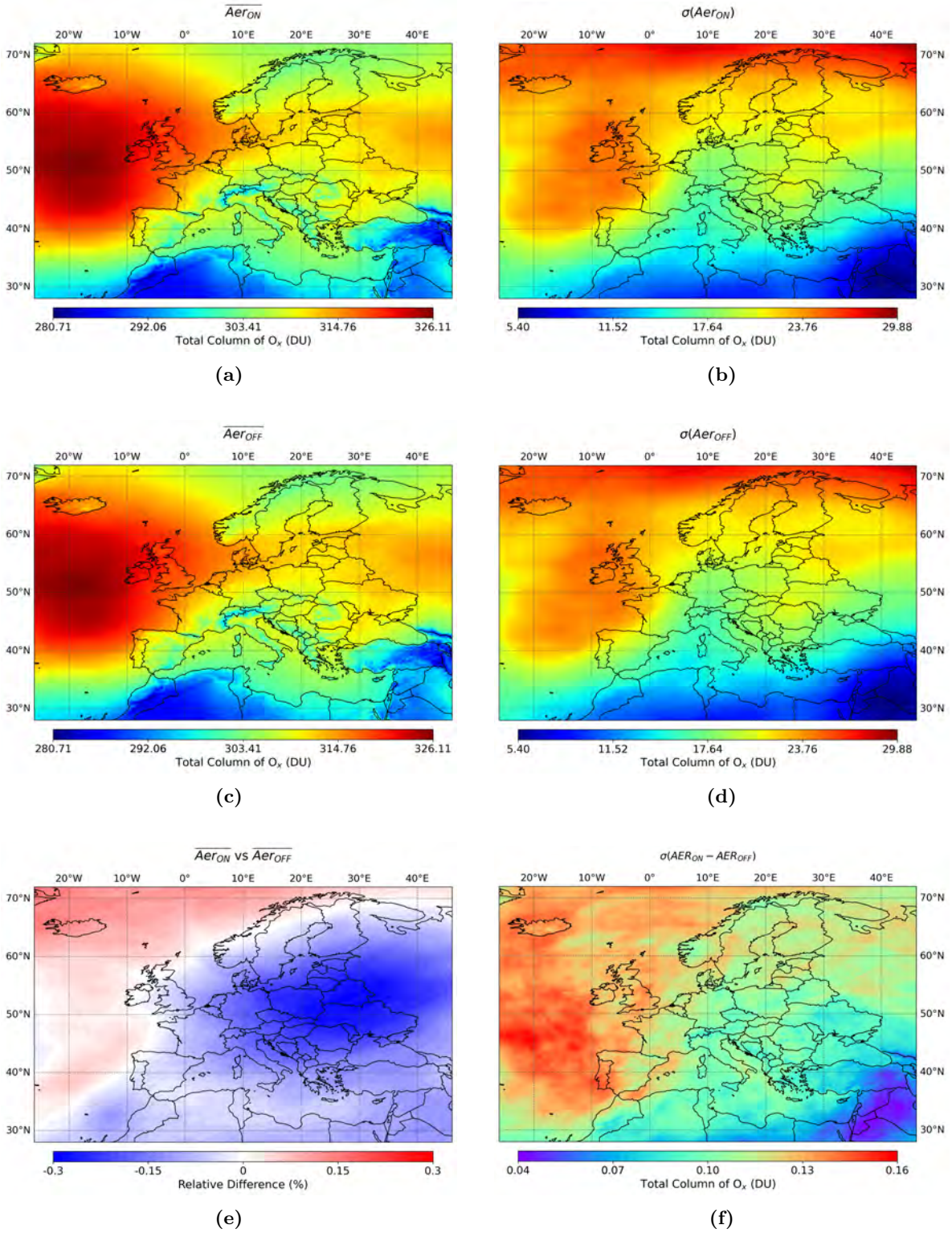
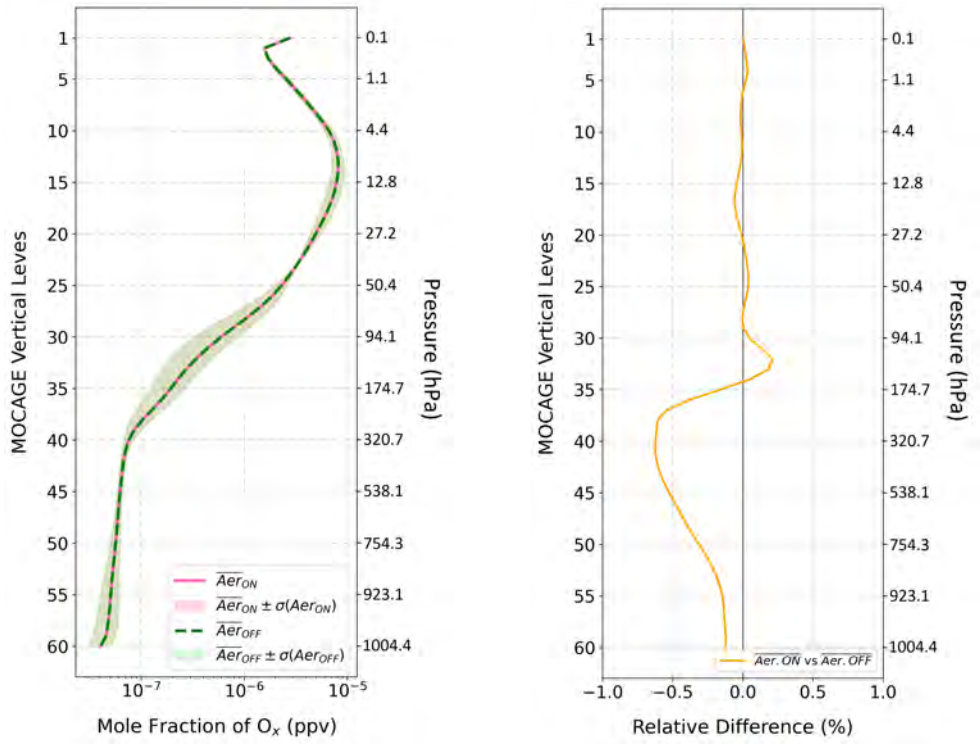


Figure 6.4: Averages of the O_x Total Columns over the 3-month period of study (1st of June up to 31st of August, 2019) for a run MOCAGE with activation of the aerosols **(a)** and without it **(c)**. Their percentage relative difference is in **(e)**. The corresponding standard deviations and the standard deviation of the differences are in **(b)**, **(d)** and **(f)**, respectively.



(a) **(b)**

Figure 6.5: Average over the 3-month period of study (1st of June up to 31st of August, 2019) and, for each of the 60 MOCAGE model levels, on the MACC01 domain **(a)** for both runs with ($\mathrm{Aer_{ON}}$) and without ($\mathrm{Aer_{OFF}}$) aerosols activation. The percentage relative difference between the averages is shown in **(b)**.

6.2 Nature Run

6.2.1 Nature Run Settings

Since producing a good quality NR is crucial for the success of the whole experiment, this step deserves a special focus.

As already introduced above, we chose to assimilate radiances acquired by the IASI instrument into the MOCAGE global domain GLOB11, in a configuration of 60 vertical levels. The corresponding MACC01 domain, where no assimilation is performed, benefits from the assimilation through the boundary conditions provided by the global domain. The output from MACC01 is, then, taken as the sought NR.

For the purpose of this part of the study, IASI L1c data are exploited, gathered from the EUMETSAT Earth Observation Portal[3].

Although the topic of the direct assimilation of IASI L1c data into MOCAGE is still in course of refinement, in the last few years different efforts have been made in order to implement it and test it, as in Emili et al. (2019) and El Aabaribaoune et al. (2021). The settings we chose for the assimilation are the same exploited in the literature just mentioned. A total of 284 channels, belonging to the band of the MWIR sensitive to the ozone, have been assimilated. The assimilation algorithm exploited is the 3D-Var, with an hourly assimilation window. Observation errors are computed through Desroziers diagnoistic, while background errors are evaluated as a percentage of the ozone profiles themselves. In the following subsection, this configuration will be illustrated in detail.

6.2.2 Study on the Assimilation of IASI L1c Data in Preparation for the Nature Run

In the context of the preparation of the NR for the OSSE needed for the present PhD, a study has been performed in order to compare the global configuration, used to produce the NR, to real observations (OMI and ozonesondes) and to evaluate its reliability. The quality of this run has been found to be similar to the quality of the results obtained in literature.

At the same time, parallel tests were carried out in order to bring a contribution to further improving and refining the method of assimilating L1 radiances within a CTM (for IASI, but also for other IR hyperspectral sounders). The possibility of using different background errors than those proposed so far has been explored, together with the evaluation of the impact of thinning the group of 284 assimilated channels. Since this effort provided interesting results for the further development of the methodology, it has been the topic of a scientific paper submitted to the *Journal of Geophysical Research: Atmospheres*, which is reported in full in the following.

[3] `https://eoportal.eumetsat.int` (last access: 2 October 2022)

Further improvements of IASI radiance assimilation in MOCAGE for ozone: exploring the impact of background errors and channel subsets

Francesca Vittorioso[1], Vincent Guidard[1], Nadia Fourrié[1], Clément Lavollée[1]

[1]CNRM, Université de Toulouse, Meteo France, CNRS, Toulouse, France

Key Points:

- Ozone
- MOCAGE
- atmospheric chemistry

Corresponding author: Francesca Vittorioso, francesca.vittorioso@meteo.fr; francesca.vittorioso@outlook.com

Abstract

Satellites provide the most global view of the atmospheric composition, in particular for ozone. In the last few years, efforts have been made in order to implement the direct assimilation of Infrared Atmospheric Sounding Interferometer (IASI) L1 data for ozone into the MOCAGE chemistry transport model, using 284 channels in the ozone band. This study aims to further investigate the assimilation of IASI L1 data into a recent version of MOCAGE with two goals: studying the impact of the background errors $\mathbf{B}$ and optimizing the assimilation through the reduction of the channel number exploited. Two $\mathbf{B}$ matrices have been computed through two different methods: one is a percentage of the ozone profiles themselves ($\mathbf{B}_{\mathrm{perc}}$), while the other is computed using the so-called NMC method ($\mathbf{B}_{\mathrm{nmc}}$). Each matrix has then be tested in the assimilation process separately. Comparison of the results to independent data from OMI and radiosoundings showed improvements all over the globe with the use of $\mathbf{B}_{\mathrm{nmc}}$ rather than $\mathbf{B}_{\mathrm{perc}}$. The error reduction when comparing to ozonesondes is localised in the higher levels of the atmosphere (circa 200 ppbv error reduction) and at the UTLS. To achieve the second objective, different groups of channels have been tested, using the $\mathbf{B}_{\mathrm{nmc}}$: 284 and two subsets of 76 (rp. 34) representing 90% (rp. 83%) of the total ozone Degrees of Freedom for Signal. Similar or better quality is achieved when assimilating a smaller number of channels, especially in the stratosphere and at the UTLS, with OMI and ozonesondes as verification data.

1 Introduction

Ozone is a trace gas and a secondary species (i.e. not emitted) resulting from photochemical reactions. It is mainly found in stratosphere (almost 90%), resulting from photodissociation of oxygen. The stratospheric ozone layer is of paramount importance in shielding harmful ultraviolet solar radiation and it makes possible life on our planet. On the other hand, a smaller percentage of ozone is found in troposphere. It comes in small part from the stratosphere itself and in larger amounts from reactions in the lower layers involving primary compounds of natural or anthropogenic origin. Tropospheric ozone can cause issues to human health and the ecosystem, as well as to agriculture and material goods due to its high oxidant power.

Since ozone plays such a key role in the Earth's radiative balance, atmospheric chemistry and air quality, it is crucial to monitor and predict its atmospheric concentration and transport. To do so, prediction from numerical models and observations, *in situ* or from remote sensing, are used individually or more often combined through the data assimilation process.

Satellite data are those that provide a more global view of the atmospheric situation. They are assimilated into models in different forms. Level 2 (L2) ozone retrievals (products) deduced from observed data are by now efficiently assimilated into Chemical Transport Models (CTMs) (e.g. Lahoz et al., 2007; Massart et al., 2009; Emili et al., 2014; Flemming et al., 2017). This kind of process, however, may be subject to errors introduced by the *a priori* information used to produce the retrievals. Level 1 (L1) radiances, on the other hand, are already efficiently assimilated in numerical weather prediction models (NWP) in order to correct meteorological variables such as temperature and water vapour profiles (Bouyssel et al., 2022; Bormann et al., 2019). In that case, the radiative transfer is usually part of the observation operator in the assimilation process, avoiding the introduction of biases due to the *a priori* information used for satellite retrievals and, as a consequence, the quality of the assimilation itself.

In the last few years, efforts have been made in order to implement the direct assimilation of Infrared Atmospheric Sounding Interferometer (IASI) L1 data into the Model Of atmospheric Chemistry At larGE scale (MOCAGE) (i.e. Emili et al., 2019; El Aabaribaoune et al., 2021). These studies focused mainly on implementing the L1 assimilation process in

the CTM, assessing the impact of assimilation of radiances compared to the one of L2 products and, in a second step, computing the appropriate observation error matrices.

The present study aims to further investigate the assimilation of IASI L1 data into the most recent version of MOCAGE with two main goals: studying the impact of the covariance matrix of the background errors **B** and optimizing the assimilation efficiency. Targeted upgrades compared to the previous studies have been set in order to further refine the system and to examine specific topics more carefully. To do so, we test two **B** matrices computed through different methods and thus representing the background errors differently. The first one is obtained, just as in a big part of literature about this topic, as a percentage of the ozone profiles themselves. A second **B** matrix, on the other hand, is computed using the method introduced by the National Meteorological Centers (NMC) (Parrish & Derber, 1992) and using forecasts valid at the same time. Although some studies exist using an "advanced" **B** matrix for the assimilation of L2 data (Inness et al., 2013), to the best of the authors' knowledge, no literature exists about background error studies for assimilation of radiances into a CTM.

MOCAGE being an operational model, a second goal of this paper is to understand whether, by reducing the number of channels assimilated in the model, the process could be optimised while maintaining a good efficiency. To do so, after choosing the "best" **B** matrix, different groups of channels are tested.

This paper is organized as follows. Section 2 describes the sources from which we picked up the sets of data that we exploited. Data type, instruments and networks are described for data used in both assimilation and validation. In Section 3 the theory behind the methods is illustrated. More in detail, the characteristics of the MOCAGE model in which we perform the assimilation, the assimilation algorithm we use and the techniques for evaluating the error matrices are described. In Section 4, we illustrate the choices made in order to set-up the experiment, namely the case study, the data thinning for IASI radiances and finally the choice of the spectral band and channels to be assimilated. Section 5 groups the results. In a first stage, a study is shown aiming to evaluate the impact of using two different **B** matrices but the same number of channels. In a second phase, for the same **B** matrix, the impact of using different channel groups is shown. Finally, conclusions and perspectives are drawn at the end of the paper.

2 Data

2.1 IASI L1C

The hyper-spectral IASI is a Fourier-Transform Spectrometer (FTS), main payload of the European Meteorological Operational Satellite (Metop) series operated by the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT) (Siméoni et al., 1997; Blumstein et al., 2004). This interferometer acquires spectra of atmospheric emission within 645 and 2760 cm^{-1} (3.6 - 15.5 µm), with a spectral apodized resolution of 0.5 cm^{-1} and a 0.25 cm^{-1} spectral sampling. Consequently, it measures at 8461 wavelengths (or channels) in a 12 km diameter footprint at nadir.

The first IASI instrument has been launched on board the Metop-A satellite in 2006. Two supplementary instruments followed on board Metop-B (launched in 2012) and Metop-C (2018). In the time period ranging from July 2019 until the retirement of Metop-A in November 2021, the three instruments provided operational exploitable data simultaneously, with a nearly global coverage.

The IASI mission was originally meant to profile atmospheric temperature and humidity for weather forecasting applications mainly, but also for measuring atmospheric compounds like ozone. It showed over time its high potential to be exploited in monitoring and forecasting atmospheric chemical composition and for climate studies too. Thanks to its fine

spectral resolution, signal-to-noise ratio and wide spectrum range, it is a precious resource for detecting trace gases like ozone, methane and carbon monoxide, as well as clouds, aerosols and greenhouse gases (Phulpin et al., 2002; Clerbaux et al., 2009; Hilton et al., 2012).

For the purposes of this study, IASI Level 1C (L1C) data are exploited and gathered on the EUMETSAT Earth Observation Portal (`https://eoportal.eumetsat.int`, last access: 2 October 2022) in a NetCDF format including the full spectral range. The data pre-treatement and the spectral area covered in this study will be illustrated, together with the experimental set-up, in Section 4.

2.2 Datasets for Comparison and Validation

In order to validate the results of IASI L1C data assimilation in MOCAGE, different sets of remote sensed and *in situ* data have been used. Their source and main features are briefly described here below.

2.2.1 OMI

The Ozone Monitoring Instrument (OMI) is an ultraviolet/visible (UV/VIS) nadir solar backscatter spectrometer (Levelt et al., 2006), flying since 2004 on board the National Aeronautics and Space Administration (NASA) Aura spacecraft. It provides, from a polar Sun-synchronous pattern, a nearly global daily coverage. OMI is able to measure trace gases, aerosols, cloud pressure and coverage with a footprint of 13x24 km^2 at nadir. In addition to monitoring the ozone layer, the data it collects are used with success in a wide range of research and operational applications, such as pollutants, air quality and climate change (e.g. Levelt et al., 2018).

Daily ozone total columns recorded by OMI have been used in this study to have a satellited sensed source of O_3 for validation (`https://acdisc.gesdisc.eosdis.nasa.gov/data/Aura_OMI_Level3/OMTO3e.003/`, last access: 2 October 2022).

2.2.2 Ozone Radiosounding

Meteorological services launch ozonsondes with an almost weekly frequency, in order to acquire *in situ* measurements of the atmospheric state. This kind of instrument provides accurate measurements of O_3 profiles up to heights of the order of 10 hPa and with a vertical resolution of 150−200 m.

The radiosonde network is not homogeneous across the globe. It is strongly dependent on the procedures of the meteorological sites in the measurements campaigns themselves and on the weather conditions that do or do not allow the sondes to fly efficiently. For these and many other reasons, the radiosounding geographical distribution is also uneven and in general more representative in the northern mid-latitudes.

As the most accurate measurements of tropospheric O_3 obtainable for years, they served, for many applications, including the validation of satellite products (e.g. Worden et al., 2008; Boxe et al., 2010; Dufour et al., 2012) and models (e.g. Geer et al., 2006).

The ozone profiles used in this study in order to validate the model simulations, are those collected and distributed by the World Ozone and Ultraviolet Data Centre (WOUDC; `https://woudc.org`, last access: 2 October 2022). The localisation of the 25 stations available is displayed in Figure 1. For the period of study that has been taken into account (4.1), there is a total amount of 296 available observed profiles.

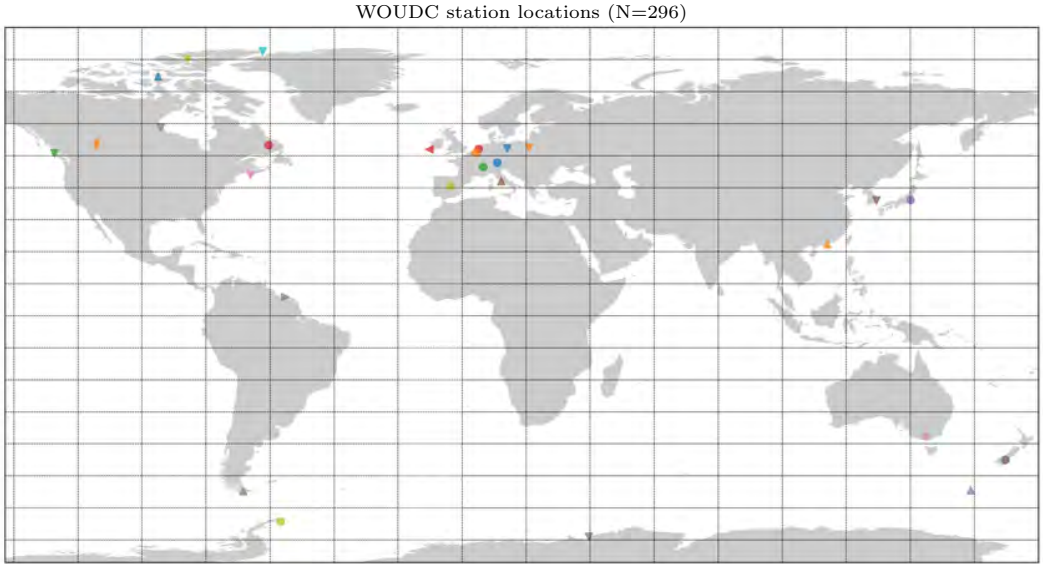


Figure 1. Localisation of the 25 stations provided by the World Ozone and Ultraviolet Data Center (WOUDC). Each station is depicted through a different symbol and color. In this case, for the time period treated, N=296 observations are available.

3 Methods

3.1 Chemical Transport Model: MOCAGE

The MOCAGE model is an off-line three-dimensional CTM that has been developed at Météo-France since 2000 (e.g. Josse et al., 2004; Sič et al., 2015; Guth et al., 2016). It has been exploited along the two last decades for a wide range of operational and research applications. For instance, it served for several studies aiming to evaluate the climate change impact on atmospheric chemistry (e.g. Teyssèdre et al., 2007; Lacressonnière et al., 2012; Lamarque et al., 2013; Watson et al., 2016), as well as the trace gases transport throughout the troposphere (Morgenstern et al., 2017; Orbe et al., 2018). MOCAGE has been used to explore the Upper Troposphere – Lower Stratosphere (UTLS) structure and to evaluate the impact of biomass burning plumes (e.g. Cussac et al., 2020) or to estimate the impact of volcanic sulfur emissions on the tropospheric sulfur budget (Lamotte et al., 2021). MOCAGE is also a precious resource for the air quality monitoring and forecasting on the French Prev'Air platform (Rouil et al., 2009) and over Europe within the Monitoring Atmospheric Composition and Climate (MACC) project (Marécal et al., 2015).

This model uses different chemical schemes in order to reproduce the atmospherical chemical composition: the REactive Processes Ruling the Ozone BUdget in the Stratosphere (REPROBUS) is used for the stratosphere (Lefevre et al., 1994), while, for the tropospheric representation the Regional Atmospheric Chemistry Mechanism (RACM) is exploited (Stockwell et al., 1997). Through the combination of the two aforementioned schemes (called RACMOBUS), MOCAGE is able to simulate 118 gaseous species, 434 chemical reactions, primary aerosols and secondary inorganic aerosols.

At present, MOCAGE supplies two geographical configurations for the use. It covers in operational mode the global scale (GLOB11), down to a smaller continental domain (MACC01), using two-way nested grids capacity. The GLOB11 configuration has a 1° longitude × 1° latitude horizontal resolution. The MACC01 presents a thinner resolution of 0.1° longitude × 0.1° latitude (approximately 10 km at the latitude of 45°N) and it is bounded in 28°N, 26°W and 72°N, 46°E.

For the vertical levels, MOCAGE uses σ-pressure vertical coordinates (Eckermann, 2009). Through this system, on the calculation of which we will not enter into detail in this paper, each grid point is represented by a value above the surface and the topography is not a constraint. The model hence has a non-uniform vertical resolution: 47 vertical altitude-pressure levels from the surface up to 5 hPa. The levels are denser near the surface, with a resolution of about 40 m in the lower troposphere and 800 m in the lower stratosphere. A 60 hybrid levels version is also used in research mode. This consists of the 47 levels computed as just described, plus 13 additional levels going up to 0.1 hPa. Resolutions in upper stratosphere is around 2 km.

MOCAGE CTM runs in an off-line mode. Depending on the application it can be coupled with a general circulation climate model (for climate studies), or with NWP models (e.g. for near real time applications). The core of the chemical reactions used in MOCAGE is also exploited on-line into the Integrated Forecast System (IFS) (Huijnen et al., 2019; Williams et al., 2022).

In this study MOCAGE has been used off-line and the meteorological forcing comes from the Météo-France's operational global NWP model ARPEGE (Action de Recherche Petite Echelle Grande Echelle) (Courtier et al., 1991). The global configuration (GLOB11) has been used on 60 hybrid vertical levels. Indeed, simulating the full atmosphere, including the ozone, will be beneficial for radiance simulations through the radiative transfer.

3.2 Assimilation algorithm

The assimilation system used within MOCAGE has been originally defined in the ASSET (Assimilation of Envisat) project (Lahoz et al., 2007). It was jointly developed by CERFACS (Centre Européen de Recherche et de Formation Avancée en Calcul Scientifique) and Météo France and further refined over the years. It has already been exploited for many studies on the assimilation of chemical data [e.g Massart et al. (2009); Emili et al. (2014, 2019); El Aabaribaoune et al. (2021)], and also on aerosols assimilation (e.g., Sič et al., 2015; Descheemaecker et al., 2019; El Amraoui et al., 2022), in the study of the exchanges between troposphere and stratosphere (e.g., El Amraoui et al., 2010) and in many other fields.

For this work we use the three-dimensional variational (3D-Var) method, the aim of which is to look for the best representation of the atmospheric state, or in other words the best compromise between the background model state and the observations. This is done by minimising the cost function that follows:

$$J(\mathbf{x}) = \frac{1}{2}\,(\mathbf{x}-\mathbf{x}^{\mathrm{b}})^{\mathrm{T}}\,\mathbf{B}^{-1}\,(\mathbf{x}-\mathbf{x}^{\mathrm{b}}) + \frac{1}{2}\,[\mathbf{y}-H(\mathbf{x})]^{\mathrm{T}}\,\mathbf{R}^{-1}\,[\mathbf{y}-H(\mathbf{x})] \tag{1}$$

where $\mathbf{y}$ is the vector of the observations, while $\mathbf{x}_b$ and $\mathbf{x}$ the *a priori* background and the model state vector respectively. The state that minimises the cost function $J(\mathbf{x})$ will then be defined as $\mathbf{x}_a$, i.e. the analysis state. H is the observation operator, that is usually a Radiative Transfer Model (RTM) whose function is to transform a model state to a vector comparable to the observed radiances (or *vice versa*). In this work this function is covered by the Radiative Transfer for TOVS (RTTOV) version 12 (Saunders et al., 2018) in clear-sky conditions (the scattering effect of clouds and aerosols are not taken into account). Last but not least, the $\mathbf{B}$ matrix and the covariance matrix of the observation errors $\mathbf{R}$ are two essential components in the equation, since they allow each term to be given its proper weight.

The 3D-Var has been used with a hourly assimilation window.

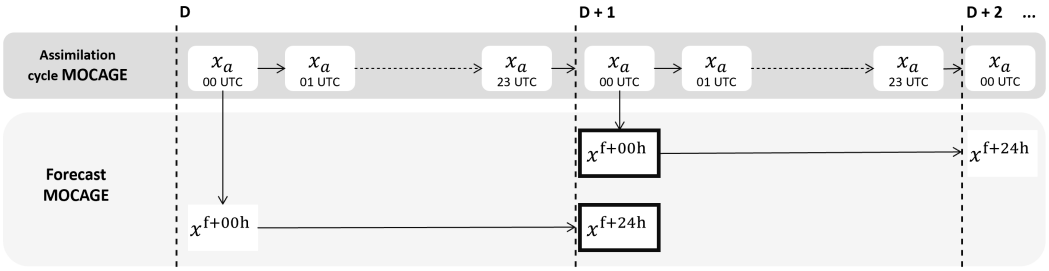


Figure 2. An illustrative chart of the NMC method using MOCAGE. x_a represents the hourly analyses obtained from MOCAGE. $x^{\mathrm{f+00h}}$ is the forecast at 0h initialised with the assimilation x_a at time 00 UTC (the two of them are superimposable). Finally, $x^{\mathrm{f+24h}}$ is the forecast at +24h. The forecasts $x^{\mathrm{f+24h}}$ for day D and $x^{\mathrm{f+00h}}$ for day D+1 are valid at the same time.

3.3 Error Estimation

It is extremely important to assign the correct error to all the parameters that are involved in the cost function minimisation. This is why the error estimation becomes a crucial step in the method. Accordingly, the present subsection will describe the choices the authors made for the purpose of this project.

3.3.1 Background Errors

In previous studies the background-error standard deviation was assumed to be proportional to the ozone concentration itself. Emili et al. (2014) and Peiro et al. (2018) chose to use an error varying along the vertical column and expressed as a percentage of the O_3 background profile: a percentage of 15% was attributed to the troposphere and a smaller one of 5% to the stratosphere. Emili et al. (2019), comparing the standard deviation of a free model simulation against independent observations, actually, show that the error is lower in stratosphere, larger in free troposphere with the highest values near the tropopause. In this latter study and in the follow-up El Aabaribaoune et al. (2021), however, these percentages have been refined up to the value of 2% above 50 hPa and 10% below, since the model itself had been upgraded compared to the prior works. In the present study, as a more recent version of the model MOCAGE is used, we prescribe 2% all over the entire atmospheric column in order to compute the background standard deviation, i.e. the square root of the diagonal of the first $\mathbf{B}$ we evaluate (named $\mathbf{B}_{\mathrm{perc}}$ from here on).

As previously explained, one of the aims of this paper is to assess the impact of using a $\mathbf{B}$ matrix evaluated in a way that differs from previous studies in literature, in the wake of which this work stands. Accordingly, also a second strategy has been chosen for the computation of background errors, relying on a method introduced by the National Meteorological Centers (NMC) (Parrish & Derber, 1992). This is a technique that defines background errors using differences in numerical weather prediction (NWP) forecasts valid at the same time.

To apply this method, we run a cycle of hourly assimilations in MOCAGE over the 3-month period used in this work as case study (see Subsection 4.1 for more details). As shown in Figure 2, each day D will provide an assimilation x_a for every hour of the day. At the same time, we run a MOCAGE forecast up to a 24-hour forecast range for each day D, initialised with the x_a assimilation of the same day at 00 UTC. The forecasts $x^{\mathrm{f+24h}}$ for day D and $x^{\mathrm{f+00h}}$ for day D+1 are valid at the same time. Thereafter the covariance of the differences between each couple are computed, averaged over time and longitudes. This produces a 2D-field that represents the standard deviations of the background errors, varying along latitudes and vertical levels.

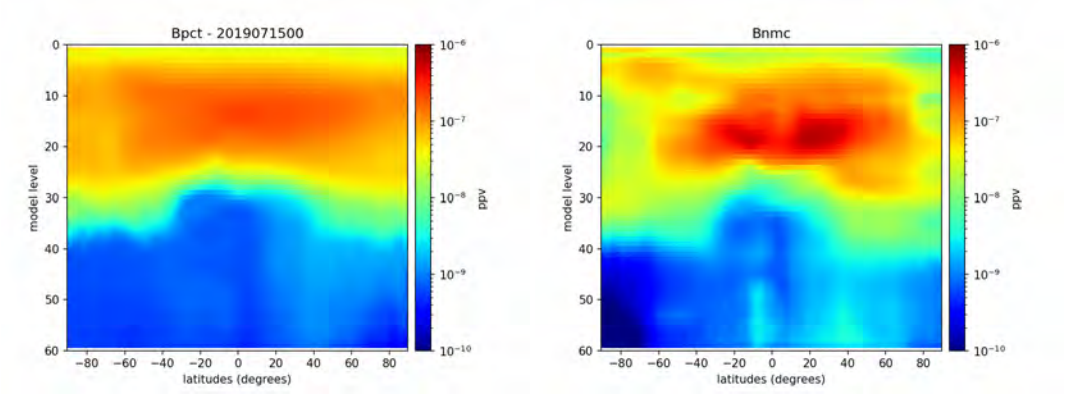


Figure 3. Background-error standard deviations (σ^B) for $\mathbf{B}_{\text{perc}}$ in the left panel and $\mathbf{B}_{nmc}$ on the right. One random day (15 July 2019 at 00 UTC) is chosen for $\mathbf{B}_{\text{perc}}$, representative of the general behaviour of the period of study, and values are averaged over the longitudes.

The background-error standard deviations (σ^B) are shown in Figure 3 for both $\mathbf{B}_{\text{perc}}$ and $\mathbf{B}_{\text{nmc}}$. While a 2D-field is obtained and displayed from the above-mentioned NMC-method implementation, we select one random day for $\mathbf{B}_{\text{perc}}$, that is representative of the general behaviour of the period of study (values are averaged over the longitudes). The panels show similar structures, but different intensities in many areas. The $\mathbf{B}_{\text{nmc}}$ case reveals stronger values at the Tropics and Mid-latitudes (50°S-50°N) and weaker at the Poles. The error gradient at the UTLS is lower in the atmospheric column for $\mathbf{B}_{\text{nmc}}$ compared to $\mathbf{B}_{\text{perc}}$. Furthermore, the NMC case clearly shows stronger tropospheric σ^B values around the Tropics and Northern Mid-latitudes (20°S-90°N). On the other hand, over the Antarctic region the tropospehric σ^B is weaker in the NMC case compared to the other. Notice that the higher the σ^B values, the lower the weight attributed to the background during the assimilation process and, consequently, the larger the role played by the observations.

3.3.2 Observation Errors

For the estimation of a full $\mathbf{R}$ matrix, just as in El Aabaribaoune et al. (2021), we apply the diagnostic procedure developed by Desroziers et al. (2005) and currently exploited by many NWP centres. Through this method, variances and covariances of the the observation errors are computed using *observation-minus-analysis* and *observation-minus-background* statistics:

$$\mathbf{R} = E\left\{ \left[\mathbf{y} - H(\mathbf{x}_{\text{a}})\right] \left[\mathbf{y} - H(\mathbf{x}_{\text{b}})\right]^{\text{T}} \right\} \tag{2}$$

where E is the statistical expectation operator.

Values comparable to what described by El Aabaribaoune et al. (2021) are obtained in this study.

4 Experimental set-up

4.1 Period of Study

The experiment covers the time period running from the 15th of May to the 31st of August 2019. The first 16 days of this period are used as spin-up period and will not be included in the evaluations carried out all over the the study. The following will focus on

the three-month period ranging from the 1st of June to the 31st of August. During this period, data from two or three IASIs are simultaneously available depending on the day.

Compared to the time period covered by Emili et al. (2019) and El Aabaribaoune et al. (2021), the one used in this paper is more recent and wider (although all the cases concern Northern Hemisphere summer). This does not make our study directly comparable to the previous ones, but still allows conclusions to be drawn on the new tests performed and the contribution to the overall method.

4.2 Data Thinning

In treating IASI L1c data to feed the model, a pre-treatment has been applied to the original dataset.

A thinning of the data has been planned in order to make the assimilation more efficient and to avoid geographically redundant data. First, only pixel number 1 has been considered out of the four pixels in each IASI field of regard. Additionally, one must take into account that the MOCAGE GLOB11 resolution ($1^\circ \times 1^\circ$) is much wider than the IASI pixel size. Consequently, to avoid any spatial averaging of the IASI data, only one pixel per 1-degree box is kept in each one-hour slot.

In order to avoid cloud contamination, which reduces RT accuracy and increases errors in assimilation, the data were filtered using the Cloud Mask information from the Advanced Very High Resolution Radiometer (AVHRR) within the IASI pixel: only pixels with less than 1% of clouds are kept. Desert dusts can also affect the IASI spectrum, thus a detection filter has been applied as well.

4.3 Spectral Range and Exploited IASI Channels

For this study three subsets of IASI channels have been taken into account. The first and largest consists of 284 channels bounded between 980 and 1100 cm^{-1}. From this spectral range, a few wavelengths have been excluded in order to remove channels strongly affected by H_2O. This approach is the same exploited in the previous studies on ozone L1 radiance assimilation (Emili et al., 2019; El Aabaribaoune et al., 2021), which was in turn in line with the choice of the spectral channels already used for IASI L2 ozone retrievals (e.g. Barret et al., 2011)].

For the present study, however, smaller subsets of channels have also been extracted from the initial group of 284. To do so the optimal selection method (see i.e. Rodgers, 1998; Rabier et al., 2002; Vittorioso et al., 2021) has been used and the *Degrees of Freedom for Signal* (DFS) was the figure of merit chosen. This latter shows the improvement of the analyse error over the background one: $\mathrm{DFS} := \mathrm{Tr}\,(\mathbf{I} - \mathbf{A}\mathbf{B}^{-1})$ where $\mathbf{I}$ is the identity matrix and $\mathbf{A}$ and $\mathbf{B}$ the covariance matrices of the analysis and background errors respectively.

Through this selection process, different groups of channels can be isolated. However, for the purposes of this work only two groups have been retained, consisting of 76 and 34 channels each. By selecting only 34 over 284, the 83% of the maximum possible total DFS value (achieved if all channels are taken) is reached. On the other hand, to have a 90% of the maximum total DFS, 76 channels are needed. Complete channel number and Jacobian lists are provided in Table B1 and Figure B1 respectively (see Appendix B).

The spectral location of the three sets of channels is shown in Figure 4. When selecting the first 76 channels, very few channels in the atmospheric windows are kept. These channels provide a redundant information and for this reason they are rejected by the algorithm of selection we exploit. On the other hand, the selection of channels in the ozone band is well distributed. When decreasing from 76 to 34 channels, the selection among the ozone and the surface spectral regions is well balanced.

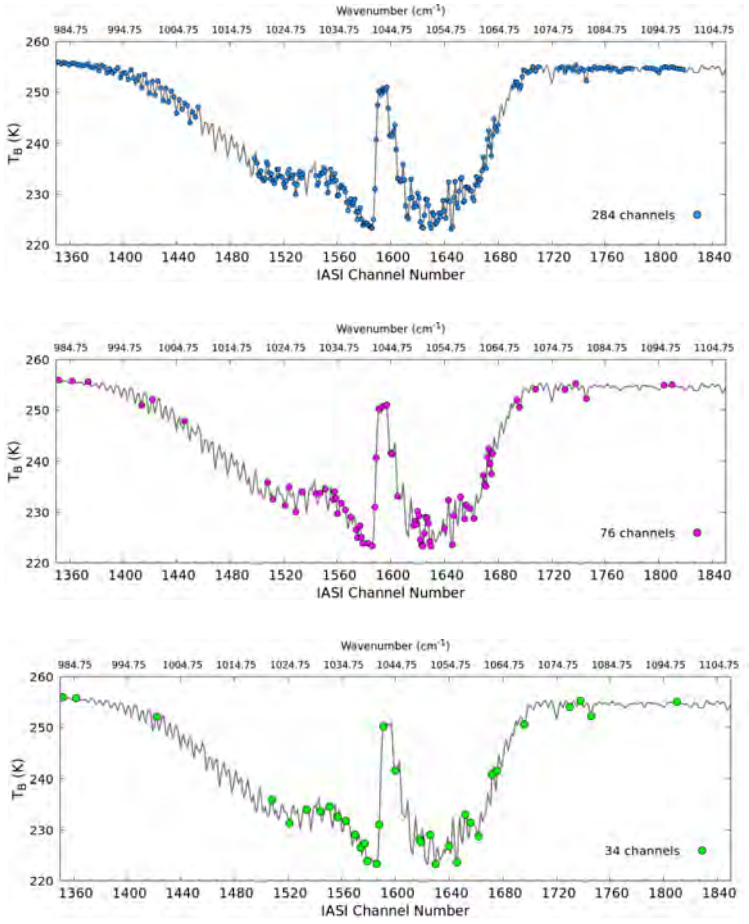


Figure 4. Spectral location of the three groups of channels exploited: 284 (top panel), 76 (middle) and 34 (bottom).

5 Assimilation Results

5.1 Impact of the B matrix ($\mathbf{B_{perc}}$ vs $\mathbf{B_{nmc}}$)

In order to assess the impact of using different correlation matrices of the background errors (i.e. $\mathbf{B}_{\text{perc}}$ and $\mathbf{B}_{\text{nmc}}$, computed as in 3.3.1) on the assimilation process, we performed two different assimilation runs of the MOCAGE model. Both runs assimilate the largest set of 284 channels and use the full $\mathbf{R}$ matrix (3.3.2). The input data and the experimental setups only differ in the $\mathbf{B}$ matrix exploited.

The average of the ozone field difference between the two experiments over the three-month period (1st June - 31st August 2019) has been computed and it is shown as zonal cross section in Figure 5. It is observed that the vertical structure is modified when using $\mathbf{B}_{\text{nmc}}$. More ozone is observed in the upper troposphere at all latitudes, while smaller quantities are found in lower stratosphere. In the stratosphere, on the other hand, the structure is mixed along the latitude bands, alternating increases and decreases in the ozone field. Values vary between -800 and 800 ppbv. This variation is of the same magnitude of what obtained by El Aabaribaoune et al. (2021) when studying the impact of improving the observation-errors. In the Tropics and Northern-Hemisphere (NH) the difference is slightly negative (up to -2 ppbv), indicating that the ozone values are smaller when using the $\mathbf{B}_{\text{nmc}}$.

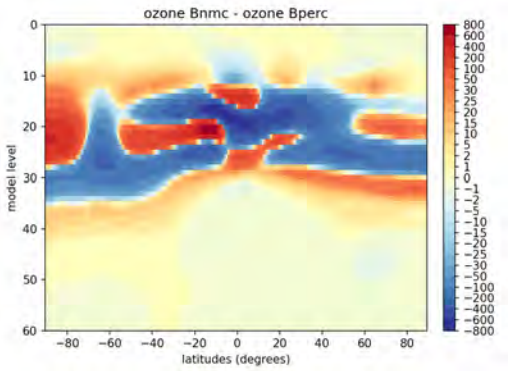


Figure 5. Average of the ozone field difference between experiments using $\mathbf{B}_{\mathrm{nmc}}$ and $\mathbf{B}_{\mathrm{perc}}$, from the 1st of June to the 31st of August 2019. Values are in ppbv.

5.1.1 Verification against OMI

The outcomes of these experiments have first been evaluated in comparison to OMI data. Figure 6 shows three maps of ozone total columns averaged over the period of study ranging from the 1st of June to the 31st of August 2019. These fields are obtained from OMI observations (top panel) and the two IASI observation assimilations in MOCAGE using different $\mathbf{B}$ matrices ($\mathbf{B}_{\mathrm{perc}}$ bottom left and $\mathbf{B}_{\mathrm{nmc}}$ right). The two assimilation runs present, at first sight, very similar spatial structures. However, a more quantitative analysis shows differences especially at mid-latitudes. Several areas over Atlantic Ocean and over the USA and Europe show lower values of ozone total columns in the case of $\mathbf{B}_{\mathrm{nmc}}$ use, locating this case slightly closer to the real observations.

A direct inter-comparison of the mean bias (MB - for statistical indices definition see Appendix A) averaged over the 3-month period has been performed as well and it is displayed in Figure 7 (top panels). From examination of these two charts, it is noticeable that, when using the $\mathbf{B}_{\mathrm{nmc}}$ matrix for the assimilation run, there is a global reduction in bias compared to observations than when using $\mathbf{B}_{\mathrm{perc}}$. Table 1 summarises some of the statistics computed with respect to OMI data, that have been averaged over the period of study, but also over the entire globe. Results show that a bias reduction of 4 DU larger is found when using $\mathbf{B}_{\mathrm{nmc}}$.

A measure of the error occurring between the MOCAGE assimilation runs using a different $\mathbf{B}$ and the OMI observations themselves, is obtained through the computation of the root mean square error (RMSE - see Appendix A). Figure 7 (bottom panels) show the RMSE values averaged over the 3-month period of study. Looking at these results immediately reveals very clearly the reduction of the error in the case exploiting $\mathbf{B}_{\mathrm{nmc}}$ compared to the one using the $\mathbf{B}_{\mathrm{perc}}$. This is true not only at mid-latitudes, but also over the rest of the globe. This result is confirmed when looking again at the statistics averaged over the globe (Table 1). The RMSE values, just as for the MB, result weaker when assimilating IASI channels using $\mathbf{B}_{\mathrm{nmc}}$. A global error reduction of 4 DU on average is thus observed trough the exploitation of this matrix rather than the $\mathbf{B}_{\mathrm{perc}}$. On the other hand, Pearson's Correlation values remain the same in both cases.

5.1.2 WOUDC Ozonesondes

In order to get an overview of what happens along the atmospheric column, a comparison between each of the assimilation runs and the data collected from WOUDC ozonesondes has been carried out as well. Top left panel in Figure 8 shows the RMSE computed between

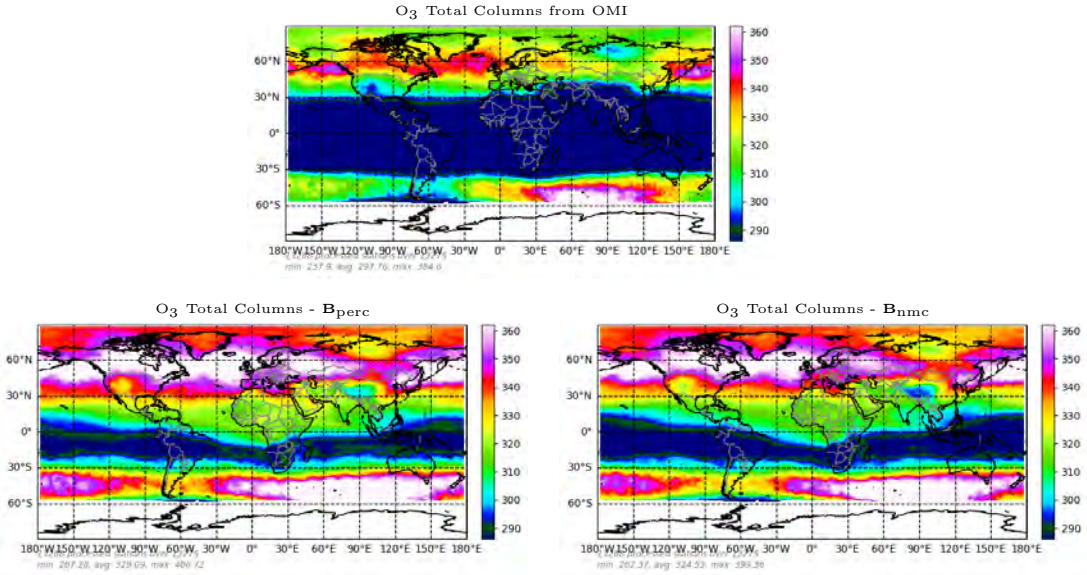


Figure 6. O_3 total columns for OMI observations (top panel), from assimilation in MOCAGE using $\mathbf{B}_{\text{perc}}$ (bottom left) and from assimilation using $\mathbf{B}_{\text{nmc}}$ (bottom right). The values are averaged over the time period ranging from the 1st of June to the 31st of August 2019, in Dobson Units (DU).

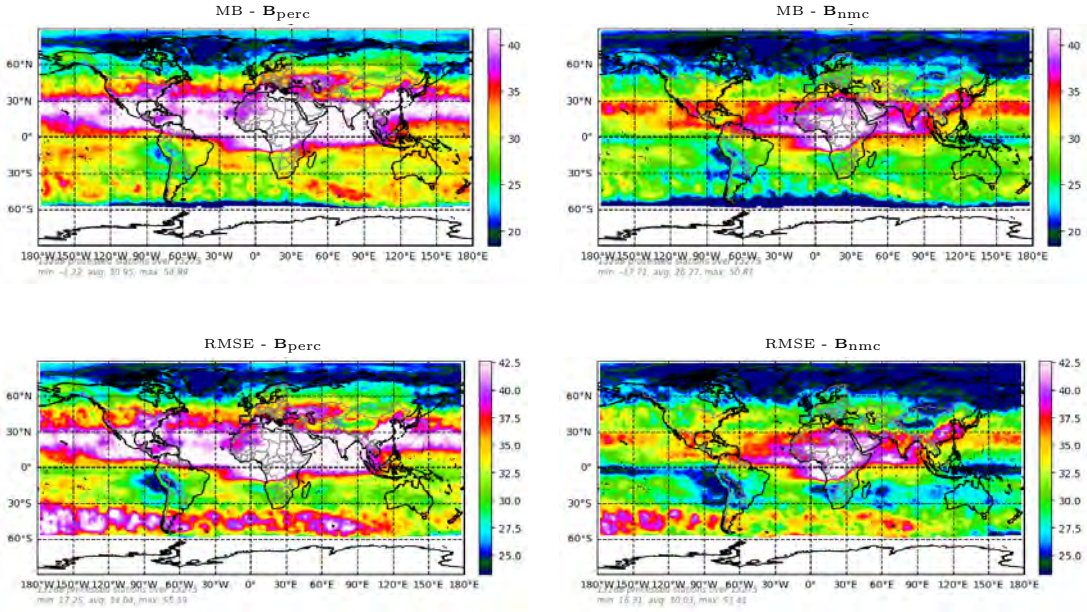


Figure 7. MB (top panels) and RMSE (bottom) values computed with respect to OMI data for the assimilation run outcomes using the $\mathbf{B}_{\text{perc}}$ (left) and the $\mathbf{B}_{\text{nmc}}$ (right). Values averaged over the time period ranging from the 1st of June to the 31st of August 2019, in Dobson Units (DU).

Table 1. Statistics from the comparison of OMI observations and the assimilation of the 284 IASI channels into MOCAGE using $\mathbf{B}_{\text{perc}}$ and $\mathbf{B}_{\text{nmc}}$. The values here listed are averaged over the globe and the 3-month case study period (1st of June - 31st of August 2019). For the definition of the statistical indicators see Appendix A.

	MB (DU)	RMSE (DU)	Correlation (%)
$\mathbf{B}_{\text{perc}}$	31	34	80
$\mathbf{B}_{\text{nmc}}$	27	30	80

each assimilation run results and the measurements from ozonesondes, averaged over the globe along the period of study. The red and blue curves refer to the $\mathbf{B}_{\text{perc}}$ and the $\mathbf{B}_{\text{nmc}}$ case respectively. A noticeable error reduction (of the order of 200 ppbv) is observed between 30 and 70 hPa when using the $\mathbf{B}_{\text{nmc}}$. Another reduction, even if weaker, is observed around 100 hPa. In the lower levels, on the other hand, the curves are very close to each other.

When examining the other panels in Figure 8, which show the RMSE computed over different latitude bands, it is observed that the largest contribution to the results over the globe is given by the Northern-Hemisphere (NH) mid-latitudes. It is from this geographical area, in fact, that most of the used ozonesonde measurements come from: 196 out of 296 observations come from the stations located along this latitude band. On the other hand, if referring to the σ^Bs illustrated in Figure 3, we observe that in the NH the background error values for $\mathbf{B}_{\text{nmc}}$ are higher compared to $\mathbf{B}_{\text{perc}}$. As already explained, the higher the background-error the higher the weight attributed to the observations. This means that in this region, the observations play a stronger role when using $\mathbf{B}_{\text{nmc}}$, thus causing a stronger reduction in the error compared to the $\mathbf{B}_{\text{perc}}$ case. Indeed the improvement is located at same height corresponding to the maximum sensitivity of the Jacobians (see Figure B1).

In the Southern-Hemisphere (SH) region the improvement brought by $\mathbf{B}_{\text{nmc}}$ is less evident. There is a degradation between 20 and 90 hPa. At the Tropics there is a stronger impact between 15 and 40 hPa.

Finally, it should be emphasised, that the improvements in each region made by using $\mathbf{B}_{\text{nmc}}$, correspond to the altitudes at which, in the region in question, the highest ozone concentrations are found on average. This is of course related also to the UTLS location, which varies depending on the geographical area.

The results presented so far show a global reduction in the error if a $\mathbf{B}_{\text{nmc}}$ matrix is used for the representation of the background error. $\mathbf{B}_{\text{nmc}}$ will thus be used for the later part of the study.

5.2 Evaluation of the Channel Groups

After selecting the best $\mathbf{B}$ matrix, the study goes on with an investigation of the number of channels used for the assimilation. Three different assimilation runs are performed testing the three channel groups introduced in 4.3 (284, 76 and 34) and using the $\mathbf{B}_{\text{nmc}}$ matrix for background errors in each case.

5.2.1 Verification against OMI

Also for this part of the work, a first assessment was made by comparing data obtained from the assimilation to those retrieved by OMI.

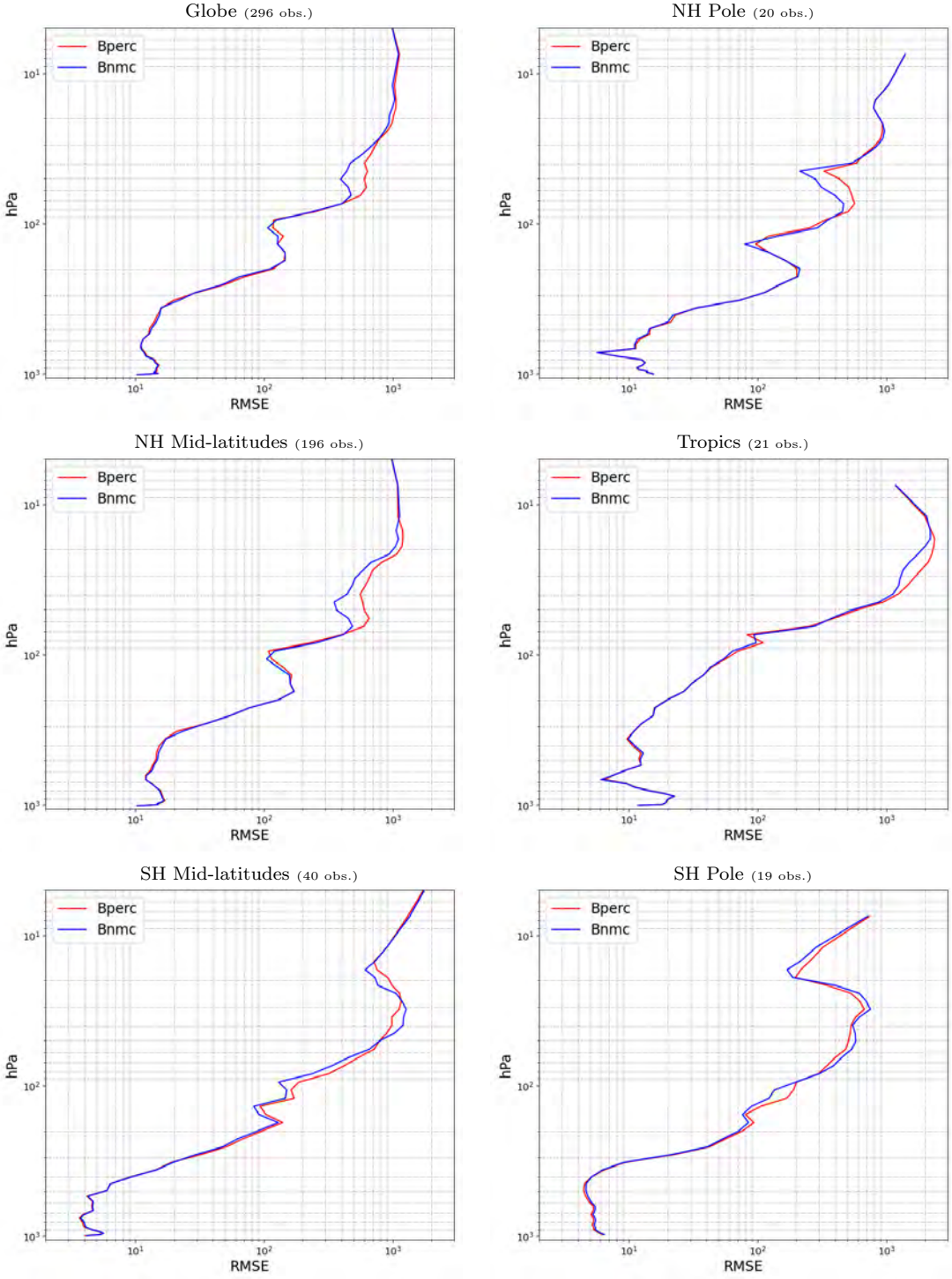


Figure 8. RMSE (in ppbv) between each assimilation run and the observations from the WOUDC ozonsondes. Each each plot represents the results referring to a certain geographical regions: Globe, Northern Hemisphere (NH) Pole, NH Mid-latitudes, Tropics, Southern Hemisphere (SH) Mid-latitudes, SH Pole. The red curve always refers to the values related to the $\mathbf{B}_{\text{perc}}$ and the blue one to the $\mathbf{B}_{\text{nmc}}$. The values are averaged over the time period from the 1st of June to the 31st of August 2019.

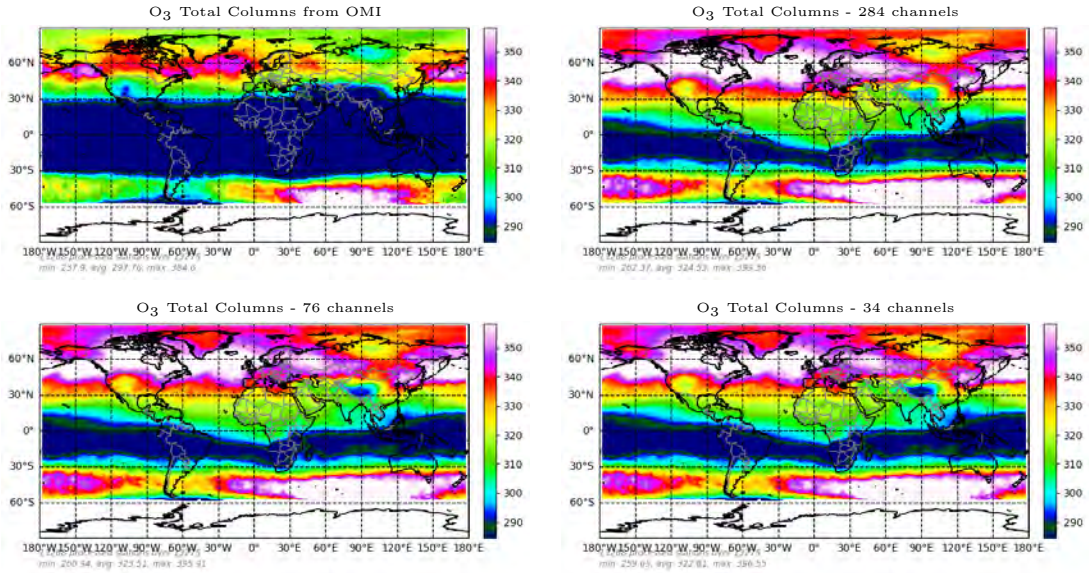


Figure 9. O_3 total columns for OMI observations (top left panel), from assimilation in MOCAGE, using $\mathbf{B}_{\text{perc}}$, of 284 (top right), 76 (bottom left) and 34 (bottom right). The values are averaged over the time period ranging from the 1st of June to the 31st of August 2019.

In Figure 9 the O_3 values in total columns from OMI and the output of the MOCAGE with assimilation of each channel group are displayed. All of the three experiments using assimilation of IASI channels overestimate the ozone field almost all over the globe, except in the Southern Tropical region. However, looking at the output obtained through the use of the smallest subsets (76 and 34 channels) one can observe values that are slightly closer to the OMI observations. Between these two experiments, the one providing results closest to reality, is the 34 channel group. This can be witnessed especially over the geographical region between 0° and 30°N. A similar conclusion can be drawn, but more in detail, by examining the MB charts in Figure 10. Regions with smaller bias values can be found in the 34 channel assimilation case. Examples are observed between 30° and 60°S or over North America and Pacific and Atlantic Oceans. However, if we compare the average over the entire domain of the statistical parameters resumed in Table 2, the same value of MB is obtained for both 34 and 76 assimilated channels. This means that there are areas where one channel subset performs better then the other and *vice versa*.

RMSE for each group of channels compared to OMI observations is shown in Figure 10. As for the MB, the 284 assimilation case is the one producing larger values. The highest reduction in the error is detected when assimilating only 34 channels. Comparing the 34 to the 76 chart, many areas (e.g. Northern Pacific and North America) show a stronger reduction in the error when a smaller number of channels is assimilated. On the other hand, the Southern-Tropical regions seems to be affected by a stronger error when less channels are assimilated. If looking again at the average over the globe of this score (Table 2), the same value is found for both channel groups. Correlations values are quite close to each other for the three cases.

5.2.2 WOUDC Ozonesondes

A further comparison has been performed against ozonesondes data from WOUDC network and the RMSE values computed are resumed in Figure 11.

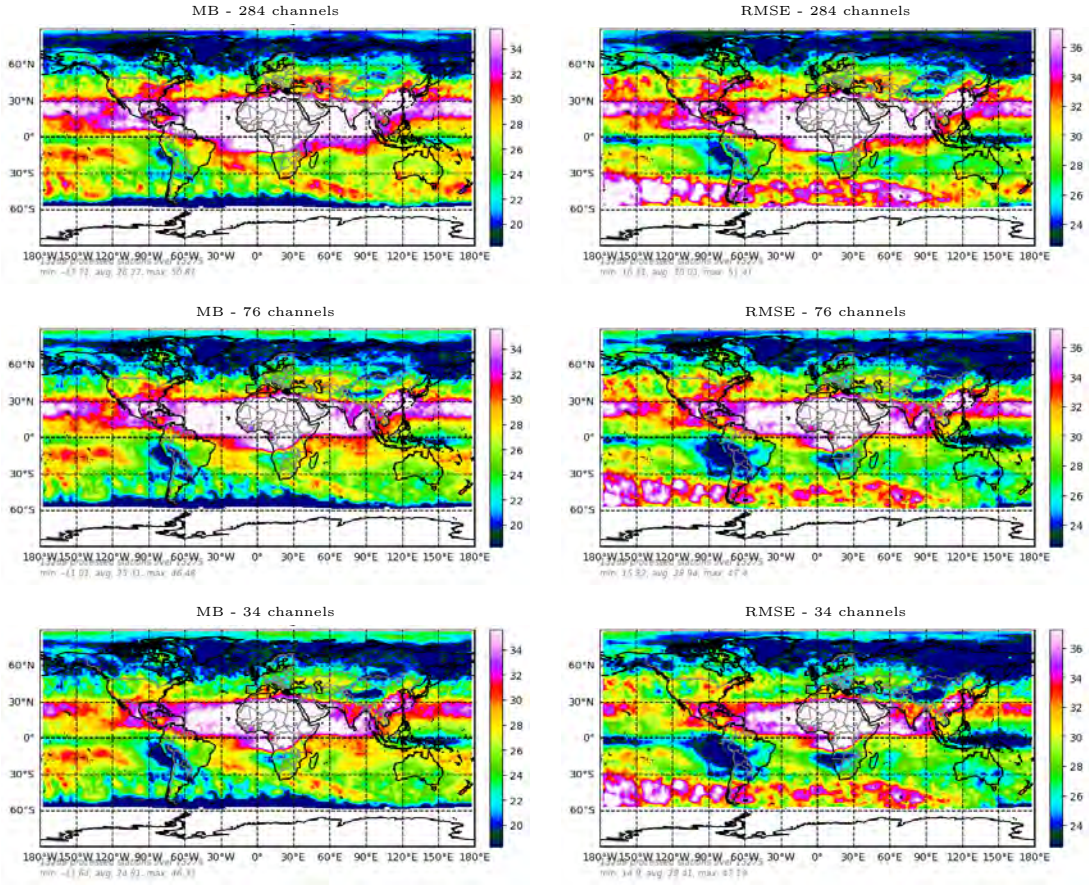


Figure 10. MB (left panels) and RMSE (right) values computed with respect to OMI data for MOCAGE outcomes assimilating 284 (top), 76 (middle) and 34 (bottom) IASI channels. Values averaged over the study time period from the 1st of June to the 31st of August 2019.

Table 2. Statistics of the comparison of OMI observations and the assimilation of 284, 76 and 34 IASI channels into MOCAGE using $\mathbf{B}_{\text{nmc}}$. The values here listed are averaged over the globe and the 3-month case study period (1st of June - 31st of August 2019). For the definition of the statistical indicators see Appendix A.

Channels	MB (DU)	RMSE (DU)	Correlation (%)
284	27	30	80
76	25	28	81
34	25	28	81

Looking at the results over the globe, a stronger error reduction is observed when using the two smaller subsets. More in details, improvements are registered between 15 and 35 hPa for the 76 selection and between 15 and 70 hPa for the 34. The reduction reaches a maximum of 300 ppbv between 20 and 30 hPa3. A moderate improvement is also detected in the band between 100 and 200 hPa for both subsets, but larger for the 34. Some difference in lower troposphere is also found, but it may not be significant.

The global scores are for sure influenced by the contribution of each different latitude region. However, as already explained in 5.1.2, due to the greatest amount of observations provided by the NH Mid-latitudes, it is this area that most influences the global results. An improvement in stratosphere can be also found in Southern Mid-latitudes and at the Tropics.

The better results obtained using the 34 channel subset can be explained observing the Jacobians in Figure B1. The information appears very redundant for the 284 subset between 10 and 70 hPa. Numerous channels present maxima peaks at the same altitude, with the same vertical behaviour. Consequently, using the smaller 34 group can bring the same information as the full 284 subset. If the observation error was perfectly described and assimilation process fully optimal, the result should be the same in all cases. Nevertheless, as the numerical framework and the error estimation are not always optimal, one can observe a gain when assimilating fewer channels. Selecting sub-groups of channels is of common usage in Numerical Weather Prediction (NWP).

6 Conclusions and Perspectives

In the last few years, efforts have been made in order to implement the direct assimilation of IASI L1c data into the CTM MOCAGE. These studies showed first how to move from L2 products assimilation to L1 (Emili et al., 2019), then how to optimally evaluate the observation errors (El Aabaribaoune et al., 2021). Following on from these previous studies, the present work focused on the impact of using different background errors and on the optimisation of the assimilation process itself through the possible reduction of the number of assimilated channels.

Assessing the impact of the background error is of crucial importance in an assimilation process and it becomes even more interesting when it comes to the assimilation of radiances into a CTM. Although studies already exist using an advanced $\mathbf{B}$ matrix for the assimilation of L2 products (Inness et al., 2013), to the best of the authors' knowledge, no literature exists about background error studies targeted to the assimilation of radiances in a CTM. We assimilated IASI L1C data into MOCAGE by exploiting two different $\mathbf{B}$ over a 3-month period (June to August 2019). The first matrix has been computed using 2% of the ozone profiles themselves ($\mathbf{B}_{perc}$). The second one, on the other hand, was obtained using the NMC method, which allows background errors to be assessed from forecasts valid at the same time ($\mathbf{B}_{nmc}$).

Proceeding to the assimilation of 284 IASI channels, we examined the results obtained using these two matrices by comparing them to data from the OMI satellite instrument. An overall reduction in the error was clearly observed when using the $\mathbf{B}_{nmc}$. A reduction of 4 DU in RMSE was obtained on average over the globe through the exploitation of a $\mathbf{B}_{nmc}$ compared to when using a $\mathbf{B}_{perc}$. From an inter-comparison with ozonesondes data issued by the WOUDC network, a general reduction of the error was again observed when using the $\mathbf{B}_{nmc}$. This improvement, however, was localised in the higher levels of the atmosphere and at the UTLS. Since the $\mathbf{B}_{nmc}$ matrix was the one better performing at this stage, we decided to keep using it the continuation of the study.

The second goal of this paper was to evaluate the impact of the number of channels assimilated. In order to assess the feasibility of reducing such number, three groups of channels have been put to test over the aforementioned period of study. The 284 channels

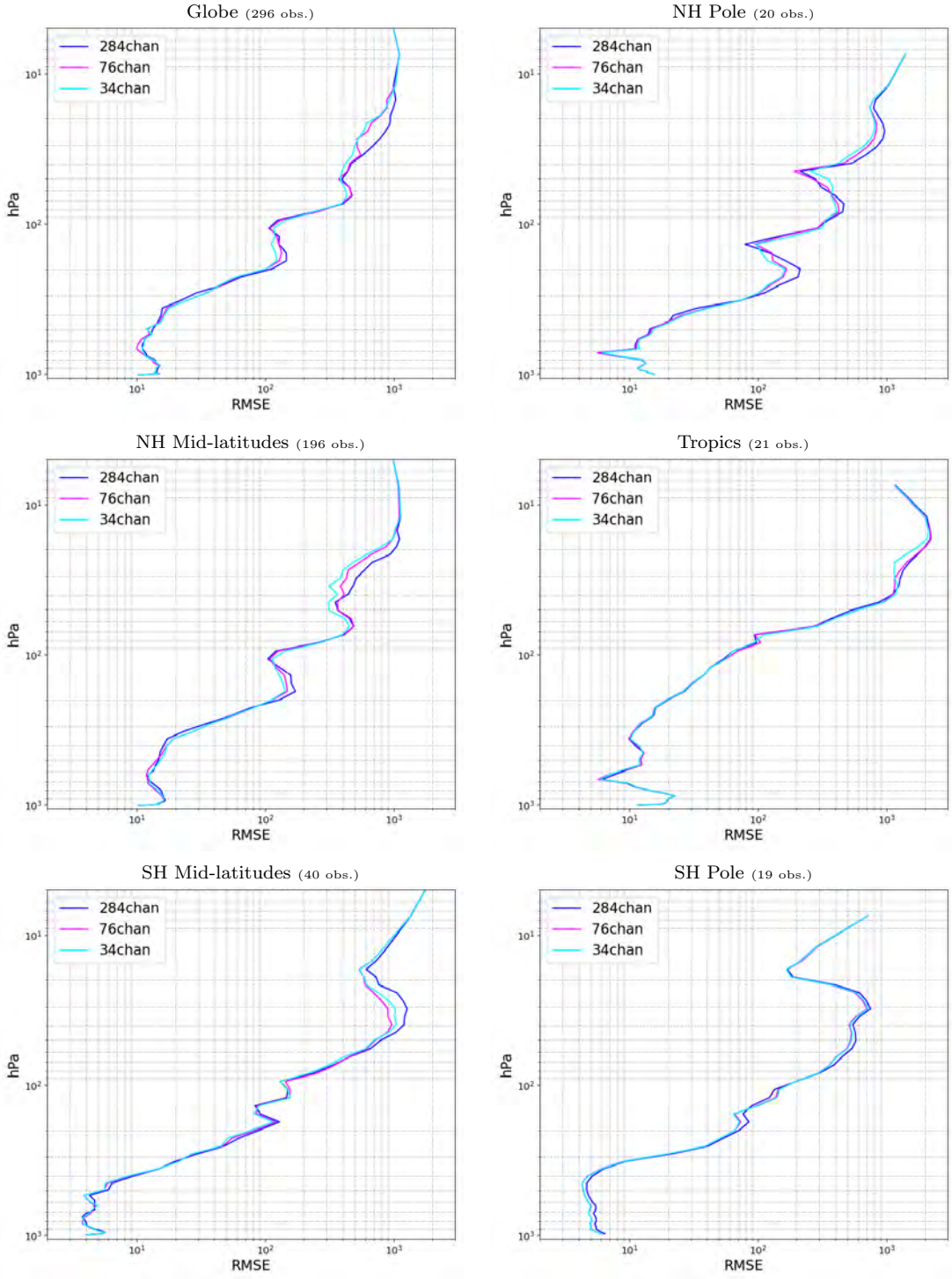


Figure 11. RMSE (in ppbv) between each assimilation run and the observations from *in situ* measurements via WOUDC ozonsondes. Each plot represents the results referring to a specific region: Globe, Northern Hemisphere (NH) Pole, NH Mid-latitudes, Tropics, Southern Hemisphere (SH) Mid-latitudes, SH Pole. The values are averaged over the time period from the 1st of June to the 31st of August 2019.

normally used in literature, plus two smaller subsets of 76 and 34 channels, representing the 90% and 83% of the total DFS respectively, have been used.

From a comparison to OMI data, the smallest subset of channels appeared to be the one providing better results over many areas of the globe. On an average, however, both smaller subsets (76 and 34) provided the same scores for Mean Bias, RMSE and Correlation. The assimilation has also been tested against ozonesondes data. Similar results were found for many areas over the globe and along the atmospheric column by using each of the three subsets. However, an error reduction has been observed in the stratosphere (15 to 70 hPa) and at the UTLS when using fewer channels. The improvement is even stronger when assimilating 34 channels than 76.

We have been able to reduce the number of channels while maintaining a good quality of the analyses. This reduced the computational time of assimilation itself (gain of 50% when assimilation 34 channels instead of 284) and could make operational analysis more efficient and easier to implement.

Overall, the present study helps to improve the process of the assimilation for IASI L1 data into a CMT. More work, however, is still needed to refine the definition of the background-error statistics, for example by using an ensemble based method. This could help to have a refined **B** matrix of the day. Another way to improve the ozone analysis could be the assimilation of other Infrared L1 radiances with the same approach. The Cross-track Infrared Sounder (CrIS), on board the polar orbiting National Oceanic and Atmospheric Administration (NOAA) satellites, or the future Infrared Sounder (IRS), that will fly on board the geostationary Meteosat Third Generation (MTG) series, could be used for this purpose.

Appendix A Statistical Indicators

In order to assess the impact of this study over the model performances and to establish the divergence of the outcomes, a few statistical indicators have been employed. The indicators used are listed and described here below. These are often used to measure the forecast performances of a model, by comparing it to observations. In this study, however, they are mostly used to compare the divergence of different forecasts. Consequently, the forecast role will be covered by one model output and the other model output will be considered as the reference itself.

The mean bias (MB) provides an indication of the difference between two data sets. It can range in $(-\infty, +\infty)$ and it is defined by the following formula:

$$\mathrm{MB} = \frac{1}{N} \sum_i (a_i - b_i) \tag{A1}$$

where a is the parameter to be evaluated through comparison to the reference b over a population N.

The root mean square error (RMSE) is the standard deviation of the errors occurring between a and the reference b. RMSE gives a measure of how these errors spread out. It is defined by the formula

$$\mathrm{RMSE} = \sqrt{\frac{1}{N} \sum_i (a_i - b_i)^2} \tag{A2}$$

and it ranges between $[0, +\infty)$. This indicator is strongly dominated by the largest values because of the square root operation. Accordingly, the interpretation of the RMSE and its reliability may be tricky whenever major outliers are present.

Both MB and RMSE need the knowledge of typical mean values of the involved quantities in order to be well interpreted.

Finally, the Pearson's correlation coefficient is used to test the match in the patterns of the modelled populations. It is computed through the formula:

$$Correlation = \frac{Cov(a,b)}{\sigma_a \sigma_b} = \frac{\frac{1}{N}\sum_i (a_i - \bar{a})\left(b_i - \bar{b}\right)}{\sigma_a \sigma_b} \tag{A3}$$

with $\bar{a}$ and $\bar{b}$ mean values of the tested quantities and σ_a and σ_b the associated standard deviations. This coefficient, ranging in $[-1, 1]$, reaches its maximum value of 1 when $(a_i - \bar{a}) = c\left(b_i - \bar{b}\right)$, with c a positive constant. In that case, the data sets being compared present the same kind of variability, but they are not identical unless the constant c has a value of unity all over each site.

Appendix B Channels for the Assimilation

Table B1 lists the IASI channels exploited for the assimilation performed throughout this study.

Figure B1 shows the Jacobians multiplied 10% of the ozone profile, devided per subsets (34, 76 and 284) and latitude bands.

Table B1. List of the 284 IASI channels (expressed in channel number n) used for this work. The smaller subsets of 76 and 34 channels are shown as well in bold and underlined respectively. Each smaller subset is contained in the previous bigger one.

1350	**1352**	1354	1356	1358	1360	**1362**	1364	1366	1368	1370	1372	**1374**
1376	1378	1380	1382	1384	1386	1388	1390	1392	1394	1396	1398	1400
1402	1404	1406	1408	1410	1412	**1414**	1416	1418	1420	**1422**	1424	1426
1428	1430	1432	1434	1436	1438	1440	1442	1444	**1446**	1448	1450	1452
1454	1456	1498	1500	1501	1502	1503	1504	1505	1506	1507	**1508**	1509
1510	1511	**1512**	1513	1514	1515	1516	1517	1518	1519	1520	**1521**	1522
1523	**1524**	1525	1526	1527	1528	**1529**	1530	1531	1532	1533	**1534**	**1544**
1545	1546	**1547**	1548	1549	1550	**1551**	1552	1553	1554	1555	1556	**1557**
1558	**1559**	**1560**	1561	1562	**1563**	1564	1565	**1566**	1567	1568	1569	**1570**
1571	1572	1573	**1574**	**1575**	1576	**1577**	**1578**	**1579**	1580	1581	1582	**1583**
1584	1585	**1586**	1587	**1588**	**1589**	1590	**1591**	1592	1593	**1594**	1595	1596
1597	1598	1599	**1600**	**1601**	1602	1603	1604	**1605**	1606	1607	1608	1609
1610	1611	1612	1613	1614	1615	1616	**1617**	**1618**	**1619**	**1620**	**1621**	**1622**
1623	**1624**	**1625**	**1626**	**1627**	**1628**	**1629**	**1630**	1631	1632	1633	1634	1635
1636	1637	1638	1639	**1640**	1641	1642	**1643**	1644	1645	**1646**	**1647**	1648
1649	1650	1651	**1652**	1653	1654	**1655**	**1656**	1657	1658	**1659**	1660	1661
1662	1663	1664	1665	1666	1667	1668	**1669**	**1670**	**1671**	**1672**	**1673**	**1674**
1675	**1676**	1677	1678	1679	1680	1692	1693	**1694**	1695	**1696**	1697	1698
1699	1700	1702	1704	1706	**1708**	1710	1726	1728	**1730**	1732	1734	1736
1738	1740	1742	1744	**1746**	1748	1750	1752	1754	1756	1758	1760	1762
1764	1766	1768	1770	1772	1774	1776	1778	1780	1790	1792	1794	1796
1798	1800	1802	**1804**	1806	1808	**1810**	1812	1814	1816	1818		

N.B. To compute the corresponding wave number (ν), apply the formula: $\nu = 0.25(n-1) + 645.00\,\mathrm{cm}^{-1}$.

Acknowledgments

This research has been carried out in the framework of a PhD project co-funded by Météo-France and Thales Alenia Space. We acknowledge EUMETSAT, WOUDC and NASA for providing IASI L1C, ozonesonde and OMI data respectively. We also want to thank Dr.

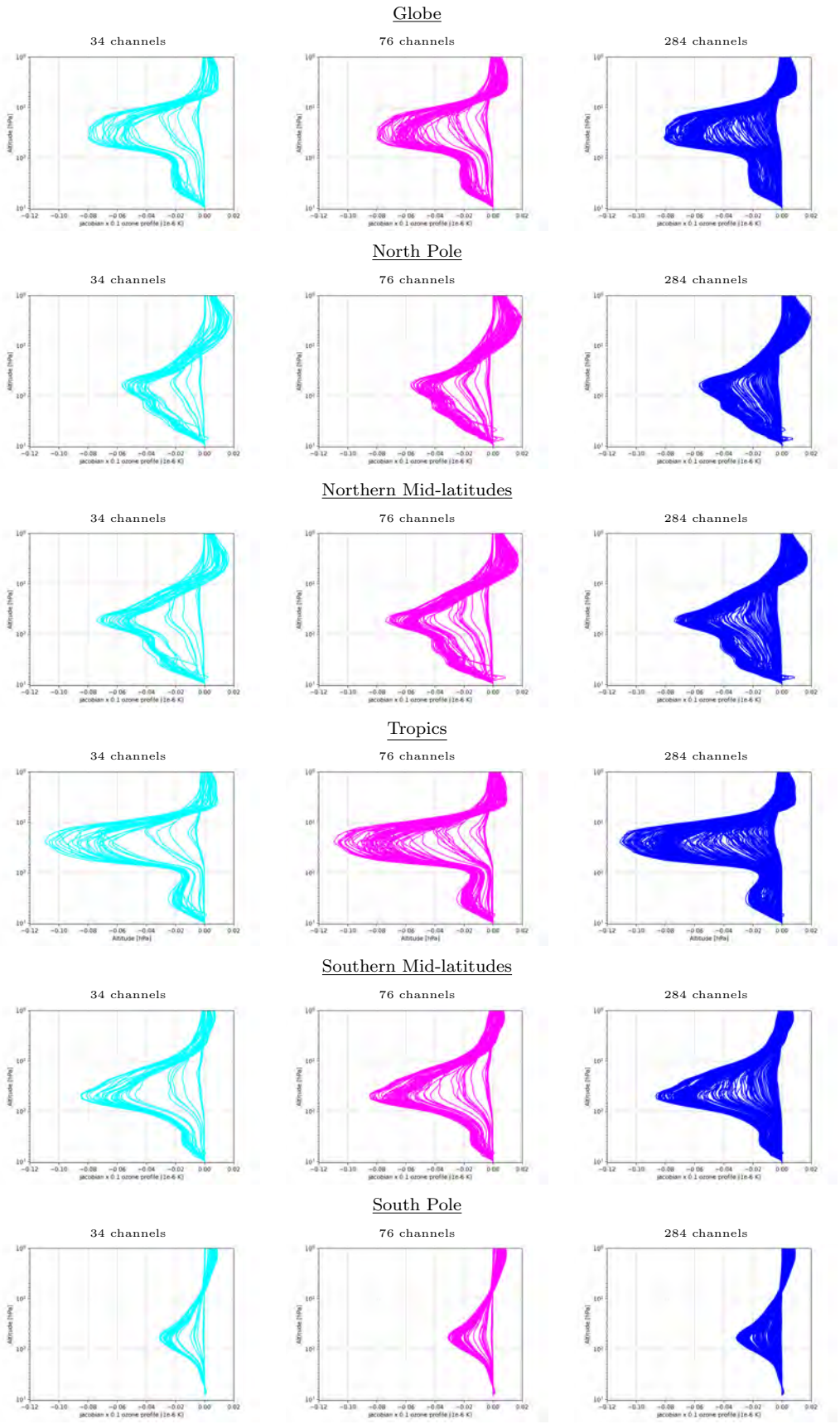


Figure B1. Ozone Jacobians times 10% of ozone profiles (K). Altitude is given in hPa. The Jacobians are displayed per latitude bands and channel groups.

Emanuele Emili and Dr. Mohammad El Aabaribaoune for the fruitful conversations, and Valentin Petiot and Dr. Mathieu Joly for their specially helpful technical support.

References

Barret, B., Le Flochmoen, E., Sauvage, B., Pavelin, E., Matricardi, M., & Cammas, J.-P. (2011). The detection of post-monsoon tropospheric ozone variability over south asia using iasi data. *Atmospheric Chemistry and Physics*, *11*(18), 9533–9548.

Blumstein, D., Chalon, G., Carlier, T., Buil, C., Hebert, P., Maciaszek, T., ... others (2004). Iasi instrument: Technical overview and measured performances. *Infrared Spaceborne Remote Sensing XII*, *5543*, 196–207.

Bormann, N., Lawrence, H., & Farnan, J. (2019). *Global observing system experiments in the ecmwf assimilation system*. European Centre for Medium Range Weather Forecasts.

Bouyssel, F., Berre, L., Bénichou, H., Chambon, P., Girardot, N., Guidard, V., ... others (2022). The 2020 global operational nwp data assimilation system at météo-france. In *Data assimilation for atmospheric, oceanic and hydrologic applications (vol. iv)* (pp. 645–664). Springer.

Boxe, C., Worden, J., Bowman, K., Kulawik, S., Neu, J., Ford, W., ... others (2010). Validation of northern latitude tropospheric emission spectrometer stare ozone profiles with arc-ions sondes during arctas: sensitivity, bias and error analysis. *Atmospheric Chemistry and Physics*, *10*(20), 9901–9914.

Clerbaux, C., Boynard, A., Clarisse, L., George, M., Hadji-Lazaro, J., Herbin, H., ... others (2009). Monitoring of atmospheric composition using the thermal infrared iasi/metop sounder. *Atmospheric Chemistry and Physics*, *9*(16), 6041–6054.

Courtier, P., Freydier, C., Geleyn, J., Rabier, F., & Rochas, M. (1991). The arpege project at météo-france. In *Proceedings of the 1991 ecmwf seminar* (pp. 193–231).

Cussac, M., Marécal, V., Thouret, V., Josse, B., & Sauvage, B. (2020). The impact of biomass burning on upper tropospheric carbon monoxide: a study using mocage global model and iagos airborne data. *Atmospheric Chemistry and Physics*, *20*(15), 9393–9417.

Descheemaecker, M., Plu, M., Marécal, V., Claeyman, M., Olivier, F., Aoun, Y., ... others (2019). Monitoring aerosols over europe: an assessment of the potential benefit of assimilating the vis04 measurements from the future mtg/fci geostationary imager. *Atmospheric Measurement Techniques*, *12*(2), 1251–1275.

Desroziers, G., Berre, L., Chapnik, B., & Poli, P. (2005). Diagnosis of observation, background and analysis-error statistics in observation space. *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography*, *131*(613), 3385–3396.

Dufour, G., Eremenko, M., Griesfeller, A., Barret, B., LeFlochmoën, E., Clerbaux, C., ... Hurtmans, D. (2012). Validation of three different scientific ozone products retrieved from iasi spectra using ozonesondes. *Atmospheric Measurement Techniques*, *5*(3), 611–630.

Eckermann, S. (2009). Hybrid σ–p coordinate choices for a global model. *Monthly Weather Review*, *137*(1), 224–245.

El Aabaribaoune, M., Emili, E., & Guidard, V. (2021). Estimation of the error covariance matrix for iasi radiances and its impact on the assimilation of ozone in a chemistry transport model. *Atmospheric Measurement Techniques*, *14*(4), 2841–2856.

El Amraoui, L., Attié, J.-L., Semane, N., Claeyman, M., Peuch, V.-H., Warner, J., ... others (2010). Midlatitude stratosphere–troposphere exchange as diagnosed by mls o 3 and mopitt co assimilated fields. *Atmospheric Chemistry and Physics*, *10*(5), 2175–2194.

El Amraoui, L., Plu, M., Guidard, V., Cornut, F., & Bacles, M. (2022). A pre-operational system based on the assimilation of modis aerosol optical depth in the mocage chemical transport model. *Remote Sensing*, *14*(8), 1949.

Emili, E., Barret, B., Le Flochmoën, E., & Cariolle, D. (2019). Comparison between the assimilation of iasi level 2 ozone retrievals and level 1 radiances in a chemical transport

model. *Atmospheric Measurement Techniques*, *12*(7), 3963–3984.

Emili, E., Barret, B., Massart, S., Le Flochmoen, E., Piacentini, A., El Amraoui, L., ... Cariolle, D. (2014). Combined assimilation of iasi and mls observations to constrain tropospheric and stratospheric ozone in a global chemical transport model. *Atmospheric Chemistry and Physics*, *14*(1), 177–198.

Flemming, J., Benedetti, A., Inness, A., Engelen, R. J., Jones, L., Huijnen, V., ... others (2017). The cams interim reanalysis of carbon monoxide, ozone and aerosol for 2003–2015. *Atmospheric Chemistry and Physics*, *17*(3), 1945–1983.

Geer, A., Lahoz, W., Bekki, S., Bormann, N., Errera, Q., Eskes, H., ... others (2006). The asset intercomparison of ozone analyses: method and first results. *Atmospheric Chemistry and Physics*, *6*(12), 5445–5474.

Guth, J., Josse, B., Marécal, V., Joly, M., & Hamer, P. (2016). First implementation of secondary inorganic aerosols in the mocage version r2. 15.0 chemistry transport model. *Geoscientific Model Development*, *9*(1), 137–160.

Hilton, F., Armante, R., August, T., Barnet, C., Bouchard, A., Camy-Peyret, C., ... others (2012). Hyperspectral earth observation from iasi: Five years of accomplishments. *bulletin of the american meteorological Society*, *93*(3), 347–370.

Huijnen, V., Pozzer, A., Arteta, J., Brasseur, G., Bouarar, I., Chabrillat, S., ... others (2019). Quantifying uncertainties due to chemistry modelling–evaluation of tropospheric composition simulations in the cams model (cycle 43r1). *Geoscientific Model Development*, *12*(4), 1725–1752.

Inness, A., Baier, F., Benedetti, A., Bouarar, I., Chabrillat, S., Clark, H., ... others (2013). The macc reanalysis: an 8 yr data set of atmospheric composition. *Atmospheric chemistry and physics*, *13*(8), 4073–4109.

Josse, B., Simon, P., & Peuch, V.-H. (2004). Radon global simulations with the multiscale chemistry and transport model mocage. *Tellus B: Chemical and Physical Meteorology*, *56*(4), 339–356.

Lacressonnière, G., Peuch, V.-H., Arteta, J., Josse, B., Joly, M., Marécal, V., ... Watson, L. (2012). How realistic are air quality hindcasts driven by forcings from climate model simulations? *Geoscientific Model Development*, *5*(6), 1565–1587.

Lahoz, W., Geer, A., Bekki, S., Bormann, N., Ceccherini, S., Elbern, H., ... others (2007). The assimilation of envisat data (asset) project. *Atmospheric Chemistry and Physics*, *7*(7), 1773–1796.

Lamarque, J.-F., Shindell, D. T., Josse, B., Young, P., Cionni, I., Eyring, V., ... others (2013). The atmospheric chemistry and climate model intercomparison project (accmip): Overview and description of models, simulations and climate diagnostics. *Geoscientific Model Development*, *6*(1), 179–206.

Lamotte, C., Guth, J., Marécal, V., Cussac, M., Hamer, P. D., Theys, N., & Schneider, P. (2021). Modeling study of the impact of so 2 volcanic passive emissions on the tropospheric sulfur budget. *Atmospheric Chemistry and Physics*, *21*(14), 11379–11404.

Lefevre, F., Brasseur, G., Folkins, I., Smith, A., & Simon, P. (1994). Chemistry of the 1991–1992 stratospheric winter: Three-dimensional model simulations. *Journal of Geophysical Research: Atmospheres*, *99*(D4), 8183–8195.

Levelt, P. F., Joiner, J., Tamminen, J., Veefkind, J. P., Bhartia, P. K., Stein Zweers, D. C., ... others (2018). The ozone monitoring instrument: overview of 14 years in space. *Atmospheric Chemistry and Physics*, *18*(8), 5699–5745.

Levelt, P. F., Van Den Oord, G. H., Dobber, M. R., Malkki, A., Visser, H., De Vries, J., ... Saari, H. (2006). The ozone monitoring instrument. *IEEE Transactions on geoscience and remote sensing*, *44*(5), 1093–1101.

Marécal, V., Peuch, V.-H., Andersson, C., Andersson, S., Arteta, J., Beekmann, M., ... others (2015). A regional air quality forecasting system over europe: the macc-ii daily ensemble production. *Geoscientific Model Development*, *8*(9), 2777–2813.

Massart, S., Clerbaux, C., Cariolle, D., Piacentini, A., Turquety, S., & Hadji-Lazaro, J. (2009). First steps towards the assimilation of iasi ozone data into the mocage-palm system. *Atmospheric Chemistry and Physics*, *9*(14), 5073–5091.

Morgenstern, O., Hegglin, M. I., Rozanov, E., O'Connor, F. M., Abraham, N. L., Akiyoshi, H., ... others (2017). Review of the global models used within phase 1 of the chemistry–climate model initiative (ccmi). *Geoscientific Model Development*, *10*(2), 639–671.

Orbe, C., Yang, H., Waugh, D. W., Zeng, G., Morgenstern, O., Kinnison, D. E., ... others (2018). Large-scale tropospheric transport in the chemistry–climate model initiative (ccmi) simulations. *Atmospheric Chemistry and Physics*, *18*(10), 7217–7235.

Parrish, D. F., & Derber, J. C. (1992). The national meteorological center's spectral statistical-interpolation analysis system. *Monthly Weather Review*, *120*(8), 1747–1763.

Peiro, H., Emili, E., Cariolle, D., Barret, B., & Le Flochmoën, E. (2018). Multi-year assimilation of iasi and mls ozone retrievals: variability of tropospheric ozone over the tropics in response to enso. *Atmospheric Chemistry and Physics*, *18*(9), 6939–6958.

Phulpin, T., Cayla, F., Chalon, G., Diebel, D., & Schlüssel, P. (2002). Iasi on board metop: Project status and scientific preparation. In *12th international tovs study conference, lorne, victoria, australia* (Vol. 26).

Rabier, F., Fourrié, N., Chafäi, D., & Prunet, P. (2002). Channel selection methods for infrared atmospheric sounding interferometer radiances. *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography*, *128*(581), 1011–1027.

Rodgers, C. D. (1998). Information content and optimisation of high spectral resolution remote measurements. *Advances in Space Research*, *21*(3), 361–367.

Rouil, L., Honore, C., Vautard, R., Beekmann, M., Bessagnet, B., Malherbe, L., ... others (2009). Prev'air: an operational forecasting and mapping system for air quality in europe. *Bulletin of the American Meteorological Society*, *90*(1), 73–84.

Saunders, R., Hocking, J., Turner, E., Rayer, P., Rundle, D., Brunel, P., ... others (2018). An update on the rttov fast radiative transfer model (currently at version 12). *Geoscientific Model Development*, *11*(7), 2717–2737.

Sič, B., El Amraoui, L., Marécal, V., Josse, B., Arteta, J., Guth, J., ... Hamer, P. (2015). Modelling of primary aerosols in the chemical transport model mocage: development and evaluation of aerosol physical parameterizations. *Geoscientific Model Development*, *8*(2), 381–408.

Siméoni, D., Singer, C., & Chalon, G. (1997). Infrared atmospheric sounding interferometer. *Acta Astronautica*, *40*(2-8), 113–118.

Stockwell, W. R., Kirchner, F., Kuhn, M., & Seefeld, S. (1997). A new mechanism for regional atmospheric chemistry modeling. *Journal of Geophysical Research: Atmospheres*, *102*(D22), 25847–25879.

Teyssèdre, H., Michou, M., Clark, H., Josse, B., Karcher, F., Olivié, D., ... others (2007). A new tropospheric and stratospheric chemistry and transport model mocage-climat for multi-year studies: evaluation of the present-day climatology and sensitivity to surface processes. *Atmospheric chemistry and physics*, *7*(22), 5815–5860.

Vittorioso, F., Guidard, V., & Fourrié, N. (2021). An infrared atmospheric sounding interferometer–new generation (iasi-ng) channel selection for numerical weather prediction. *Quarterly Journal of the Royal Meteorological Society*, *147*(739), 3297–3317.

Watson, L., Lacressonnière, G., Gauss, M., Engardt, M., Andersson, C., Josse, B., ... others (2016). Impact of emissions and+ 2 c climate change upon future ozone and nitrogen dioxide over europe. *Atmospheric environment*, *142*, 271–285.

Williams, J. E., Huijnen, V., Bouarar, I., Meziane, M., Schreurs, T., Pelletier, S., ... Flemming, J. (2022). Regional evaluation of the performance of the global cams chemical modeling system over the united states (ifs cycle 47r1). *Geoscientific Model Development*, *15*(12), 4657–4687.

Worden, H. M., Bowman, K. W., Worden, J. R., Eldering, A., & Beer, R. (2008). Satellite measurements of the clear-sky greenhouse effect from tropospheric ozone. *Nature Geoscience*, *1*(5), 305–308.

6.2.3 Nature Run vs Control Run

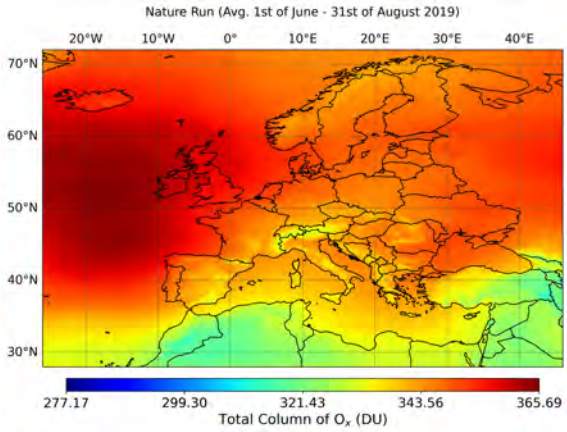
Once the Nature Run that best suited the purposes of this research was obtained, an evaluation against the CR was a mandatory step. As previously explained, indeed, it is of paramount importance that the CR is sufficiently different from the NR, even if consistent with it, in order to avoid the identical twin problem of over-optimistic results at the assimilation stage. This is particularly important for the present study since the same model was used for both the runs. An inter-comparison would also be needed to verify that the precautions taken to differentiate the different configurations were sufficient. To recall the differences in settings and inputs provided to MOCAGE for the two different frameworks CR and NR, please refer to Table 6.1.

6.2.3.1 Total Columns

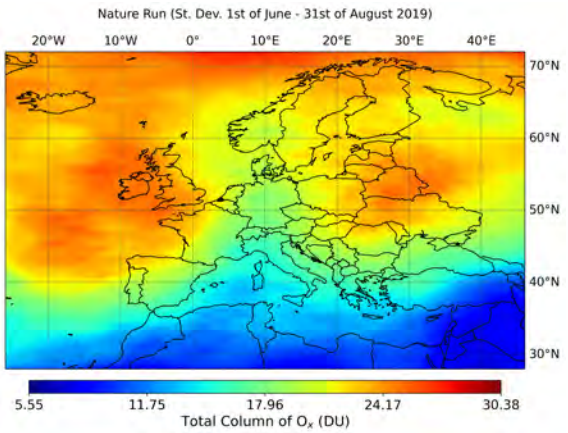
At first, O_x concentrations in total column are averaged over the 3-month period of study, from June to August, for both the NR (Panel 6.6a) and the CR (Panel 6.6c). By examining the two maps, it is possible to see that the runs are quite different from each other in terms of concentration magnitude: concentrations are actually smaller in the case of CR. They do, however, show a spatially distributed variability that is entirely consistent on both cases. When assessing the variations between the runs by the percentage computing relative difference (Panel 6.6e), using the NR as a reference, one can estimate that these variations are always negative. In other words, the CR shows concentrations of O_x are always lower than for the NR. More specifically, the deviation between the average concentration ranges between a minimum of 10.5% and maximum of 13.5%. The largest variations seem to take place between 40°N-20°E and 50°N-40°E. The errors associated with each run, shown in 6.6b and 6.6d, report the same type of spatial variability and similar intensities. The main differences are over the continental Europe (between 20 and 35°E, 50 and 60°N,) where the NR shows slightly higher standard deviations (of the order of a few DU) and at the border South and North of the domain, where, on the other hand, is the CR to present weakly stronger values. Even over the Atlantic Ocean, between 50 and 60°N the NR seems to report a finer structure of values compared to the CR.

At this stage, statistics[4] are also computed between the total columns of ozone obtained through the two runs (NR is used as the reality in this comparison). The results are displayed in Figure 6.7 on an average per month and for the whole period (June to August). An evaluation of the modified normalised mean bias (MNMB), shows that the bias tends to increase with time of execution, while the average values for the three months together are centred around -0.12 (CR weaker in intensity of O_x total columns than NR). The FGE values indicate that the error also increases with time, ranging from average values of around 0.10 in June to values around 0.15 in August.

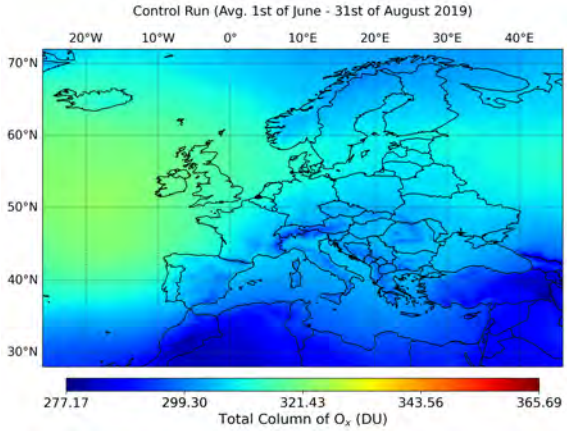
[4]Please notice that the characteristics of each indicator are listed more in detail in Appendix B.



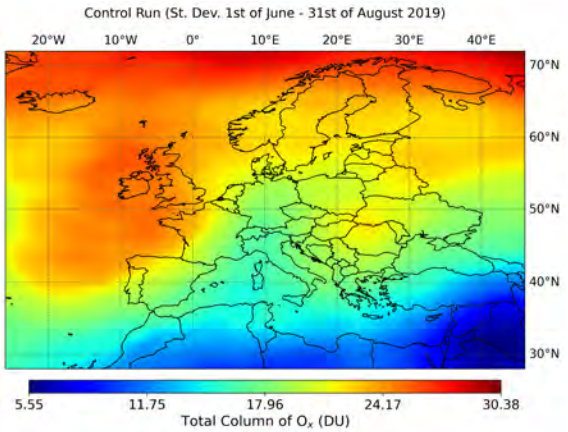
(a) $\overline{\text{NR}}$



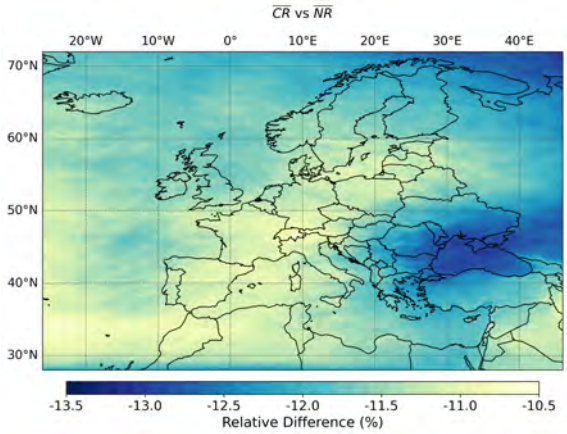
(b) σ(NR)



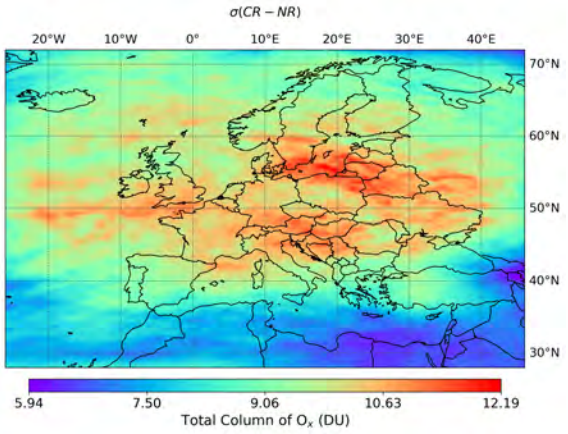
(c) $\overline{\text{CR}}$



(d) σ(CR)



(e) Relative Difference $\overline{\text{CR}}$ vs $\overline{\text{NR}}$



(f) $\sigma(\text{CR} - \text{NR})$

Figure 6.6: Average over the 3-month period of study (1st of June up to 31st of August, 2019) of the O_x Total Columns from the NR **(a)** and the CR **(c)**. Corresponding Standard Deviations are in **(b)** and **(d)**. Panel **(e)** shows the relative difference between the averages of the two runs $D_{rel\%} = [(\overline{CR} - \overline{NR})/(\overline{NR})] \cdot 100$, while **(f)** is the Standard Deviation of their differences.

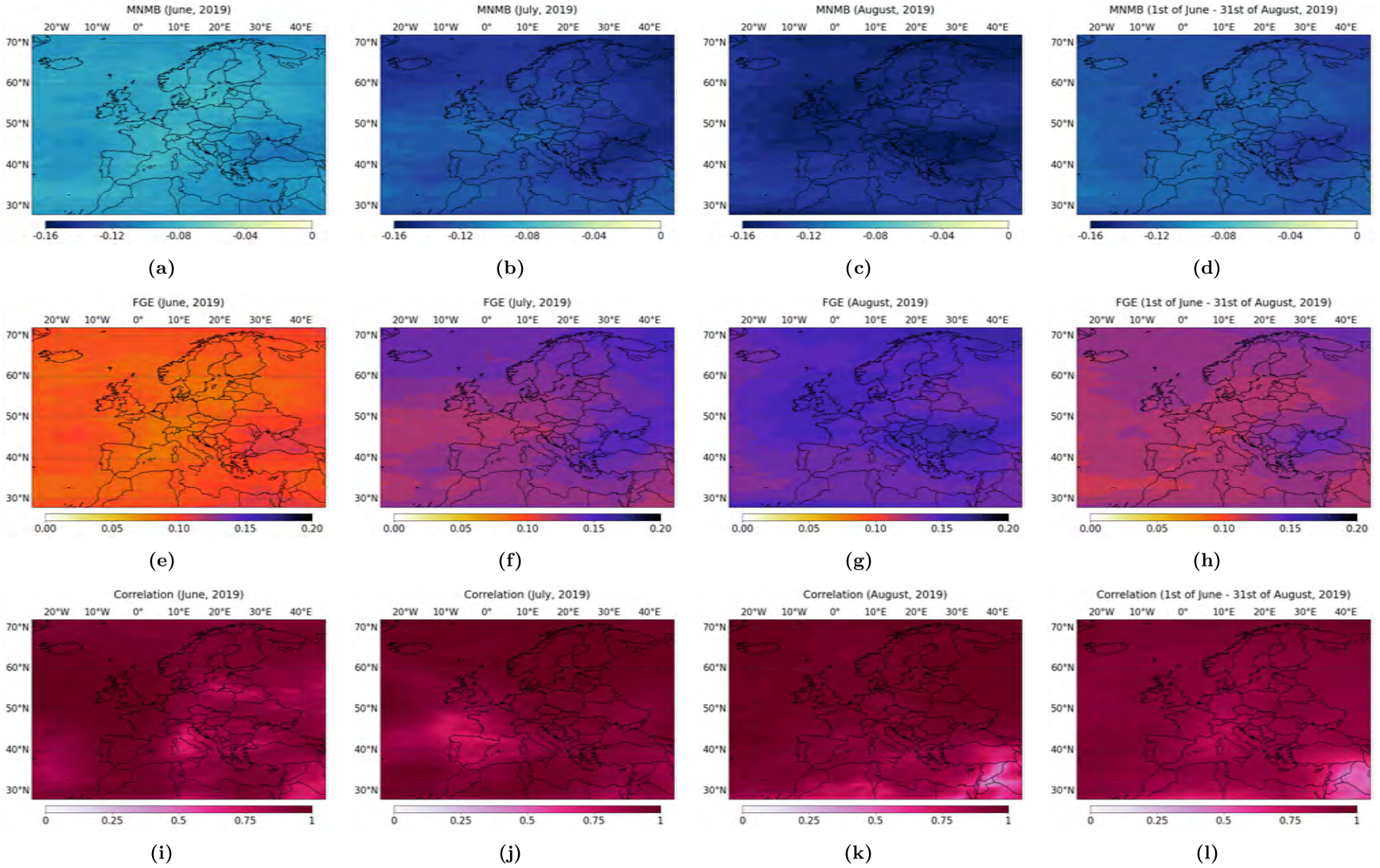
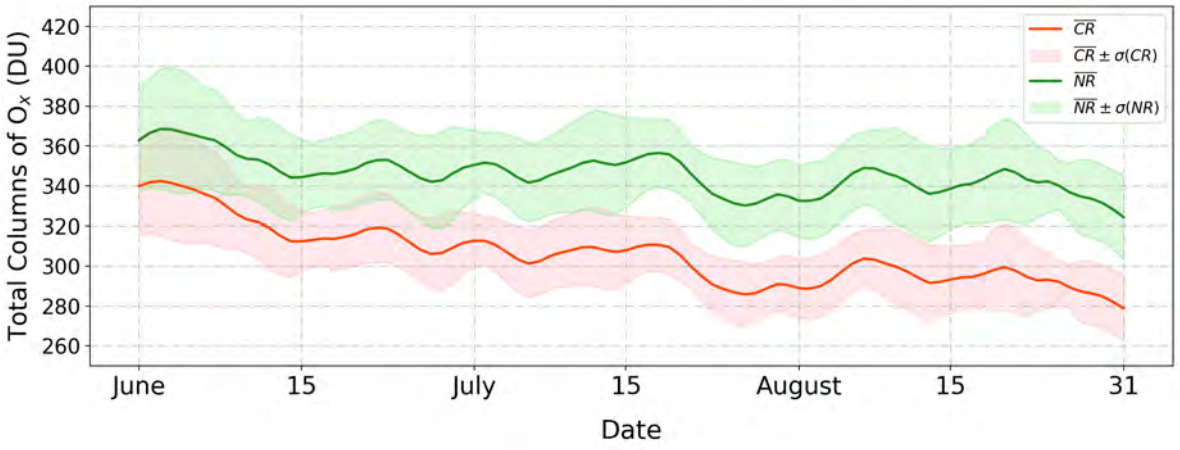


Figure 6.7: Statistics for the O_x Total Columns, computed using NR as reality. Modified Normalised Mean Bias (MNMB) averaged over the month of June **(a)**, July **(b)**, August 2019 **(c)** and over the 3 months **(d)**. Fractional Gross Error (FGE) for the corresponding periods are in Panels from **(e)** to **(h)**, while Pearson Correlation in **(i)** to **(l)**.



(a)

(b)

Figure 6.8: Time series of the averages per day of the O_x Total Columns obtained from the CR (red) and NR (green), plus and minus standard deviation **(a)**; relative differences between the averages of the two runs $[(\overline{CR} - \overline{NR})/(\overline{NR})] \cdot 100$ is shown in **(b)**.

CR and NR, on the other hand, result to be very correlated, with a Pearson Correlation Coefficient that tends to 1 as the months progress. This means they are spatial consistent.

Figure 6.8a shows the total column values averaged for each day in the period of study (from June to August) and over the entire domain. The time series of the averages are shown with the respective standard deviation added and subtracted. The relative differences between the averages are also shown in 6.8b. The two runs present the same daily variability, but, again, the CR values are weaker. The curves tend to assume a stable divergence after July. This could be due to how the global run, from which the regional NR was extracted, was constructed. In fact, from July 2019 data of three IASIs, instead of two, became available and were thus assimilated. The variation between the total columns from the two runs ranges always between 7% and 15%.

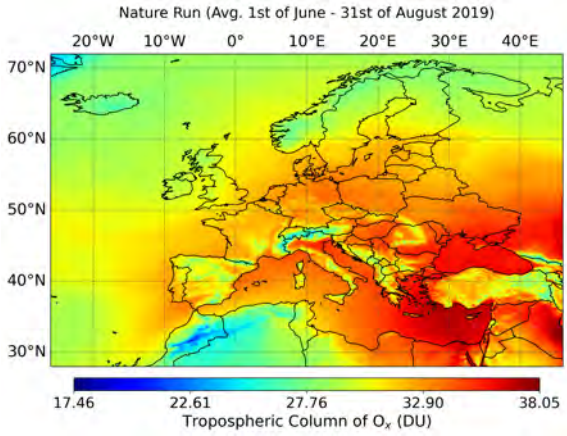
6.2.3.2 Tropospheric Columns

As explained in the opening Chapter, typical ozone concentrations in the troposphere are lower than those found in stratosphere. When performing a study on the total columns, thus, the contribution of stratospheric ozone will be the one that mainly arises. In order to better assess what happens in troposphere, then, a study limited to the tropospheric layer must be led. After empirically assessing the average position of the tropopause, and cross-comparing it with the vertical levels provided by MOCAGE, it has been decided to define in this manuscript as *Tropospheric Column* the one running from the surface up to about 300 hPa, which corresponds to MOCAGE vertical levels ranging from 60 to 40 (for the equivalence level - pressure see Appendix A).

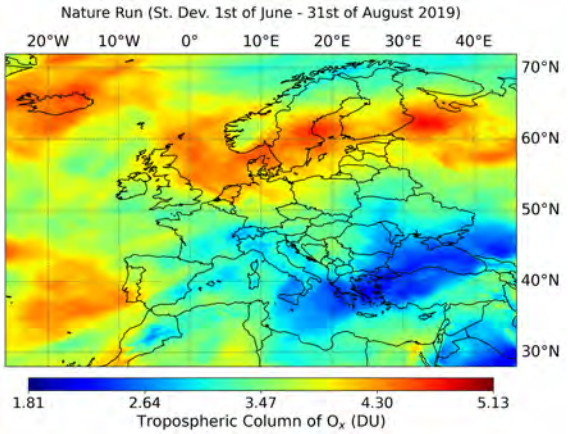
Panels 6.9a and 6.9c report the tropospheric columns thus obtained, averaged over the 3-month time-period from June to August, for the NR and the CR respectively. In comparison to what had been observed for the total columns (Panels 6.6a and 6.6c) the values are, as expected, lower in magnitude (they range from about 17 to 38 DU). When quantifying the differences occurring between the two cases, it is observed that these reach percentage values that are even higher than in the case of the total columns. The maximum variations occur in the South-East quadrant of the domain, with peaks of -20% (CR values smaller than NR) above the Black Sea.

The associated errors (Panels 6.9b and 6.9d) are, again, similar in spatial distribution, but different in intensity. In the case of CR, more extreme values are reached. The maxima are stronger than those corresponding to the NR and are located to the west of the domain, at latitudes between 10°W and the edge. The minima, on the other hand, lie in the East of the domain, between the lower edge and 50°N.

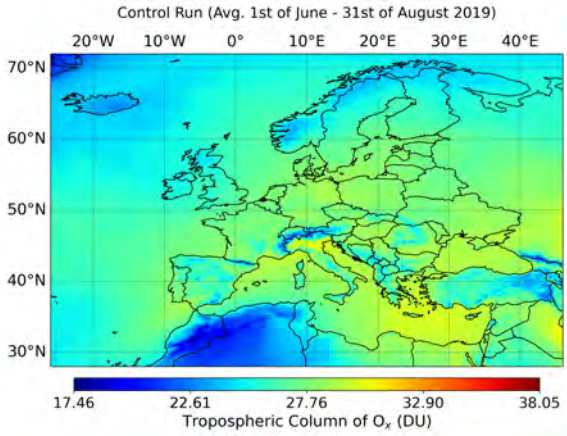
The standard deviation of the bias between the two scenarios, shown in 6.9f, reports areas of minima above Alps, Morocco and in the Middle-East area. Maxima, on the other hand, are over Netherlands and the nearby seaside area, above the western portion of the Black Sea, southward on the Mediterranean and Egypt, in the continental area around 45°N and 26°E. Some of these values may be led by stratospheric intrusions, which are known to take place in the Mediterranean basin [Lelieveld and Dentener (2000); Lelieveld et al. (2002)].



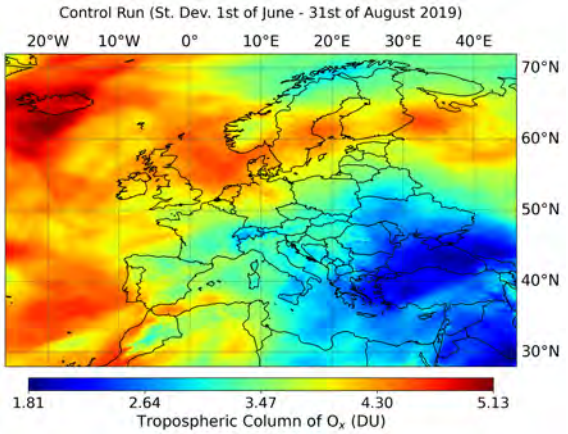
(a) $\overline{\text{NR}}$



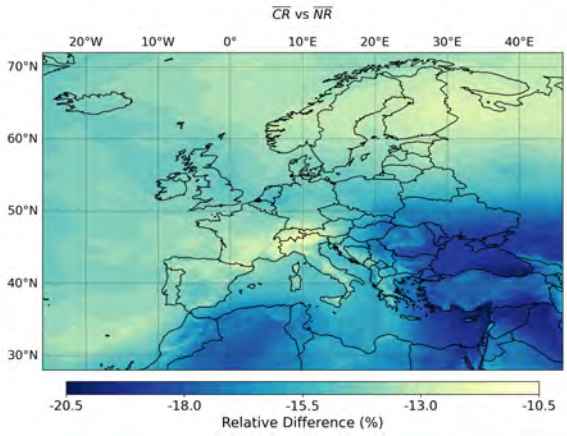
(b) σ(NR)



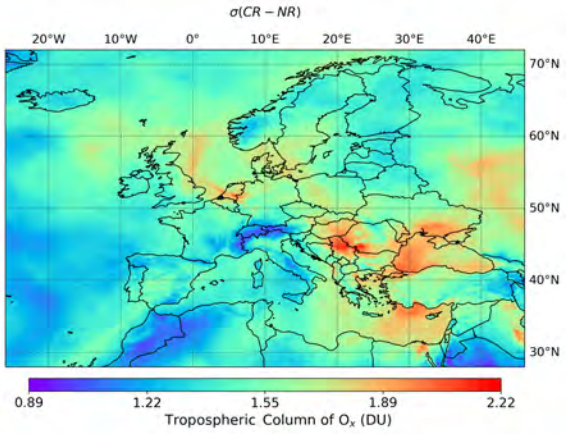
(c) $\overline{\text{CR}}$



(d) σ(CR)



(e) Relative Difference $\overline{\text{CR}}$ vs $\overline{\text{NR}}$



(f) σ(CR − NR)

Figure 6.9: Average over the 3-month period of study (1st of June up to 31st of August, 2019) of the O_x Tropospheric Columns from the NR **(a)** and the CR **(c)**. Corresponding Standard Deviations are in **(b)** and **(d)**. Panel **(e)** shows the relative difference between the averages of the two runs, while **(f)** is the Standard Deviation of their differences.

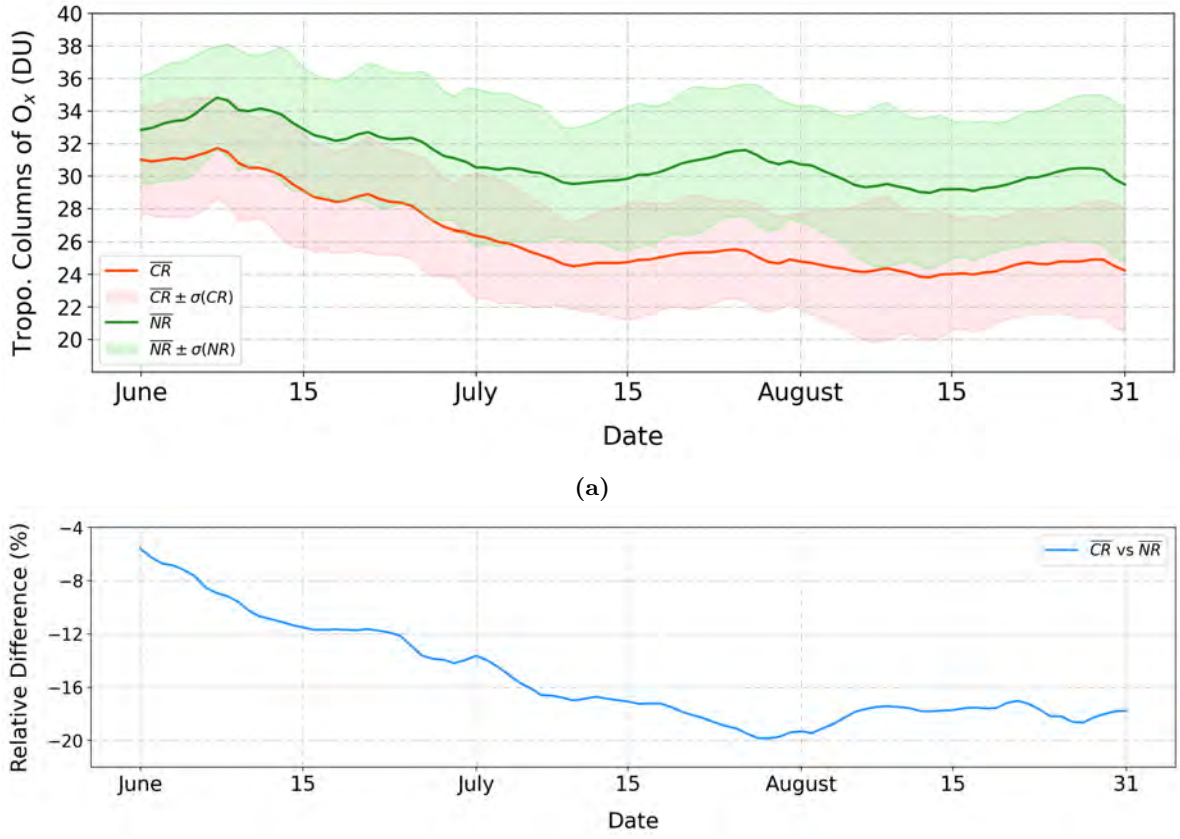


Figure 6.10: Time series of the averages, per day and over the MACC01 domain, of the O_x Tropospheric Columns obtained from the CR (red) and NR (green), plus and minus standard deviation **(a)**; percentage relative differences between the averages of the two runs are in **(b)**.

As already done for the total columns, the domain and day-averaged values of the tropospheric O_x columns were calculated (Figure 6.10). The time series of the averages, with associated standard deviations, is reported in Panel 6.10a. The relative difference between the averages is, instead, in Panel 6.10b. Compared to Figure 6.8, the magnitudes of the amount of gas are, of course, lower. But also, the two runs show a larger relative divergence compared to what was found for the total columns. Again, the NR estimates larger quantities of O_x than the CR does. Compared to the case of the total columns, however, where they differed by as much as 15%, in this case the relative differences reach peaks of 20% at the end of July, only to rise again by a few points in August.

6.2.3.3 Vertical Variations

So far, only values on atmospheric columns have been examined. Figure 6.11 gives an idea of what happens to the vertical distribution of ozone. The cross sections of the O_x mole fraction along the latitudes are displayed, averaged over the whole period of study and the longitudes. The cases of NR, CR and their relative percentage differences are shown in 6.11a, 6.11c and 6.11e, respectively. What can be immediately observed is that the largest portion of ozone is found between level 20 and 10 for both runs. However, looking more closely we find that the CR values are lower and the maxima are more shifted towards the higher atmospheric levels. This is confirmed by the relative difference showing that the maximum divergence of the runs occurs between levels 35 and 20, where values 35% stronger are found for the NR compared to the CR. Looking at the equivalence of the vertical levels with pressure values (as shown in Appendix A), it is found that this layer ranges approximately between 25 and 180 hPa, which is approximately where the ozonosphere is found. If observing the results obtained on a scale that fits the absolute maxima, it looks like the variation does not affect the lower-most layers of the atmosphere. However, when adjusting the colour scale to the values found for the levels going from the surface up the 45th (Figures 6.11b, 6.11d and 6.11f), the concentration structures and the variations taking place between the two runs can be easily recognised. Although the ozone concentrations are spatially consistent, again the values are stronger for the NR. The largest differences (up to 25%) occur in the latitude range between 30° and 37° and between level 60 to 57. The lowest variations (around 10%), on the other hand, are found from 60° and 70°.

Another inter-comparison between CR and NR was also performed by averaging, not only over the 3-month period of study and the longitudes, but all over the regional domain (both latitudes and longitudes). The averages of the O_x concentrations obtained for each of the 60 MOCAGE levels, plus and minus the respective standard deviations, are shown in Figure 6.12a. The relative difference are also evaluated between the two runs in percentage terms and are displayed in Figures 6.12b The two scenarios seem to mostly diverge in the lower troposphere and in stratosphere between levels from 40 to 25, i.e. approximately between 320 at 50 hPa (Appendix A). More in detail, variations of the order of 20% for averages occur in the lower layers. The maxima, on the other hand, are found around the level 30 ($\sim$ 90 hPa), where the NR averaged values appear to be stronger of 28% than the CR ones. On the other hand, NR and CR are very comparable for levels 20 and above.

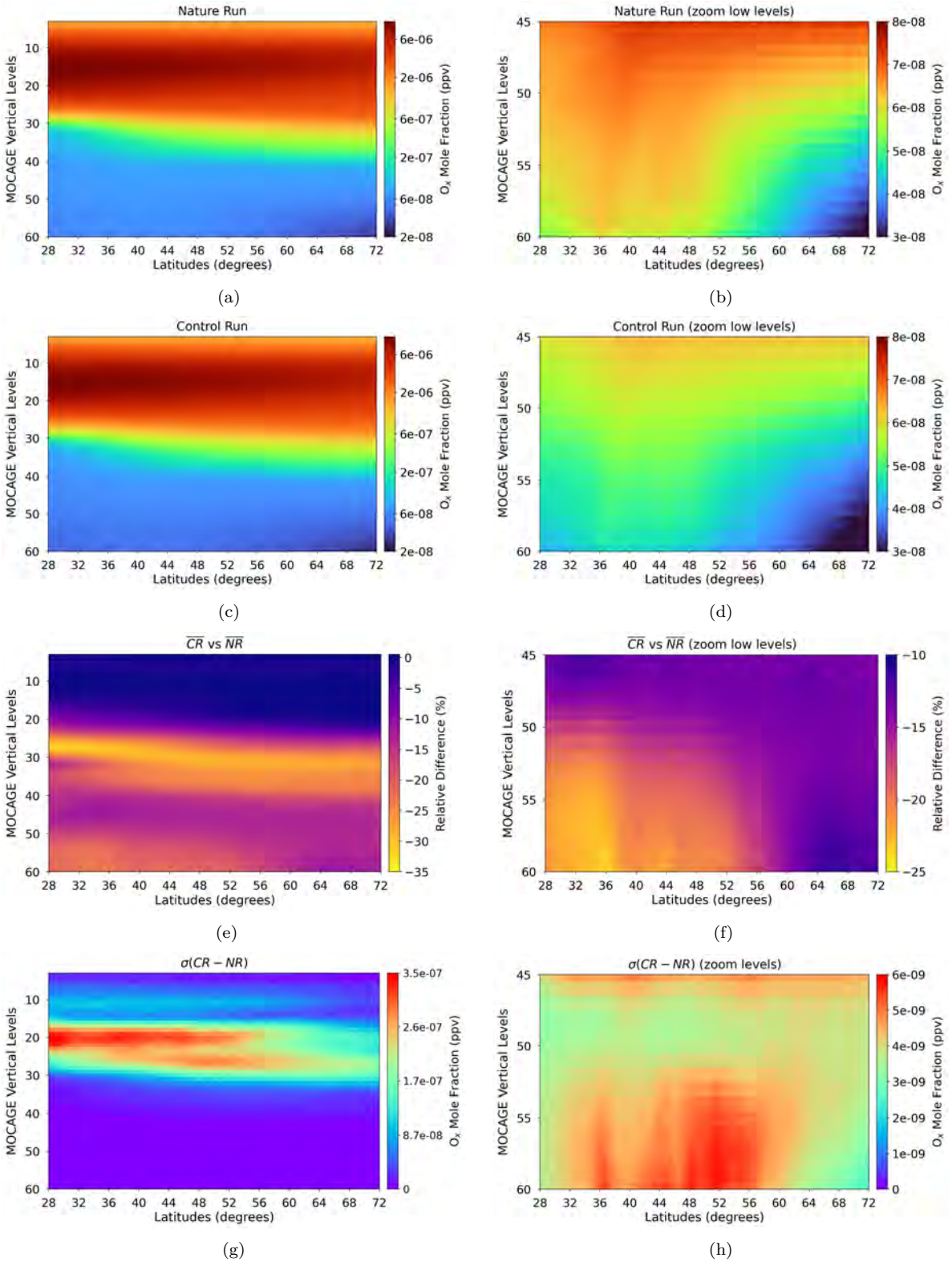
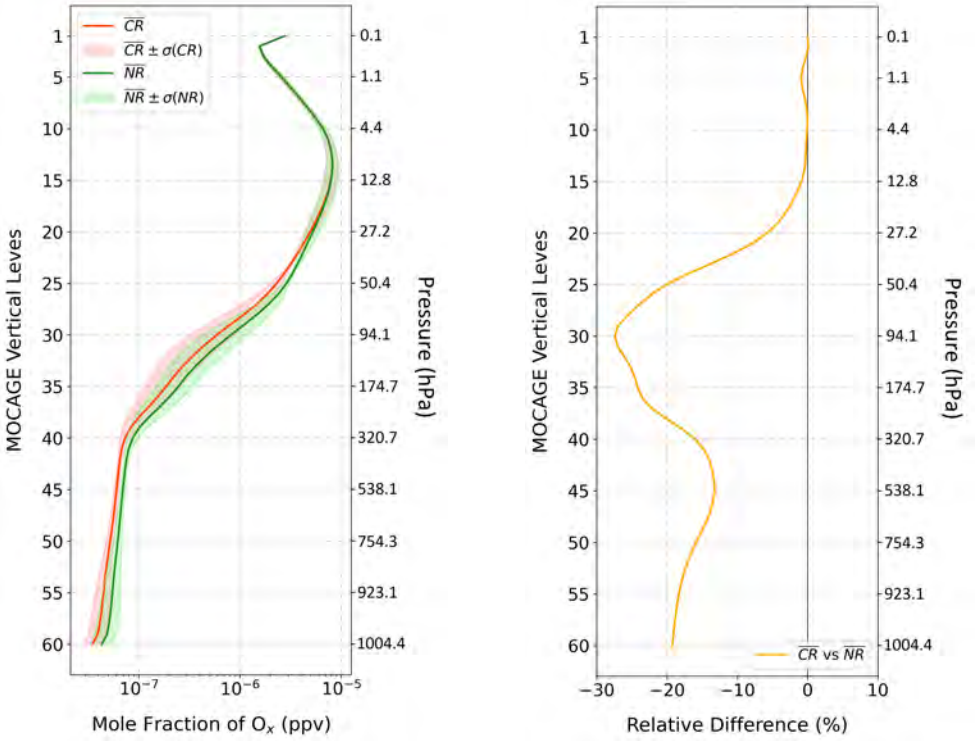


Figure 6.11: Average of the O_x Mole Fraction, for each MOCAGE vertical level, computed over the longitudes of the full domain and the time period (1st of June - 31st of August, 2019). The results are shown over the entire vertical column and in a zoom of the lower levels (from 60 to 45), for the NR in **(a)** and **(b)** and the CR in **(c)** and **(d)**. The relative differences in percentage are shown for both scales in **(e)** and **(f)**. The standard deviations of the differences are, instead, in Panels **(g)** and **(h)**. For equivalence between model and pressure levels see Figure 6.12 or refer to Appendix A.



(a) **(b)**

Figure 6.12: Average over the 3-month period of study (1st of June up to 31st of August, 2019) and, for each of the 60 MOCAGE model levels, on the MACC01 domain for both CR and NR, plus and minus their standard deviations **(a)**; relative difference between the averages is also shown in **(b)**

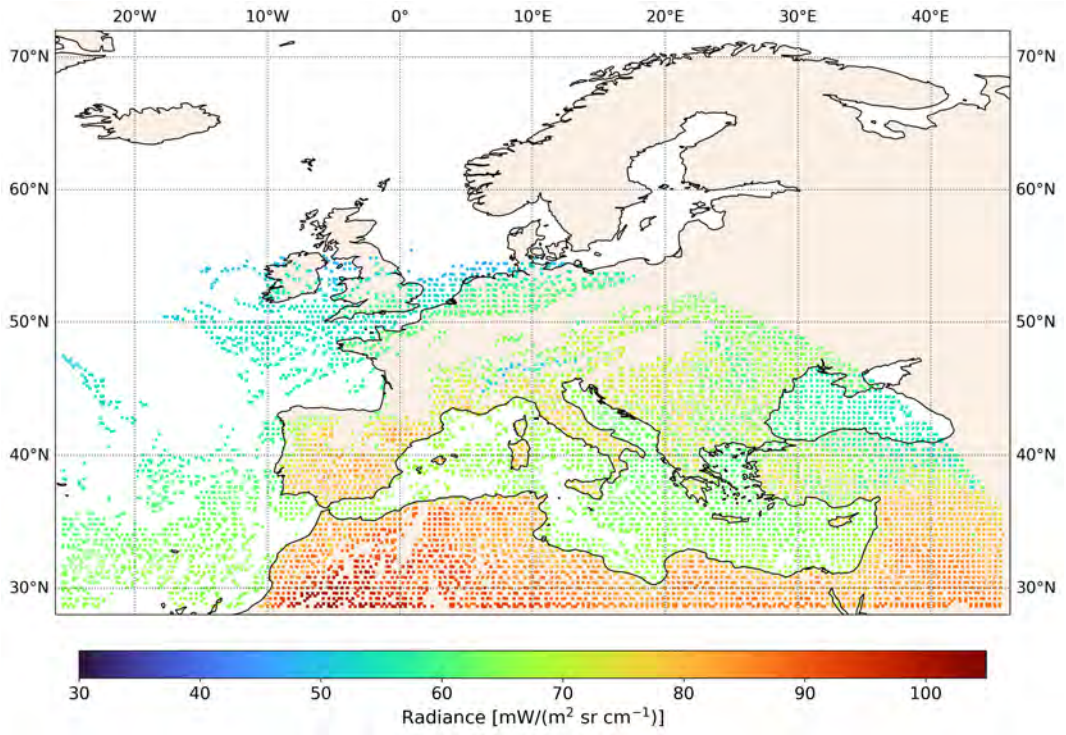


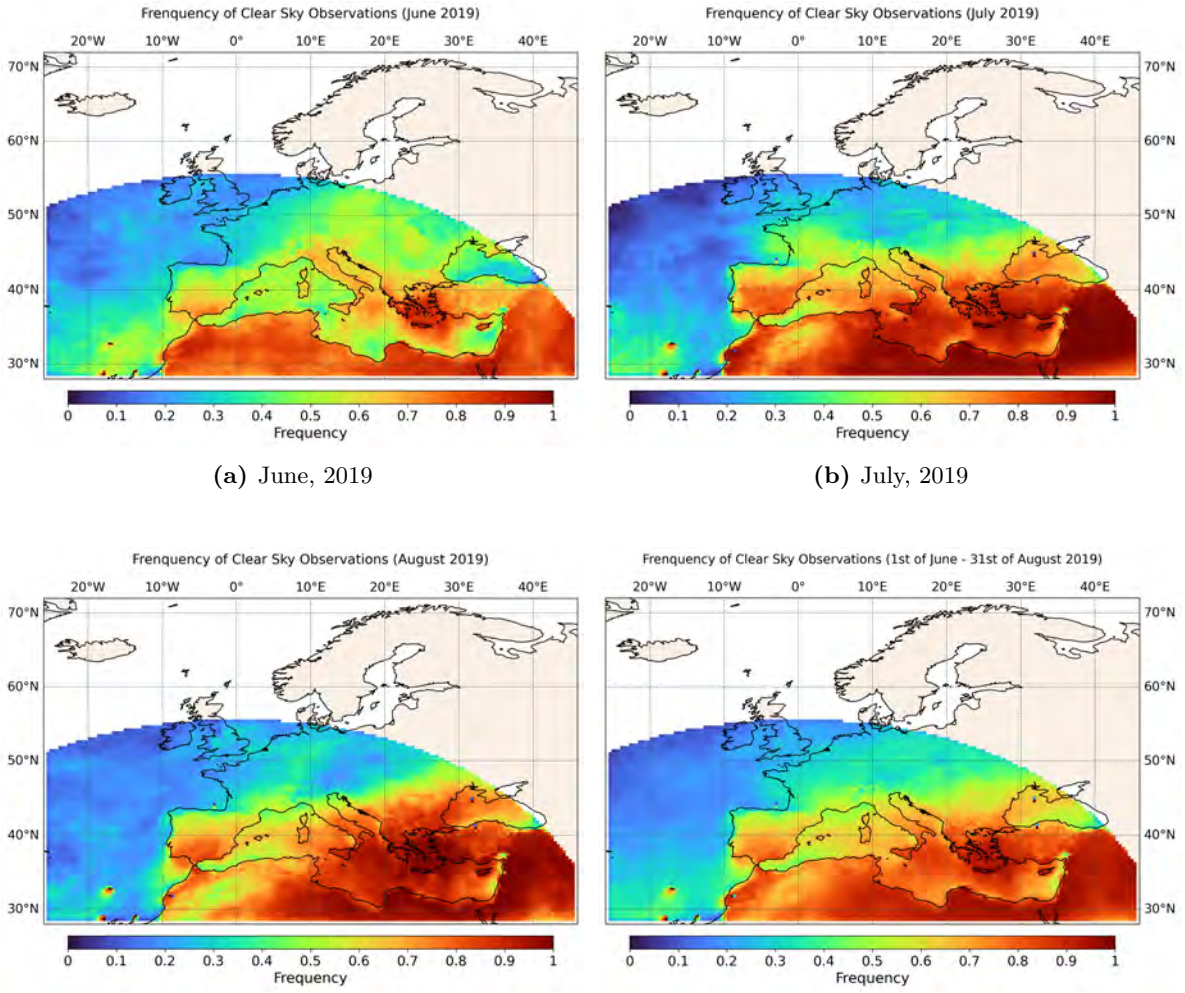
Figure 6.13: Simulated radiances for the IRS channel number 550 (1010.810 cm^{-1}) for the 1st of July, 2019 at 12 UTC.

6.3 Simulation of the Observations

As previously introduced, the IRS and IRS*2 observations have been obtained from the NR on the regional MACC01 domain, using RTTOV v12, exploiting the specific coefficients designed for each version of the instrument.

For IRS the band containing 195 contiguous channels (between 982.464 and 1099.467 cm^{-1}), pre-selected through the sensitivity studies as in 5.4.3.3, was simulated. Since, at present, only clear-sky observations are assimilated in MOCAGE, only clear-sky pixels are simulated for this work. To determine whether a pixel is clear or not, the model cloud parameters are used, which come, therefore, from the meteorological forcing exploited (i.e. ARPEGE OPER). Given the density of the observations that an instrument of the IRS's calibre will be able to provide, it becomes unlikely, from computational point of view, to be able to simulate and then assimilate such dense observations in the time available for this project. A thinning of the pixels to be simulated was therefore carried out. One pixel per 0.4° box was therefore simulated in each scenario. To the perfect observations thus obtained, the IRS instrumental noise was then added to produce the ultimate synthetic observations.

With regard to IRS*2, the corresponding band containing 361 contiguous channels (982.343 to 1099.379 cm^{-1}) was simulated. The same setting as for IRS was applied for both cloud cover and thinning. The IRS*2 observations will therefore be spatially located exactly where the IRS ones are. The noise specific for this version of the instrument (as reported in 5.4.2)



(a) June, 2019

(b) July, 2019

(c) August, 2019

(d) June to August, 2019

Figure 6.14: Frequency of the simulated clear-sky observations (the same for IRS and IRS*2) over the month of June **(a)**, July **(b)** and August 2019 **(c)**. The frequency all over the three months is shown in **(d)**.

was finally added to produce the IRS*2 synthetic observations.

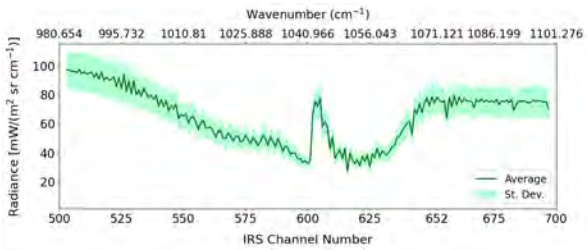
Figure 6.13 shows the distribution of the simulated pixels for a random IRS channel (channel number: 550, wavenumber: 1010.810 cm^{-1}) of simulated observations at 12 UTC on a day in the middle of the study period (1st of July, 2019). The map, which serves for illustrative purposes only, shows how the observations could potentially be distributed taking into account thinning and cloud cover, and nicely illustrates the disk that will be covered through the instrument acquisitions.

Figures 6.14, on the other hand, report the frequency of the simulated clear-sky observations (again, the same for both instrument versions) for the months of June (6.14a), July (6.14b) and August (6.14c), and all over the 3-month period (6.14d). It is evident that observations are more frequent over land and in the South-East portion of the instrument disk. This frequency is even stronger for the months of July and August, most likely due to the

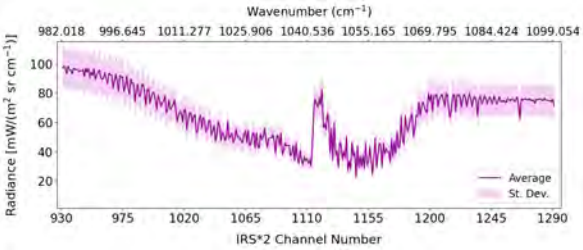
less dense cloud cover over the affected areas during these two summer months. The lowest density occurs over the Atlantic Ocean in July.

Figures 6.15a and 6.15b show the simulated radiances for both IRS and IRS*2 respectively and in the ranges of wavelengths, similar to each other, but specific for each instrument version. Only the average over one day along the whole simulated period is shown as an example (1st of July, 2019). The average is computed over the 24 hours, together with the standard deviation. IRS*2 spectrum shows a stronger frequency of the absorption lines, due of course to the thinner spectral sampling with respect to IRS. Panels 6.15c and 6.15d reports the same exact statistics, but computed over observations perturbed with the noise specific for each instrument version. The spectra corresponding to each instrument present very similar shapes, since the noise scale is quite small compared to the scale of the simulated radiances. If plotting a differences of both averages and standard deviations between perfect and noised observations (Panels 6.15e and 6.15f), it is possible to observe the differences actually occurring. More specifically, the difference in the standard deviations, it totally superimposable to the instrument noise (compare to Figure 5.9). The difference between the averaged spectra, on the other hand, even if very small, is not zero, as shown in Figures 6.15g and 6.15h.

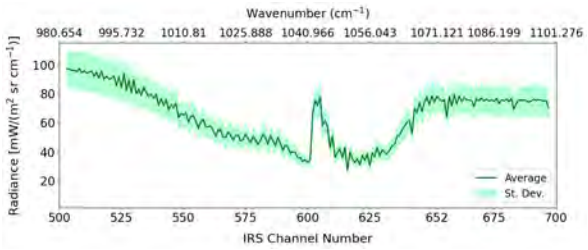
Finally, the simulated ozone Jacobians for both instruments, averaged over the regional domain, are shown in Figures 6.16. Both simple Jacobians and Jacobians times the 10% of the ozone profile itself are illustrated. For both instrument versions the highest values are located in the band between 1000 and 1200 cm^{-1}, bounded in level 50 and 35. When normalising against the ozon profile, on the other hand, this sensitivity spreads on a wider range of vertical levels and the highest values can be found around level 25 and 20. Negative values of sensitivity are also found between level 10 and 5, i.e. in stratosphere.



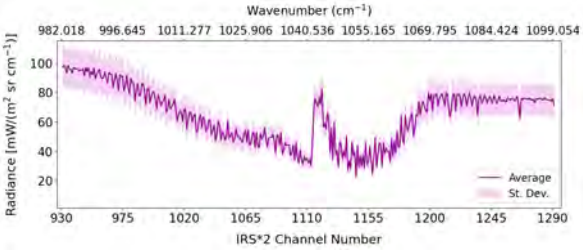
(a) IRS perfect obs.



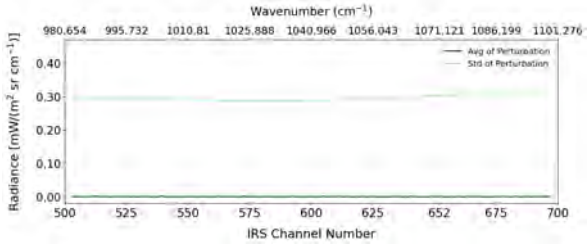
(b) IRS*2 perfect obs.



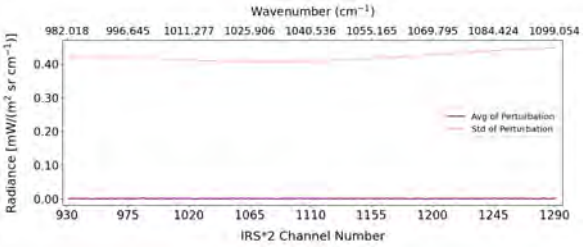
(c) IRS noised obs.



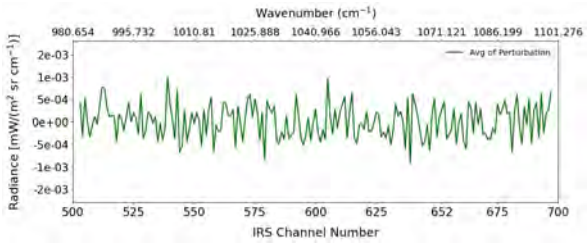
(d) IRS*2 noised obs.



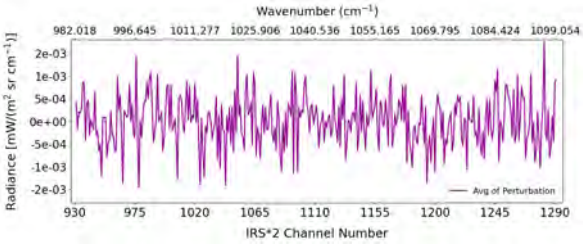
(e) Perfect vs noised IRS obs.



(f) Perfect vs noised IRS*2 obs.

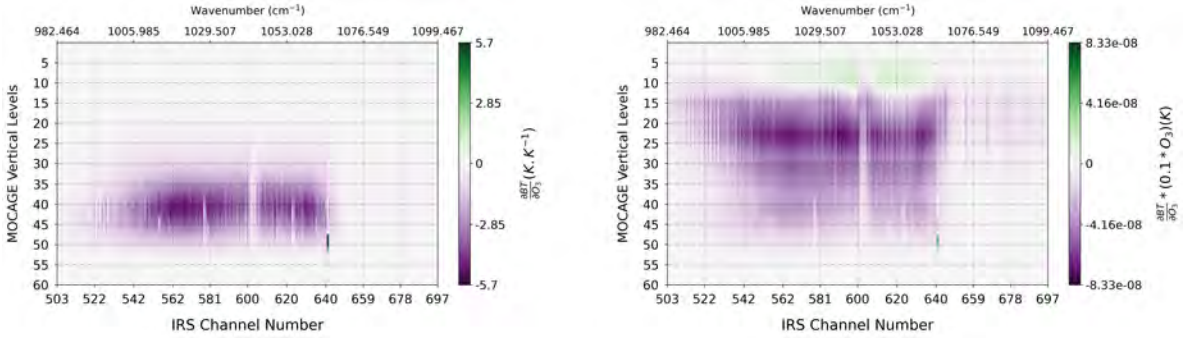


(g) Perfect vs noised IRS obs. (zoom)

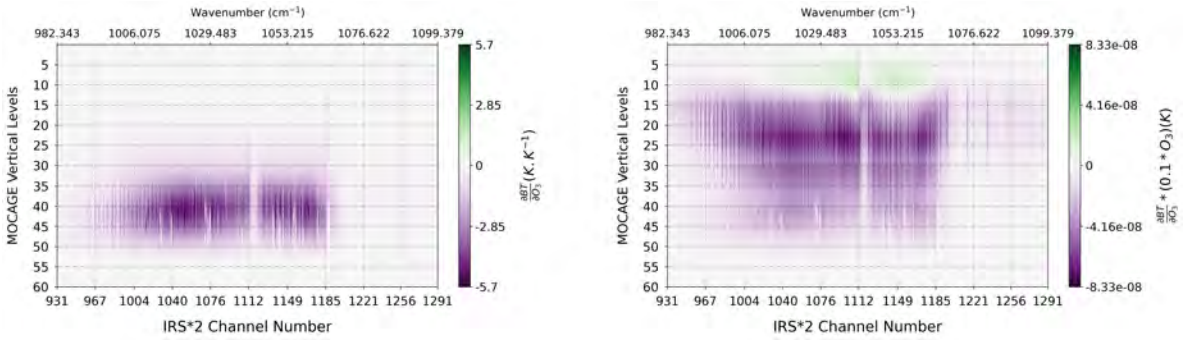


(h) Perfect vs noised IRS*2 obs. (zoom)

Figure 6.15: Simulated spectra of a day in the period of study (1st of July, 2019), averaged over the hours of the day, plus and minus standard deviation. The Panels show perfect observations for IRS **(a)** and IRS*2 **(b)**, and spectra with noise for IRS **(c)** and IRS*2 **(d)**. The differences between standard deviations before and after noise is added and between the averaged spectra are in **(e)** for IRS and in **(f)** for IRS*2. A zoom on the perturbation of the spectra is also shown in **(g)** and **(h)**.



(a) Ozone Jacobians for IRS



(b) Ozone Jacobians for IRS*2

Figure 6.16: Ozone Jacobians for the 195 channels simulated for IRS **(a)**, both simple (left) and normalised to the 10% of the ozone profile itself (right). The same is in **(b)** for the 361 channels simulated for IRS*2. The values are averaged over the regional domain.

6.4 Highlights of this Chapter

- Following the guidelines for the construction of an OSSE, the theory was adapted to the case to be studied, modifying it in order to obtain a robust framework.

- A study was carried out in order to assess the impact, on the species of interest for this study (i.e. the ozone), of the aerosol activation in MOCAGE. Given the very low impact encountered, it was possible to safely choose to set-up the model, without activating the aerosols for next steps in the OSSE construction. This reduced the computational time and made the process more efficient without impacting the quality of the output.

- The period chosen for the evaluation of the whole study runs from the 1st of June to the end of August 2019. The main reason for this choice was that 2019 saw very strong and well-documented, ozone pollution episodes. Two of them fall within the summer 2019.

- A special focus was put on the creation of the NR. It was decided to create a NR as close as possible to the real atmospheric state, by assimilating IASI radiances in the GLOB11 MOCAGE. The configuration implemented was the one suggested, so far, by the literature (i.e. assimilation of 284 channels in the LWIR ozone band; background errors computed from the percentage of the ozone profile itself; observation errors obtained by means of the Desroziers diagnostics). Once the best configuration was chosen, however, the NR designed for the OSSE was the corresponding MACC01 domain, where no assimilation was directly carried out, but which, indirectly, benefited of the assimilation performed in the global domain through the lateral boundary conditions.

  The topic of L1 data assimilation into a CTM is still being refined. A complementary study has, then, been performed with the aim of exploring the procedure of assimilation, to be exploited in a second stage also for IRS radiance assimilation, and to compare the NR framework to real observations (namely OMI and ozonesondes). This work provided interesting information about the possibility of improving the background errors evaluation and to, possibly, reduce the number of assimilated channels in future works. To support the literature, the results of this study have been submitted to the *Journal of Geophysical Research: Atmospheres.*

- The CR was set up so that no assimilation was carried out neither in the global and the regional domain, and that the surface emissions were modulated against those of the NR (referring to different years). The comparisons between these runs showed that the CR always produces smaller values of O_x compared to the NR. Although spatially consistent, they provide very different results. In terms of total columns relative differences range between 10 and 13%. For tropospheric columns the variations are rated to be between 10 and 20%. Finally, distribution of ozone concentrations along the vertical levels presents

a maximum variation of about 35%, located between approximately 25 and 180 hPa. This is considered enough to avoid the identical twin problem.

- Observations were simulated from the NR and through RTTOV version 12 for both IRS and IRS*2 (each using their specific coefficients). The radiances simulated for IRS cover the 195 contiguous wavelengths selected in the previous Chapter: from 982.464 to 1099.467 cm^{-1} [channel 503 to 697]. For IRS*2, 361 channels were simulated in the same way: from 982.343 to 1099.379 cm^{1} [channel 931 to 1291]. A horizontal thinning of the observation and a cloud filter were also applied to make the assimilation reliable and efficient in terms of computational time. The cloud filter follows the meteorological information provided by the meteorological forcing. The thinning provides one pixel simulated per box of 0.4°. This creates a good observation density along the time period of interest.

Chapter 7

Contribution of the Assimilation

In this Chapter the impact of the assimilation of both IRS synthetic observation into MOCAGE will be assessed for both simulated database. The assimilation set up, errors used and methodology will be described first. After that, the contribution of the main version of IRS to the CTM will be, at first, evaluated on its own. In the last part of the Chapter, finally, an inter-comparison between the contribution of each evaluated version of the instrument will be carried out.

7.1 Methodology

The assimilation of the synthetic observations of both IRS and IRS*2 was carried out inside the CR (on MACC01 domain), as previously introduced, in the time period going from the 15th of May, till the 31st of August, 2019. The evaluation was then performed in the three months of June, July and August. As already done during the work for the preparation of the NR described in Section 6.2 (6.2.2), the assimilation algorithm used was the 3D-Var with a hourly assimilation window. The role of the observation operator H was covered, once again, by the RTTOV version 12 in clear-sky conditions (the scattering by aerosols and clouds was not taken into account).

7.1.1 Background Errors

The $\mathbf{B}$ matrix was obtained through the same procedure as for the NR preparation study, when IASI L1c data were assimilated into the global domain of MOCAGE. The correlation terms have been modelled using a diffusion operator [as in Emili et al. (2019) and El Aabaribaoune et al. (2021)]. The diagonal terms, on the other hand, were obtained from a percentage of the ozone profiles themselves. In more detail, 2% of the ozone concentration at each atmospheric level was attributed to the background standard deviation (σ_{B}). The variances (i.e. σ_{B}^2) were then computed and attributed to the diagonal. An example of the shape of $\mathbf{B}_{\mathrm{perc}}$ is provided in Figure 7.1 for one day and hour (3rd of July, 2019, 10 UTC) representative of the general behaviour of the period of study, over the MACC01 domain (please notice that in

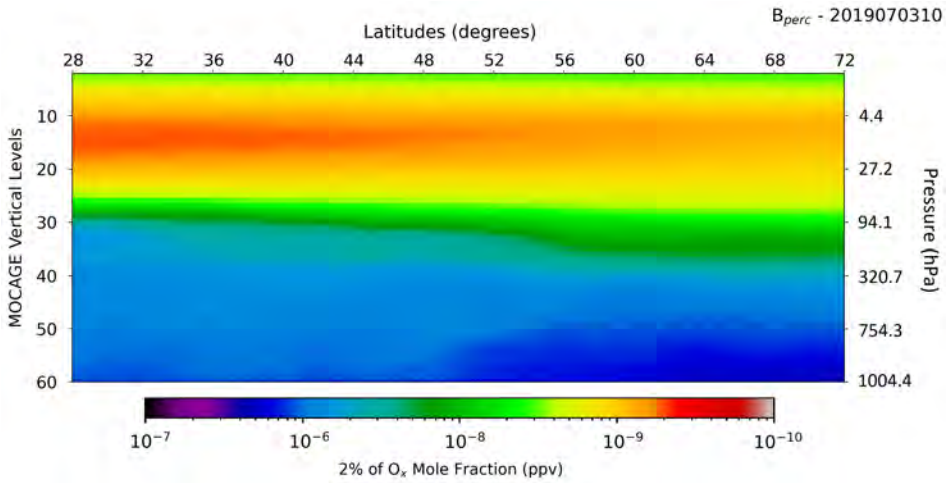


Figure 7.1: Example of $\mathbf{B}_{\text{perc}}$, i.e. 2% of the ozone concentration, for one random day and hour inside the period of study (3rd of July, 2019, 10 UTC). Values are averaged over longitudes.

Subsection 6.2.2 the same field has been shown for the GLOB11). The strongest values are found in stratosphere, between levels 20 and 10 (about 27 to 4 hPa) at latitudes between 28 and 44°. The smallest values, instead, are found in the lower-most troposphere at latitudes above 48°.

7.1.2 Observation Errors

The observation error was computed through the Desroziers's method illustrated in 4.1.4.2. As already explained, this kind of procedure is used to compute full $\mathbf{R}$ matrices, which has non-zero covariance terms, using observations, background and analysis. In order to have an initial analysis to use for this purpose, a first assimilation was performed using a diagonal $\mathbf{R}$ matrix. The variance values forming the diagonal have been determined using a fixed standard deviation $\sigma = 2.0$ mW/(m^2 sr cm^{-1}). This value was chosen so as to exceed the average values of the instrumental noise of both IRS and IRS*2.

Once the first analysis was available for both instrumental versions, the full $\mathbf{R}$ was computed through Equation 4.9.

The diagnosed standard deviations are shown in Figures 7.2a and 7.2b, together with the corresponding instrumental noise. The diagnosed σ shows different intensities depending on the version of the instrument, and thus of the noise, being treated. For IRS*2, in fact, the values have a higher variability but lower intensity than IRS. The diagnosed correlations, instead, are displayed in Figures 7.2c and 7.2d. Although the structures are found to be comparable in the two cases, higher correlation values are found for IRS. The difference is of the order of 10%, and it is complementary to the noise: the more intense the noise, the lower the inter-channel correlation.

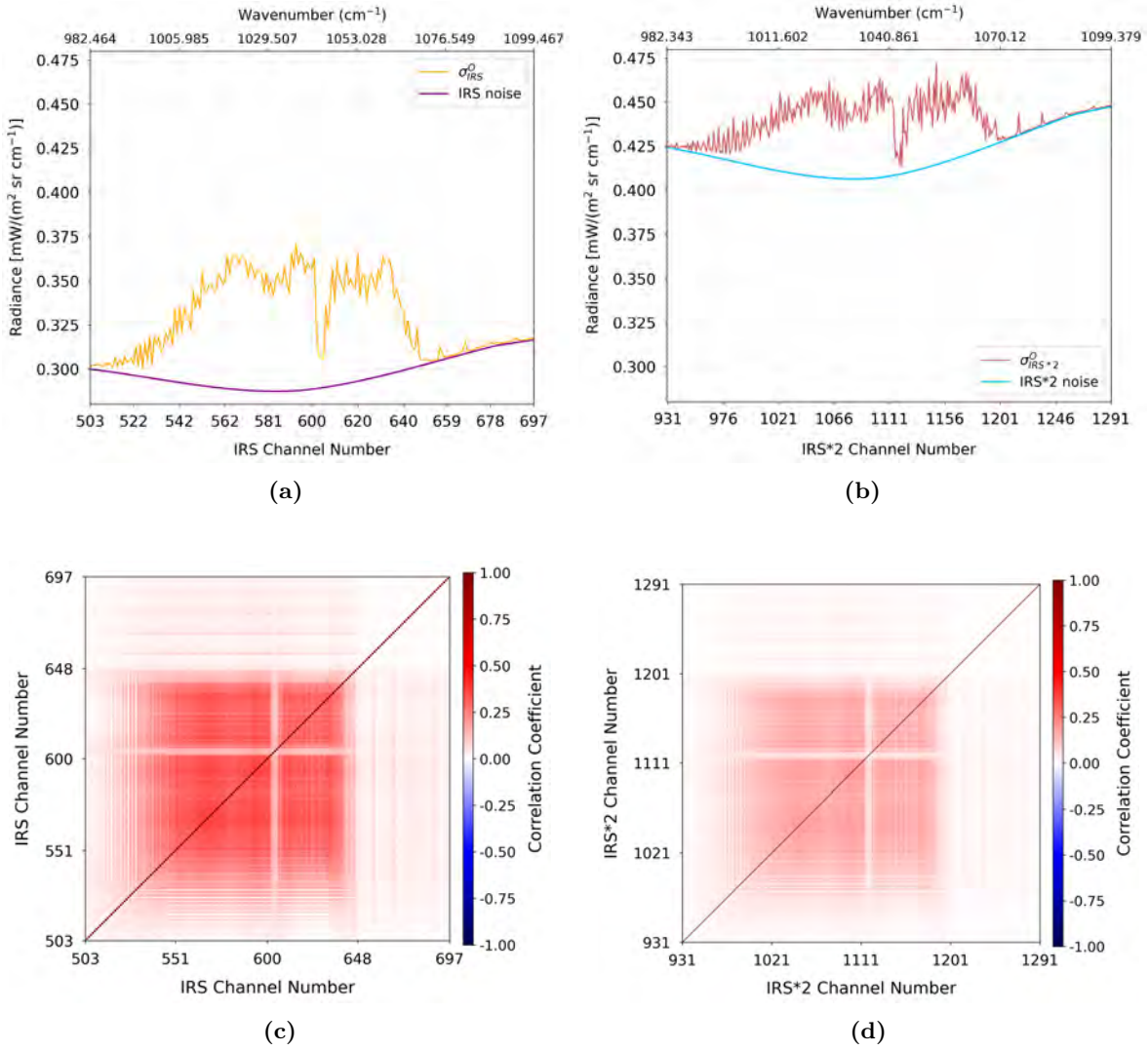


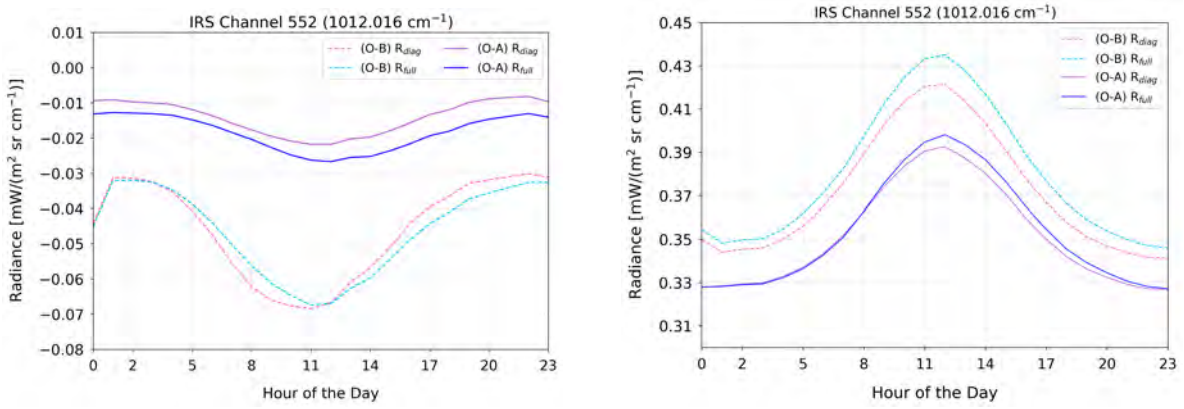
Figure 7.2: Diagnosed observation-error standard deviations and instrument noise for IRS in **(a)** and IRS*2 in **(b)**. The diagnosed error correlations are also shown in **(c)** and **(d)**, for IRS and IRS*2 respectively.

7.2 Impact of the Assimilation of IRS Observations

7.2.1 Statistics on the Observations

A first assessment of the assimilation of IRS radiances into the MOCAGE model has been carried out through the evaluation of *Observations minus Background* (O-B), i.e. innovations, and the *Observations minus Analysis* (O-A), i.e. residuals. Although the results evaluated in the rest of the Chapter will be those obtained using $\mathbf{R}_{\text{full}}$, at this stage several statistics in the space of observations have been computed for both the runs using an $\mathbf{R}_{\text{diag}}$and an $\mathbf{R}_{\text{full}}$. This insight can potentially provide some extra information about the success of diagnostics and the correct use of observation errors. In addition, despite the assimilation was performed over the period ranging from the 25th of May, until the end of August, the statistics have been evaluated over a shorter period including the three entire months of June, July and August, 2019. This choice, excluding the 6 days of assimilation, allows the system to better stabilise before an evaluation is carried out. Indeed, the assimilation trials described in this study consist in a continuous hourly assimilation cycle over the period. This means that each assimilation time creates an analysis, influenced by the observations, which is the initial state of a 1-hour forecast, that is, in turn, the background state of the next assimilation time. Effects of the observation are propagated from one assimilation to the next one, reaching some steady regime in the assimilation cycle.

Statistics of (O-A) and (O-B) by hour of day, computed over the just mentioned 3-month period, are displayed in Figure 7.3 and allow to evaluate the diurnal cycle. The results refer to an arbitrarily-selected wavelength among those simulated, i.e. IRS channel number 552 ($1012.016\,\text{cm}^{-1}$), which is representative of most of ozone-sensitive channels in the spectral range used in this study. Such a channel presents, indeed, a Jacobian picking around vertical

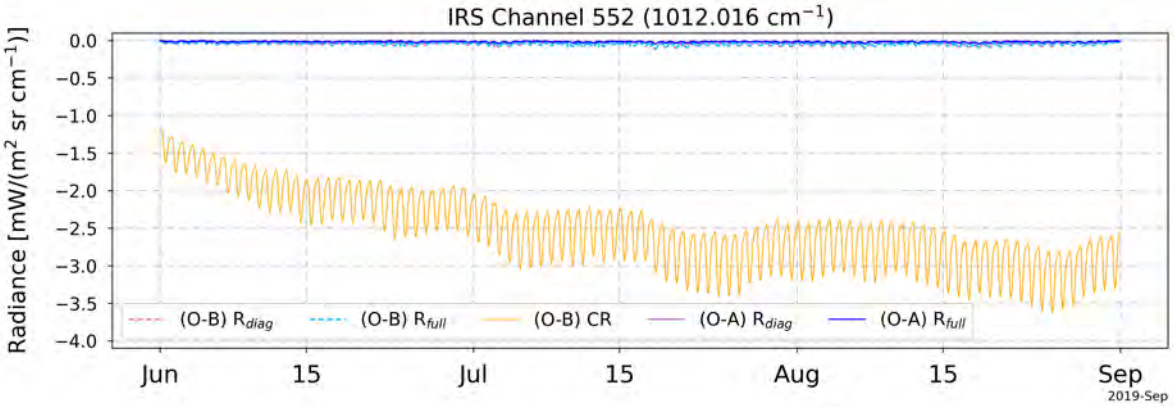


(a) Averages **(b)** Standard Deviations

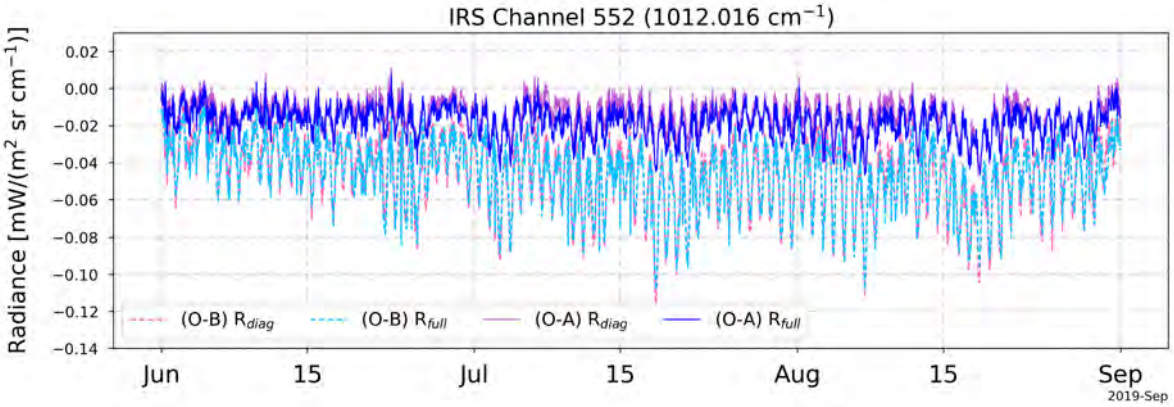
Figure 7.3: Statistics of the innovations (O-B) and residuals (O-A) computed per hour of the day over the period of study 1st of June - 31st of August, 2019. The assimilation runs using first an $\mathbf{R}_{\text{diag}}$, then the diagnosed $\mathbf{R}_{\text{full}}$. The averages are shown in **(a)**, while the standard deviations in **(b)**. Results refer to IRS channel 552 ($1012.016\,\text{cm}^{-1}$).

level 40 ($\sim$ 320 hPa), while, when weighted with the 10% of the ozone profile, it shifts its sensitivity between level 20 and 25, i.e. between 25 and 50 hPa (Jacobians in Figure 6.16). The averages (7.3a) show residuals always smaller compared to the innovations. This is an indication of successful assimilation that produces analyses closer to the observations than the background state. Comparing what obtained through the use of the two different $\mathbf{R}$ matrices, analysis is closer to observations when using $\mathbf{R}_{\text{diag}}$. This effect is due to the presence of covariance terms in the $\mathbf{R}_{\text{full}}$, causing, as expected, a slightly less efficient minimisation in terms of analysis. The differences on the innovations look to be shifted in certain hours of the day. The standard deviations of the residuals (7.3b), on the other hand, result very close in the first eight hours of the day, while slightly diverging during the central hours.

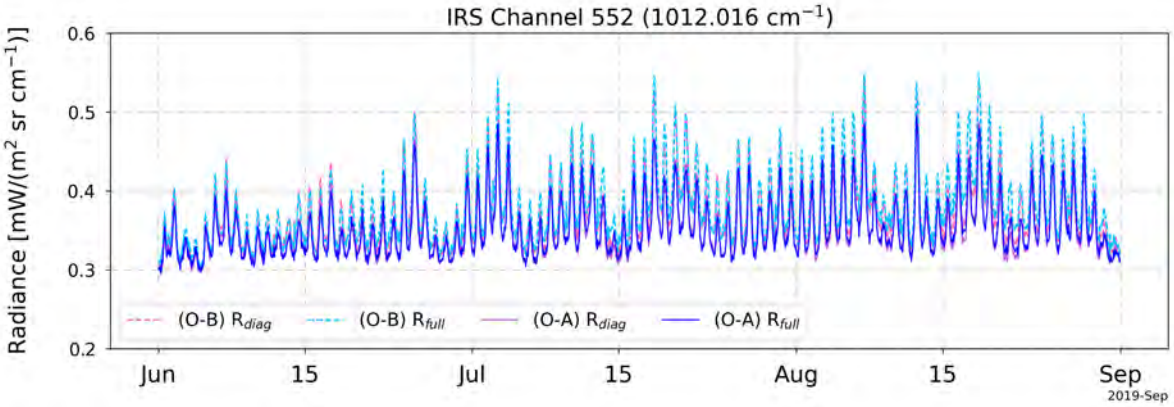
Averages, and associated standard deviations, have also also been computed per hour of the day for each day in the 3-month time period (Figures 7.4). Panel 7.4a, in addition to the averages of innovations and residuals from the runs using $\mathbf{R}_{\text{diag}}$ or $\mathbf{R}_{\text{full}}$, reports the (O-B) of a run without assimilation, i.e. the CR. Looking at the trend of the latter, in particular, one finds a divergence over time from the values assumed at the beginning of the period considered. Furthermore, such a curve assumes values up to 3.5 radiance units stronger if compared to the runs with assimilation. This means that the background state, as the run time increases, steadily diverges from the observations. When evaluating the runs were assimilation is performed, instead, the trend of the innovations remains stable and closer to zero over the entire time-series. The averages are also plotted on a y-axis scale allowing to better assess their fluctuations in Panel 7.4b. The values assumed by the residuals are again always lower than the innovations and fluctuate depending on the day and the hour. The intensity of the variations appears to be weaker in June than in the other two months. The trend of the standards deviations (7.4c) shows (O-B) and (O-A) that are comparable for both $\mathbf{R}$ exploited. Values for residuals, on the other hand, are smaller than for innovations. As for the averages, the variations are more intense in the months of July and August compared to June.



(a) Averages



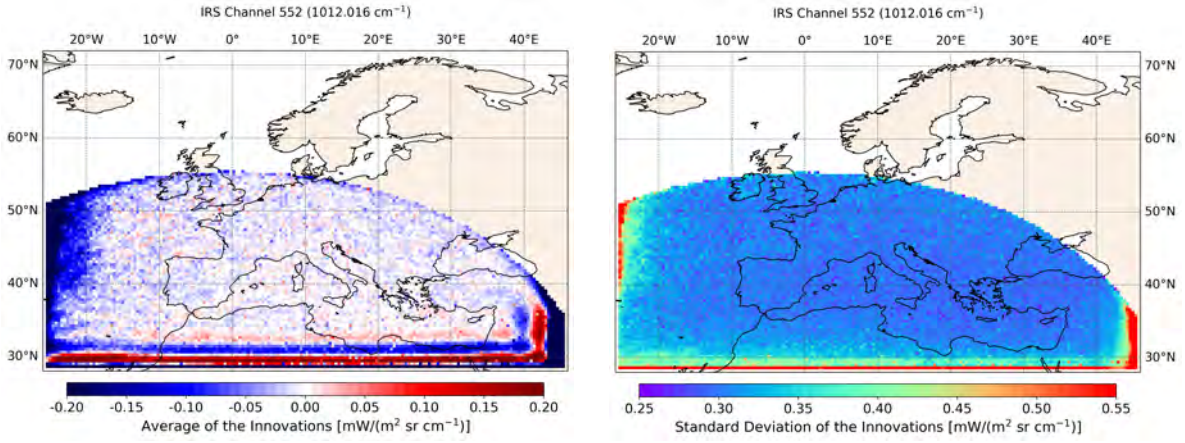
(b) Averages (zoom)



(c) Standard Deviations

Figure 7.4: Statistics of the innovations (O-B) and residuals (O-A) computed per each day and hour of the period of study (1st of June - 31st of August, 2019). Assimilation runs using first an $\mathbf{R}_{\text{diag}}$, then the diagnosed $\mathbf{R}_{\text{full}}$, and a run without assimilation (i.e. the CR) are taken into account. The averages are shown in two scales of the y-axis **(a)** and **(b)**, while standard deviations are in **(c)**. Results refer to IRS channel 552 (1012.016 cm^{-1}).

A geographical evaluation of averages and standard deviations of the innovations (O-B) has been carried out. An example is shown again on channel 552 in Figure 7.5. In both cases, extreme values are found at the edge of the domain. This could be explained by the lateral boundary conditions bringing information from outside the domain, where no IRS observations are assimilated. As already explained, due to the continuous assimilation cycle, the background in the inner part of the domain is more consistent with observations. Conversely, at the edges the ozone field from the coupling model (global) shows more discrepancies with observations. This trend will have to be taken into account in the evaluation of statistics carried out on the entire regional domain. At a later stage, one may consider performing such evaluations on a smaller domain that excludes these adjustment values.

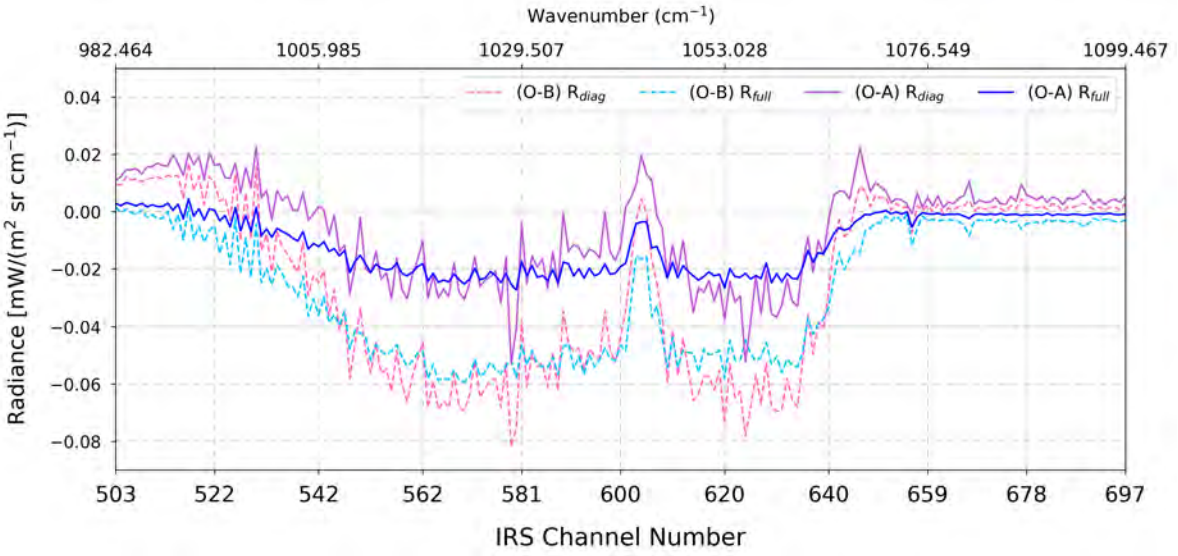


(a) Averages **(b)** Standard Deviations

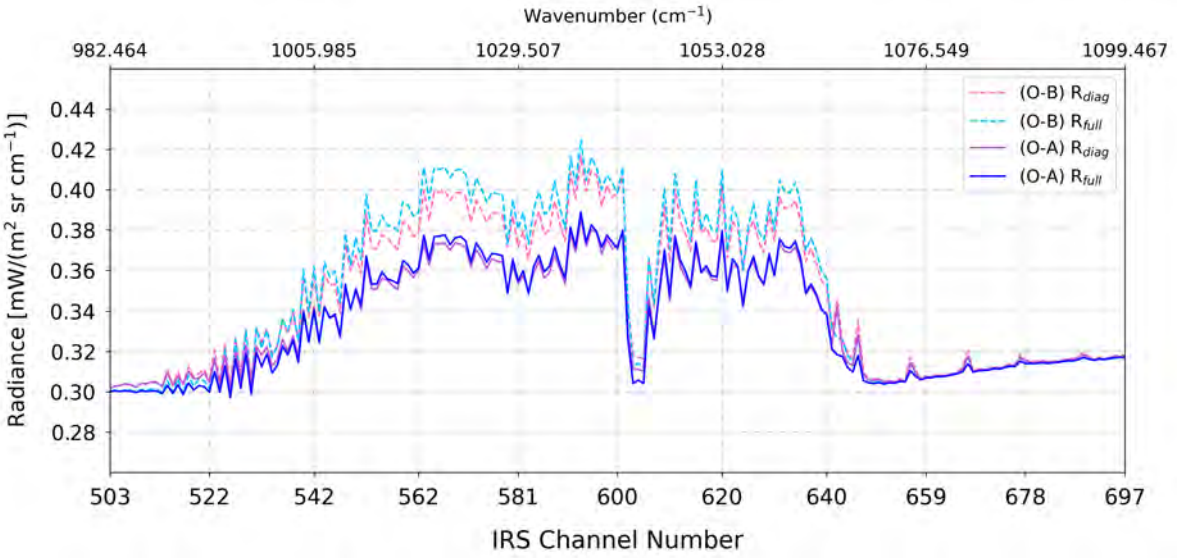
Figure 7.5: Averages **(a)** and standard deviations **(b)** of the innovations (O-B) computed over the period of study (1st of June - 31st of August, 2019). Results refer to IRS channel 552 (1012.016 cm^{-1}).

Although the results shown in the last two cases refer to only one channel, statistics were also computed per each wavelength in the range chosen for IRS. Averages and standard deviations, evaluated over the 3-month time-period, are shown in Figure 7.6.

At the beginning of the band, up to almost channel 545, innovations and residuals turn out to deviate more than for all the other wavelengths (for both $\mathbf{R}$ matrices considered). The standard deviations, on the other hand, are all found to vary little from each other in the atmospheric window area, while they deviate by a few tenths of a radiance unit at the level of the most ozone-sensitive channels. Again, the minimisation is well efficient and in the ozone-sensitive areas this is better than at the edges.



(a) Averages



(b) Standard Deviations

Figure 7.6: Statistics of the innovations (O-B) and the residuals (O-A) computed over the period of study (1st of June - 31st of August, 2019) for the range of wavelengths considered for IRS (982.64 to 1099.467 cm[−1]). Averages are shown in **(a)**, while standard deviations are in **(b)**. Assimilation runs using first an $\mathbf{R}_{\text{diag}}$, then the diagnosed $\mathbf{R}_{\text{full}}$ are taken into account.

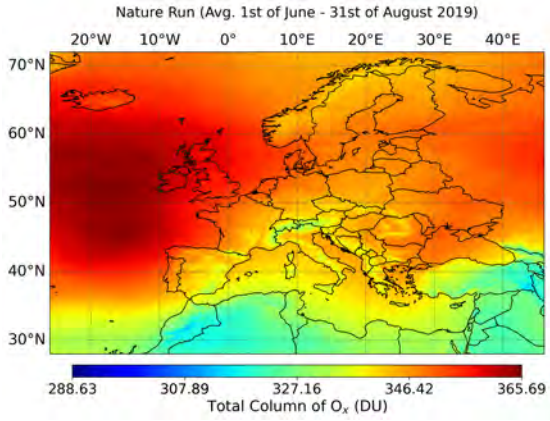
7.2.2 Evaluation of the Assimilation

A verification of the analysis against the NR was then performed in order to evaluate the impact of the IRS assimilation on the O_x field produced by MOCAGE.

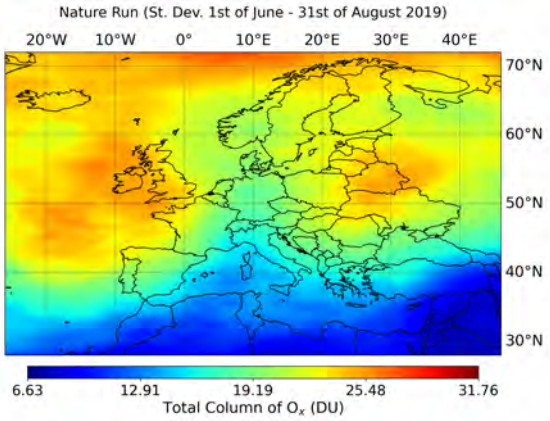
7.2.2.1 Total Columns

The values of O_x total columns in output of the assimilation run using IRS simulated data, averaged over the 3-month time period (1st of June - 31st of August, 2019), are shown in Panel 7.7c, with the corresponding standard deviation in 7.7d. To have an idea of the impact of assimilation on the ozone field, these statistics are compared to the averages values and standard deviations of what was obtained for the NR (in Panels 7.7a and 7.7b, respectively). Average and standard deviation of the differences between the two runs are, finally, shown in 7.7e and 7.7f. From an analysis of the averages, it is found that the AR is very close to the NR in the centre of the area in which the observations are assimilated. the variation, more specifically, is around 0% over the Mediterranean basin and increases progressively while approaching the edges of the assimilation area, but not exceeding -3% (i.e. NR provides slightly stronger values of Ozone total columns than the AR). Maxima of divergence of these runs are found outside the area where the observations are present (up to -13%). At the lower edge of the domain, further variations are found and can be linked to the features described for the innovations in Figure 7.5. Looking at the standard deviations of the differences (7.7f), finally, low values (on the order of a few DU) are found in the circle where the IRS observations were simulated and then assimilated. Outside this area, the values rise up, reaching maxima in the North of the domain. The impact of assimilation is therefore evident, especially if comparing what was achieved when no assimilation was not carried out, i.e. with the CR. The analysis made in the previous Chapter when comparing CR and NR, in fact, found way stronger values of variation, going from -10% up to -13% for the averages (see Figure 6.6c). Remarkable is also the the reduction in the error of the bias with respect to the NR (Panel 7.7f), which, in the area where IRS radiances are assimilated, reaches its lowest values (magenta area). The minima, around 1.82 DU, are located in the South-Est quarter of the domain, where the highest concentration of simulated, and assimilated, observations is found over the three months (look back at Figure 6.14).

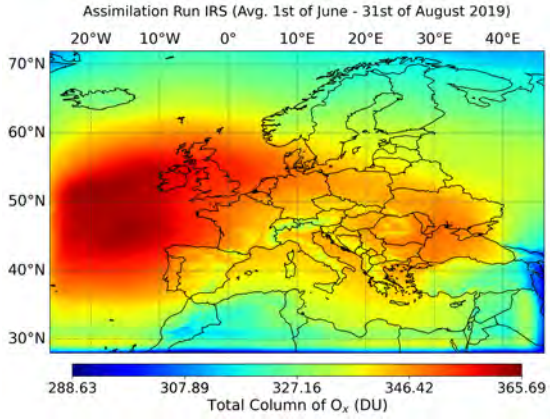
Averages on the bias, errors and correlation occurring between AR and NR have also been evaluated over the MACC01 domain, using NR, indeed, as the reality. Figure 7.8 shows the FGEs for each of the three months considered and for the three of them together. The evaluation of this parameter finds values very close to zero where IRS radiances are assimilated. The reduction in error is though clear if compared to what was obtained for the CR (values of FGE up to ~ 0.2 - Figure 6.7). Extreme values are found, once again, at the bottom margin of the domain, caused by the effects of coupling between the global and regional domain. The same kind of trend is registered when assessing the MNMB in Figure 7.9, which displays bias values very close to zero for the four charts, with slightly different structures at the top



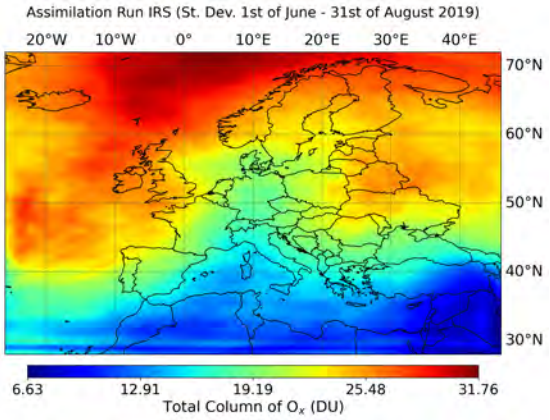
(a) $\overline{\text{NR}}$



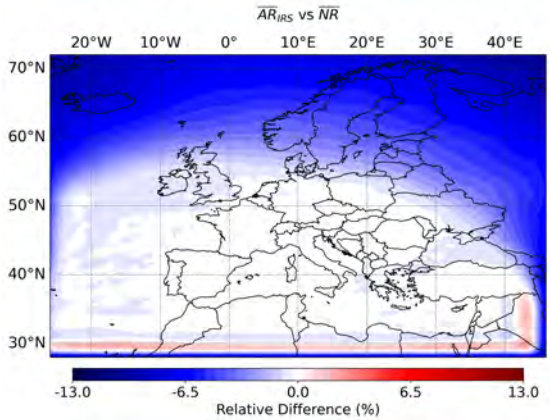
(b) $\sigma(\text{NR})$



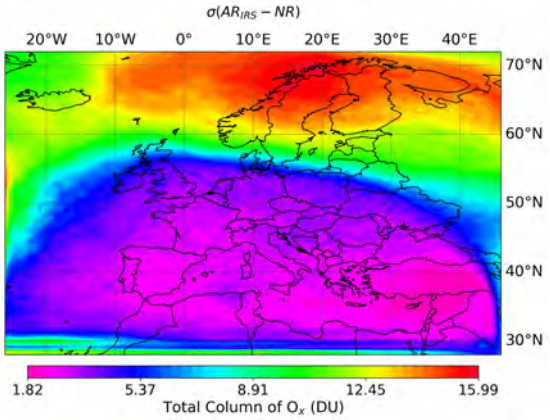
(c) $\overline{\text{AR}}_{\text{IRS}}$



(d) $\sigma(\text{AR}_{\text{IRS}})$

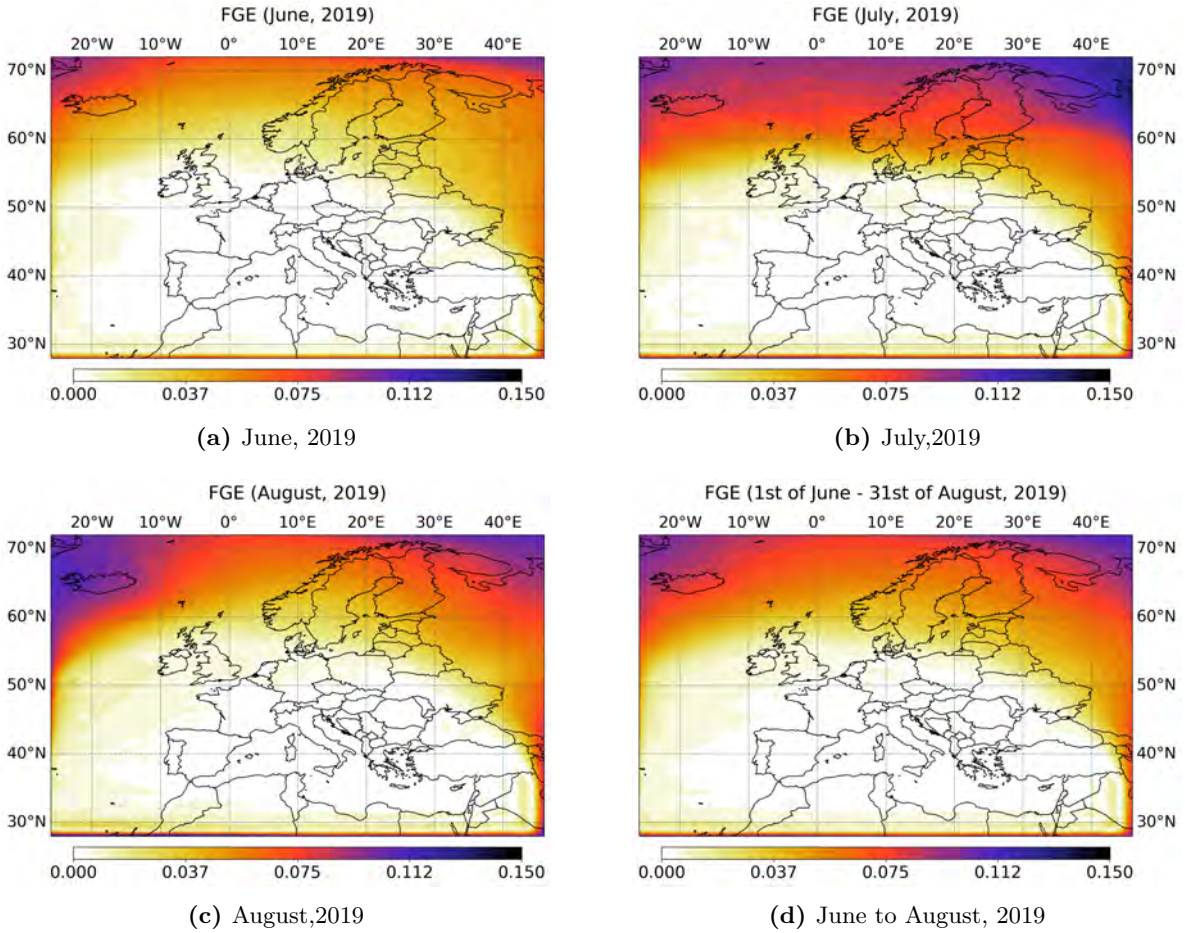


(e) Relative Difference $\overline{\text{AR}}_{\text{IRS}}$ vs $\overline{\text{NR}}$



(f) $\sigma(\text{AR}_{\text{IRS}} - \text{NR})$

Figure 7.7: Averages over the 3-month period of study (1st of June up to 31st of August, 2019) of the O_x Total Columns from NR **(a)** and AR of IRS **(c)**. Corresponding Standard Deviations are shown in **(b)** and **(d)**, respectively. Finally, **(e)** is the Relative Difference between the average O_x total columns of the AR IRS and the NR, while **(e)** represents the error of the differences between the two runs.



(a) June, 2019

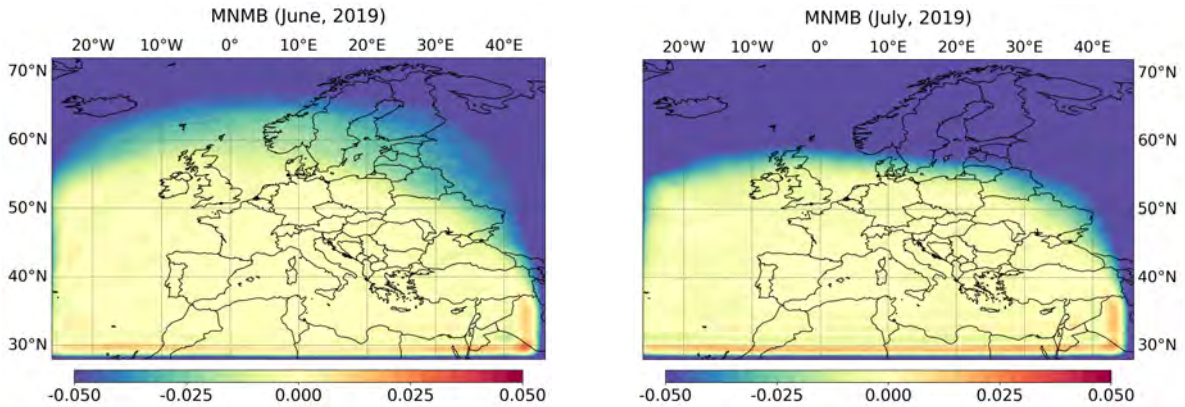
(b) July,2019

(c) August,2019

(d) June to August, 2019

Figure 7.8: Fractional Gross Error (FGE) computed between O_x Total Column from AR of IRS observations and the NR, averaged over the month of June **(a)**, July **(b)**, August 2019 **(c)** and over the 3 months **(d)**.

borders of the area where simulated observations are located. Positive biases are detected outside the area of observation assimilation, while negative values are at the lower border of the domain. The AR and NR, on the other hand, present a positive correlation almost over the whole domain for each of the months have been evaluated and for the three at the same time (Figure 7.10). Stronger values are found where the assimilated observations lay, while Weakly negative values are in the law-right border of the domain during the month of August.



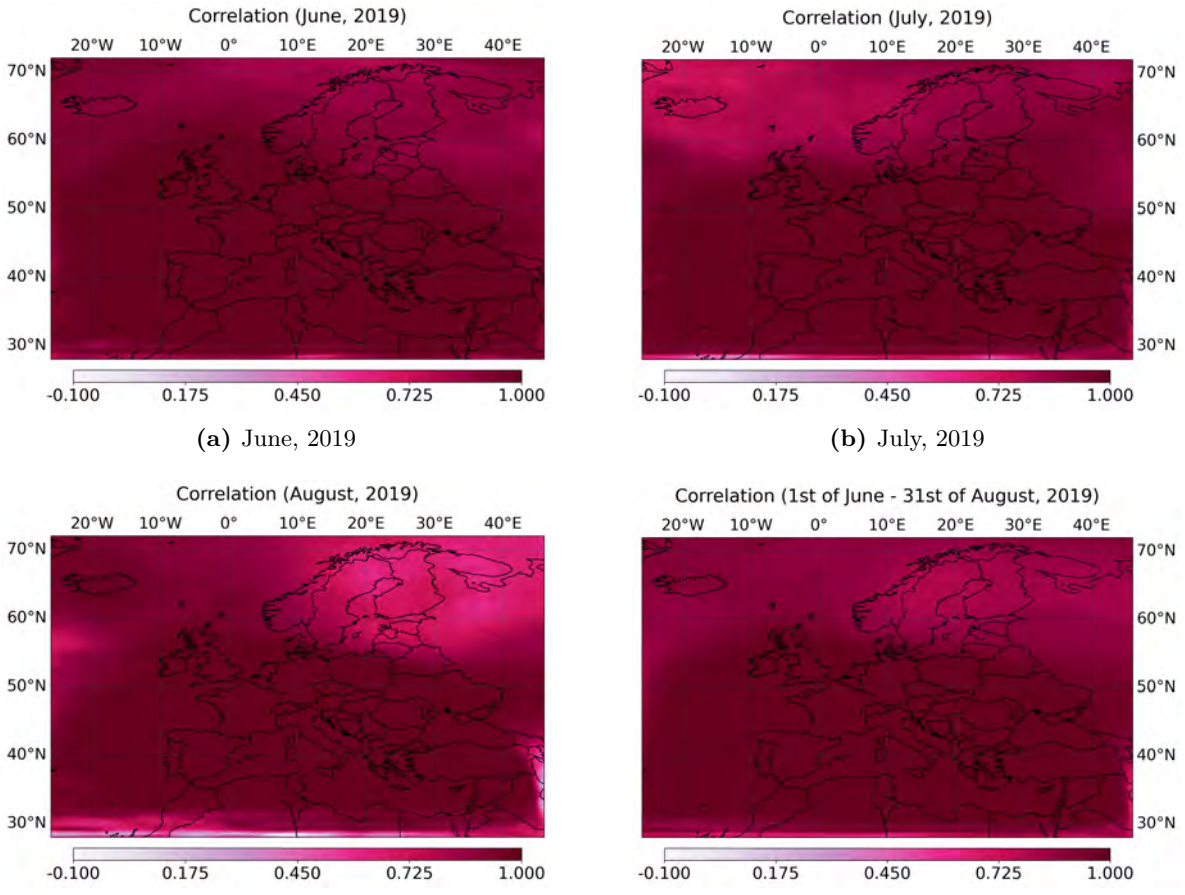
(a) June, 2019

(b) July, 2019

(c) August, 2019

(d) June to August, 2019

Figure 7.9: Modified Normalised Mean Bias (MNMB) occurring between O_x Total Column from AR of IRS observations and the NR, averaged over the month of June **(a)**, July **(b)**, August 2019 **(c)** and over the 3 months **(d)**.



(a) June, 2019

(b) July, 2019

(c) August, 2019

(d) June to August, 2019

Figure 7.10: Pearson Correlation Coefficient between values of O_x Total Column from AR of IRS observations and NR, averaged over the month of June **(a)**, July **(b)**, August 2019 **(c)** and over the 3 months **(d)**.

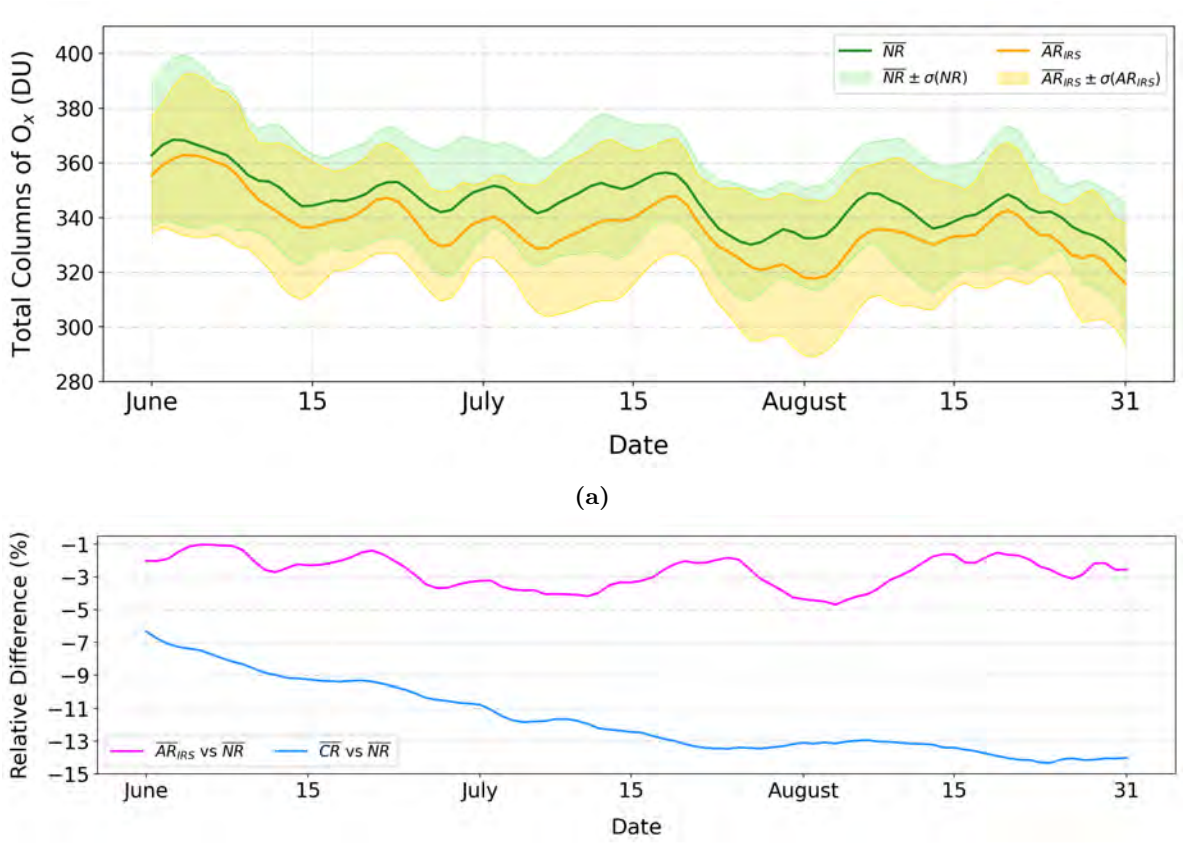


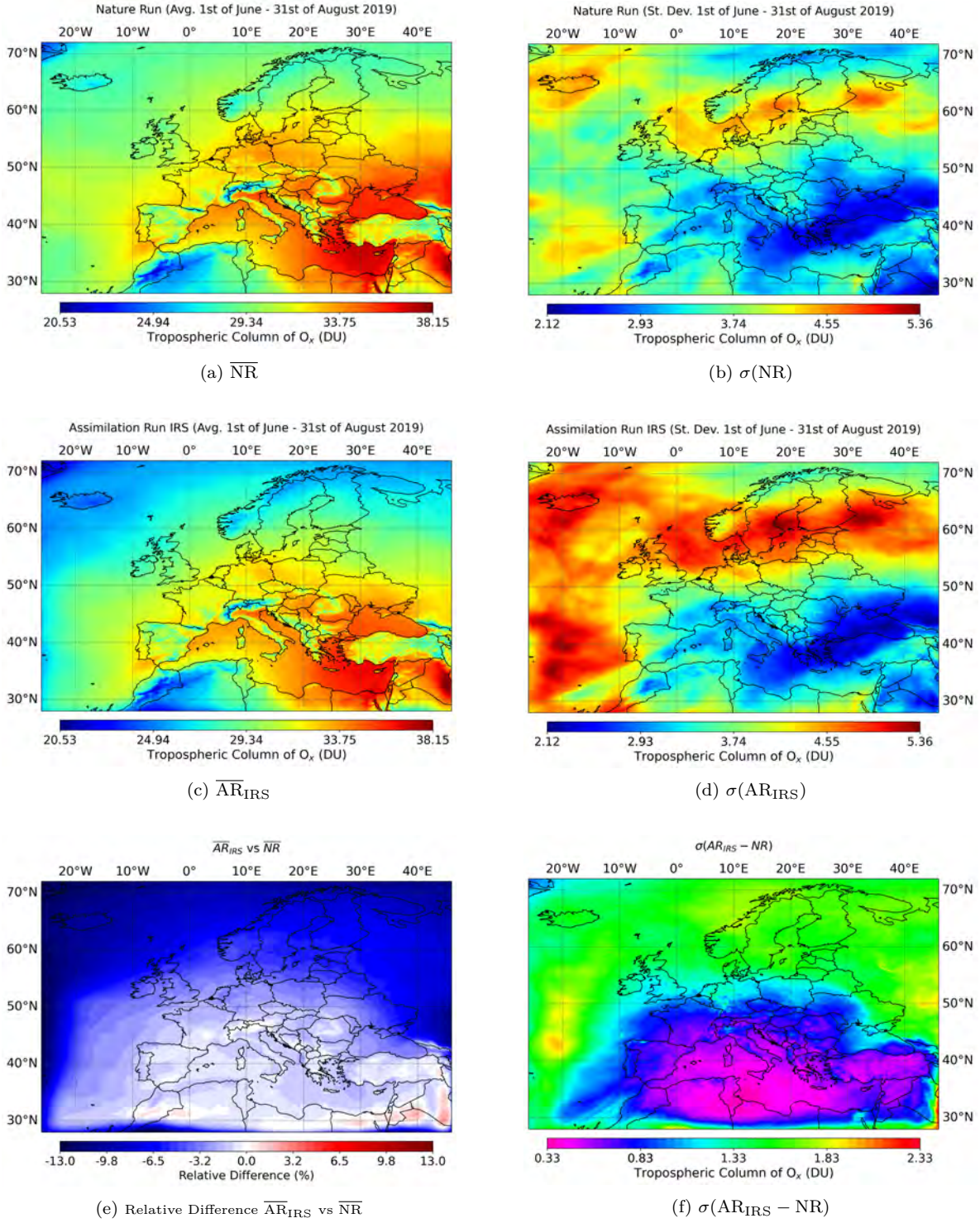
Figure 7.11: Time series of the averages per day of the O_x Total Columns retrieved by MOCAGE through the NR and AR for IRS, plus and minus the corresponding standard deviations. Percentage differences between the averages of the two runs are shown in fuchsia in **(b)**, where the differences that were obtained between CR and NR (as in 6.8b) are also shown for comparison (blue).

Time series of the daily averages of O_x total column values, also averaged over the entire MACC01 domain, are shown in Figure 7.11a, with the relative standard deviations added and subtracted in shades. The run with assimilation follows the same upward and downward trend observed for the NR. As quantitatively determined by the relative percentage difference in Figure 7.11b, the AR returns O_x total column values that are more intense than those of the NR by a minimum of 1% to a maximum of 5%. These percentages are of a lower magnitude than those encountered when comparing CR to NR in the previous Chapter (Figure 6.8). The results there obtained are here shown again (blue curve) for an easier comparison. The CR, indeed, provided values up to almost 15% smaller than the NR towards the end of the period of study. The CR, indeed, diverged because it had no constraints, such as assimilation, that keep it stable (same trend as in Figure 7.4a).

7.2.2.2 Tropospheric Columns

As already done when comparing CR and NR, we also want to assess the impact of IRS assimilation on the tropospheric column. As before, this is considered to correspond to MOCAGE levels ranging between 40, i.e. $\sim$ 220 to 330 hPa, and 60, i.e. the surface (empirical evaluation based on the approximate position of the tropopause).

Figure 7.12 displays the tropospheric columns, averaged over the period from June to August, for the NR (7.12a) and AR of IRS (7.12c), together with the corresponding standard deviations (7.12b and 7.12d). It is easy to see that much finer structures are appreciable, with differences between mountains, valleys and water surfaces. The standard deviations present the strongest values when looking at the AR of IRS, in the area outside the IRS radiance assimilation zone (above 53°N). The relative difference between the averages (7.12e) reports values further from zero than what was obtained for the total columns (7.7e), where they were around 1%. In this case, variations of the order of 2 to 3% are observed. Although the influence of the coupling between global and regional domain at the lower border is still present, it is less pronounced than in the case of the total columns. The error of the biases (7.12f) is lower than in 7.7f (also due to the lower concentrations) with values from 0.3 to 2.3 DU. The minima are encountered on the Central-Mediterranean area and Tunisia. A different structure becomes noticeable, in this case compared to the case of the total columns, on the left side of the domain between 40 and 50°N, which shows slightly higher values (about 2 DU) of error associated with the bias between the two runs.



(a) $\overline{\text{NR}}$

(b) $\sigma(\text{NR})$

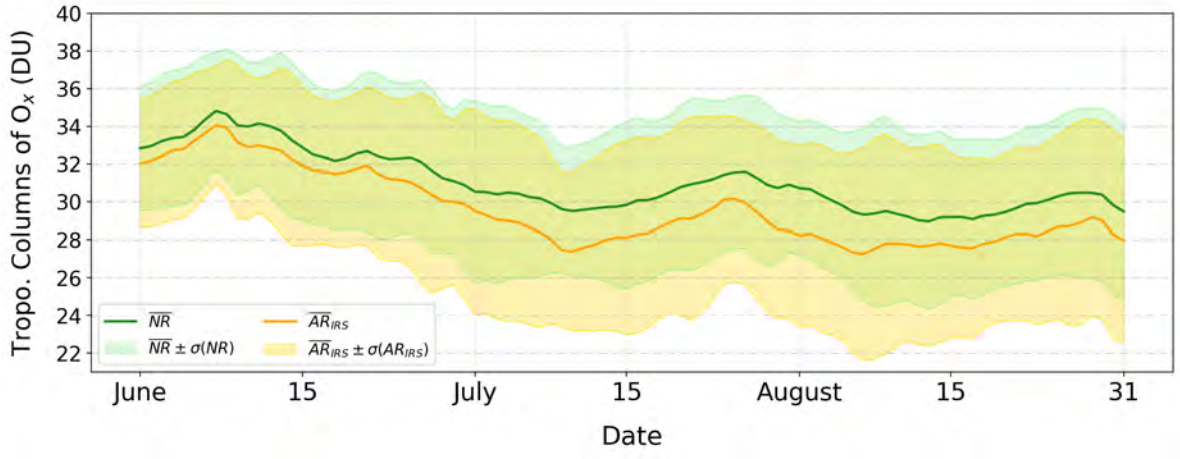
(c) $\overline{\text{AR}}_{\text{IRS}}$

(d) $\sigma(\text{AR}_{\text{IRS}})$

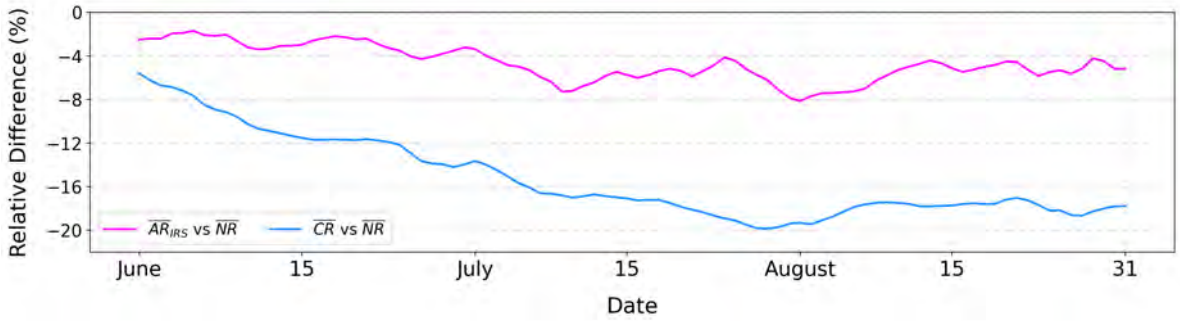
(e) Relative Difference $\overline{\text{AR}}_{\text{IRS}}$ vs $\overline{\text{NR}}$

(f) $\sigma(\text{AR}_{\text{IRS}} - \text{NR})$

Figure 7.12: Averages over the 3-month period of study (1st of June up to 31st of August, 2019) of the O_x of the Tropospheric Columns from NR **(a)** and AR of IRS **(c)**. Corresponding Standard Deviations are shown in **(b)** and **(d)**, respectively. Finally, **(e)** is the Relative Difference between the average O_x Tropospheric Columns of the AR IRS and the NR, while **(e)** represents the error of the bias between the two runs.



(a)



(b)

Figure 7.13: Time series of the averages per day of the Tropospheric Columns of O_x from the NR and AR for IRS, plus and minus the corresponding standard deviations. Percentage differences between the averages of the two runs are shown in fuchsia in **(b)**, where the differences that were obtained between CR and NR (as in 6.10b) are also shown for comparison (blue).

As already done in Figure 7.11, by averaging the values of O_x tropospheric columns over the domain and per day in the considered time period, the time series in Panel 7.13a is obtained. The relative differences between the curves of the NR and AR IRS averages, juxtaposed with the differences that were obtained when comparing CR to NR, are shown in 7.13b. The daily averages for the AR differ little from it (a maximum of 8% on the 1st of August), compared to the CR, which, on the other hand, reported values up to 20% lower than the NR.

Worth to notice are also the most sudden variations of peaks and drops. Please observe the growth in tropospheric O_x at the end of July (27th). This event, although it should be studied punctually over the locations where the event took place, corresponds to the day in which one of the ozone pollution episodes illustrated in Subsection 6.1.1 was recorded. The AR, moreover, is much closer to the NR during this period, than the CR was. Other peaks, although less marked than in the previous case, can also be observed between the 20th and the 28th of June, which correspond to another episode of ozone pollution over Europe (Subsection 6.1.1).

7.2.2.3 Vertical Variations

In addition to the columns, the impact of assimilating simulated radiances of IRS into MOCAGE was also assessed by examining what happens on the 60 atmospheric vertical levels that the model is able to simulate.

O_x mole fraction averaged over the 3-month study period and the longitudes, though, have been computed and are shown in Figure 7.14 for the NR (7.14a) and the AR for IRS (7.14c). A zoom on the situation occurring in the lower levels of atmosphere is depicted in the corresponding Panels on the right-hand side (7.14b and 7.14d), where an adapted colour scale was exploited in order to highlight the smaller concentrations and their distributions found in these layers. At first glance, one would say that the vertical distributions of ozone concentration are consistent in the higher levels of the atmosphere. The assimilation run seems to be consistent with the NR itself, although finer structures are appreciable around level 20 at latitudes around 32°. On the other hand, differences are more easily detectable in the Panels showing the lower-most levels, where the values of O_x close to the surface in the NR scenario, from 36° to about 44° of latitudes, are stronger than the values obtained in the corresponding layers and locations for the AR. In order to assess these differences more precisely, the percentage relative differences between the two runs have been computed and are shown in Panels 7.14c and 7.14f, while the standard deviation of the differences are in Panels 7.14g and 7.14h. Between levels 28 and 38, at latitudes from 60° to 72°, the AR presents values that are lower than the NR ones by up to about 15%. High values of variation are also found between levels 28 and 29, at latitudes 28° and 29°, not exceeding 27%. As already discussed in the previous Chapter (Subsection 6.2.3) the CR showed values always smaller than the NR and diverged from it up to a maximum of 35% in the range between levels 35 and 25. In contrast to the CR, instead, the AR also sometimes produces higher concentration values than the NR, and this also leads to positive relative differences. This is evident in the regions around level 29 to 20, at latitudes 29° to 36°, where relative differences up to 7%. This is, however, the area where the effects of the coupling of the GLOB11 domain and the MACC01 mostly occur. In the lower levels, on the other hand, the values of divergence from the NR remain close to zero up to 48°, while reaching slightly higher values close to the surface around latitude 52°. This means that, in these layers, the assimilation of IRS synthetic observations into MOCAGE, leads to average values of ozone mole fraction that are 15% stronger than in the NR. Please notice that, as previously broadly illustrated, the IRS observations that are assimilated are localised inside the disk of IRS acquisition, which does not cross the 60°N. Differences witnessed beyond this limit are, therefore, not attributable to the effects of the assimilation. In general, the more IRS observations, the smaller the differences between the AR and NR.

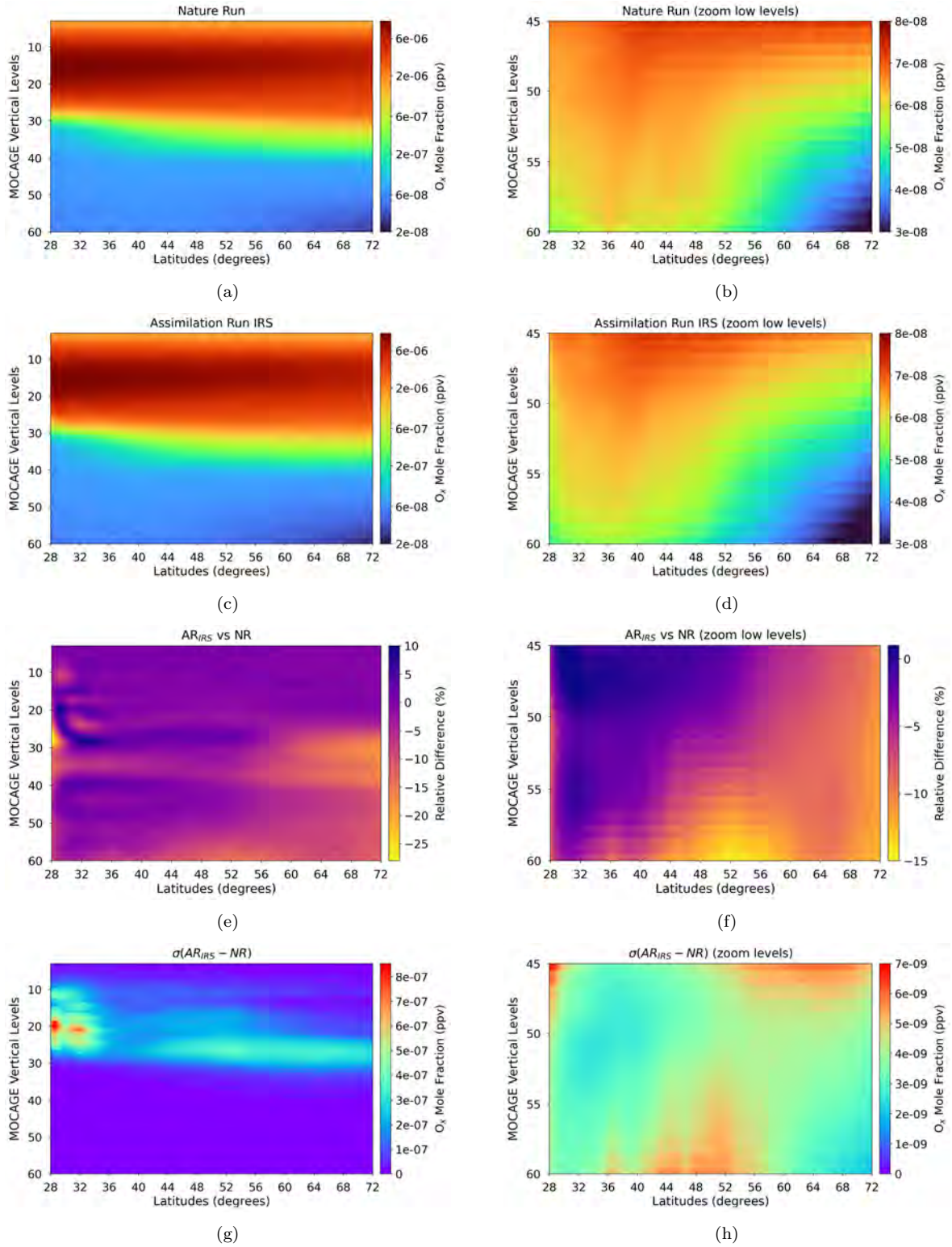


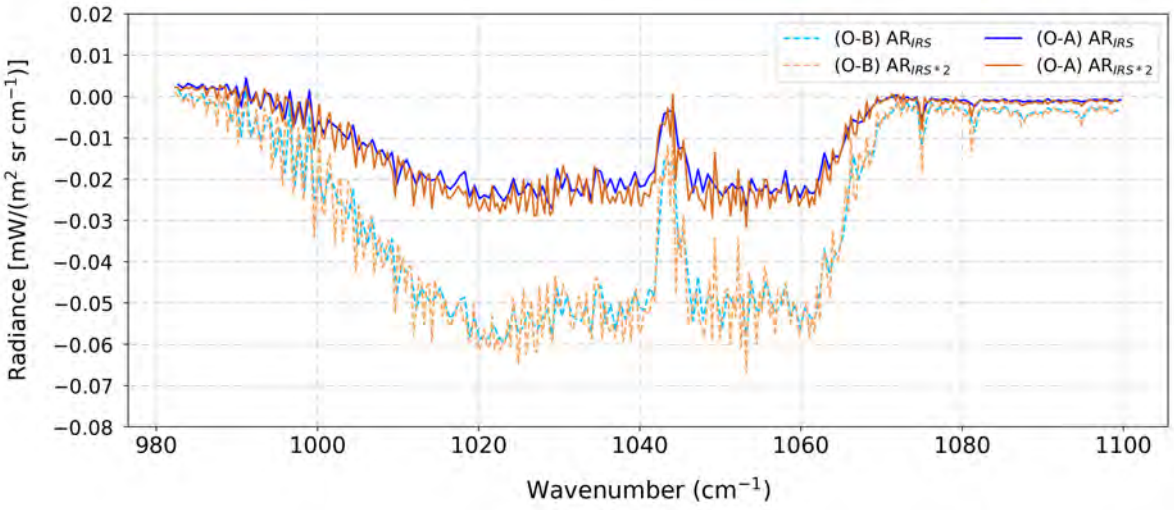
Figure 7.14: Average of the O_x Mole Fraction, for each MOCAGE vertical level, computed over the longitudes of the full domain and the time period (1st of June - 31st of August, 2019). The results are shown over the entire vertical column and in a zoom of the lower levels (from 60 to 45), for the NR in **(a)** and **(b)**, and for the AR of IRS in **(c)** and **(d)**. Relative differences between the two cases are shown in **(e)** and **(f)**. For equivalence between model and pressure levels see Figure 7.24 or refer to Appendix A.

7.3 Impact of the Assimilation of IRS*2 Observations vs IRS

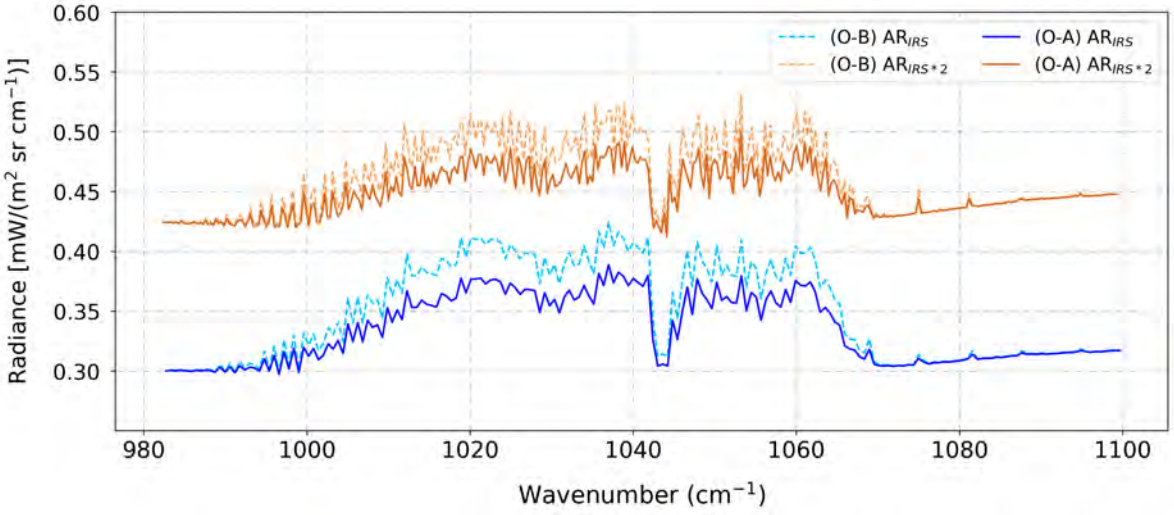
The studies shown so far for IRS assimilation have also been carried out for the second version of the IRS with an improved spectral sampling (IRS*2). However, the variations in the results obtained through the two instrument versions would not be appreciable without directly evaluating their deviation. This part of the Chapter, therefore, will focus on showing an inter-comparison between the impact of the assimilation of each IRS into MOCAGE.

7.3.1 Statistics on the Observations

Similarly to what was shown for IRS in 7.2.1, statistics on the innovations and residuals have also been computed for IRS*2. Figure 7.15 shows the statistics over the 3-month study period, for each of the channels simulated for IRS*2 (361 wavelengths). The results that had been obtained for IRS (on the corresponding range of wavelengths, 195 channels) are also shown in order to facilitate an inter-comparison. Please notice that only the statistics for the runs exploiting a $\mathbf{R}_{\text{full}}$ are here compared. For the hyperspectral IR sounders, indeed, it is more convenient to take into account the observation error correlations. An assessment of the averages (Panel 7.15a) shows comparable values, between innovations and between residuals of both IRS versions. The standard deviations, on the other hand, reveal higher values for IRS*2 compared to IRS, i.e. the minimisation is less efficient in the case of IRS*2. It is very likely the noise leading these values that are higher for IRS*2, plus the larger amount of channels for this version of the instrument.



(a) Averages



(b) Standard Deviations

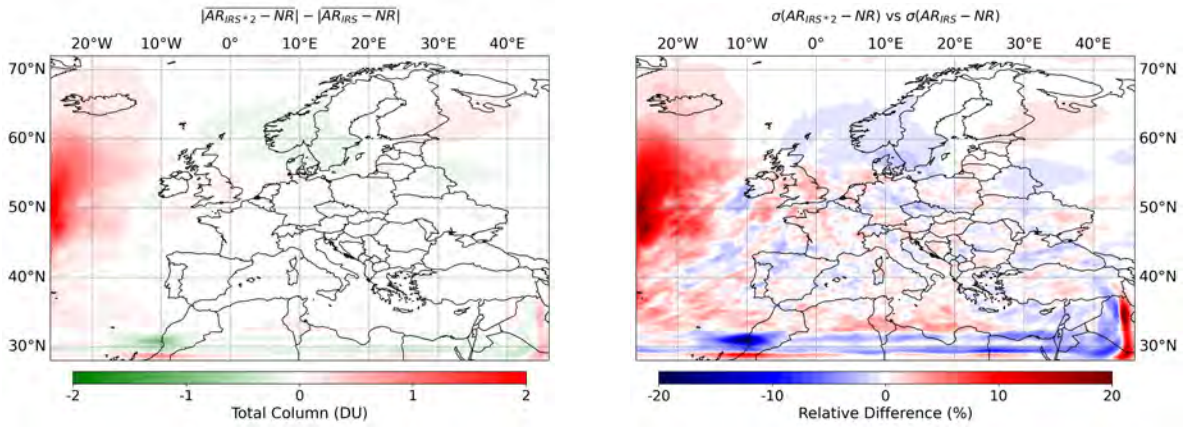
Figure 7.15: Statistics of the innovations (O-B) and the residuals (O-A) computed over the period of study (1st of June - 31st of August, 2019) for both AR of IRS and IRS*2 using $\mathbf{R}_{\text{full}}$. Averages are shown in **(a)**, while standard deviations are in **(b)**.

7.3.2 Evaluation of the Assimilation

A verification of the impact of IRS*2 assimilation on the O_x field retrieved by MOCAGE has been carried out against the NR and then compared to what was obtained for IRS. The differences occurring between the two cases will be displayed in the following.

7.3.2.1 Total Columns

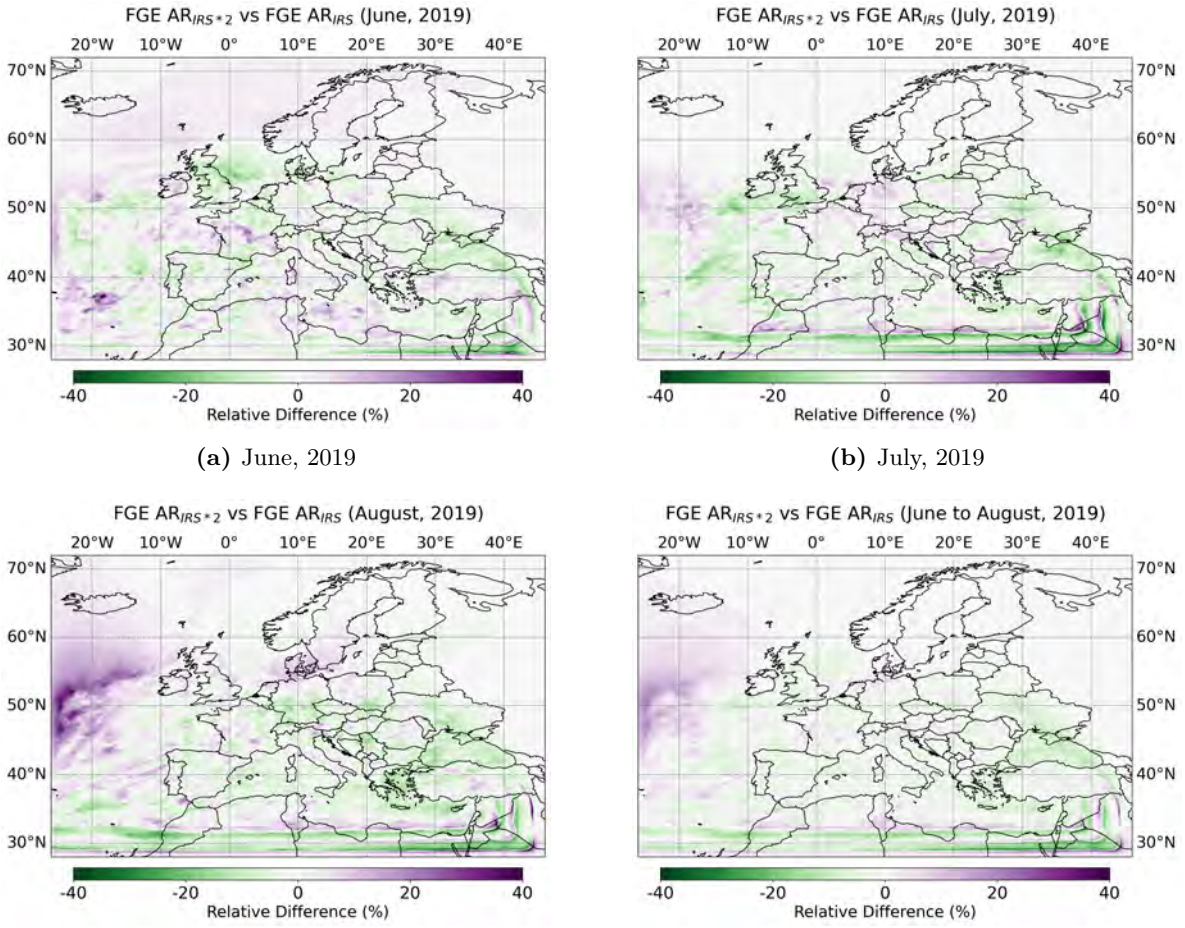
Figure 7.16 summarises the deviation of the ozone field, in total columns, provided by each ARs with respect to the NR. Panel 7.16a shows the difference between absolute values of these average bias, over the whole 3-month period evaluated so far. The colour scale has been chosen so that green shaded areas indicate where the AR of IRS*2 radiances is closer to the NR than the AR of IRS. Red shades, on the other hand, indicate AR IRS closer to the NR than that of IRS*2 As it is easy to observe, in the area with the highest frequency of assimilated observations, the difference is located around zero (white or clear shades). Areas with slightly stronger values, even if only of the order of a few DU, are found around 50°N over the Atlantic Ocean (red spot), where IRS is closer to the NR than IRS*2, and on the lower edges of the domain, where, as already introduced, the effects of the coupling between domains is remarkable.



(a) **(b)**

Figure 7.16: Difference between the absolute values of the variations of the AR IRS*2 from the NR Total Columns, and those concerning AR IRS, over the period ranging from the 1st of June to the 31st of August, 2019 **(a)**. In **(b)** is the relative difference between the standard deviation of the differences occurring between each AR and the NR.

The relative difference (in %) between the standard deviations of each AR and NR, shown in Panel 7.16b, tells that the errors are also very close to each other in the area where the observations are assimilated. On the other hand, the areas presenting more extreme values (up to plus and minus 20%) are the same that had shown stronger values of difference between the averages. In other words, the zone where the AR of IRS observations is closest to the NR, on average, is the same where the AR of IRS*2 presents the strongest standard deviation



(a) June, 2019

(b) July, 2019

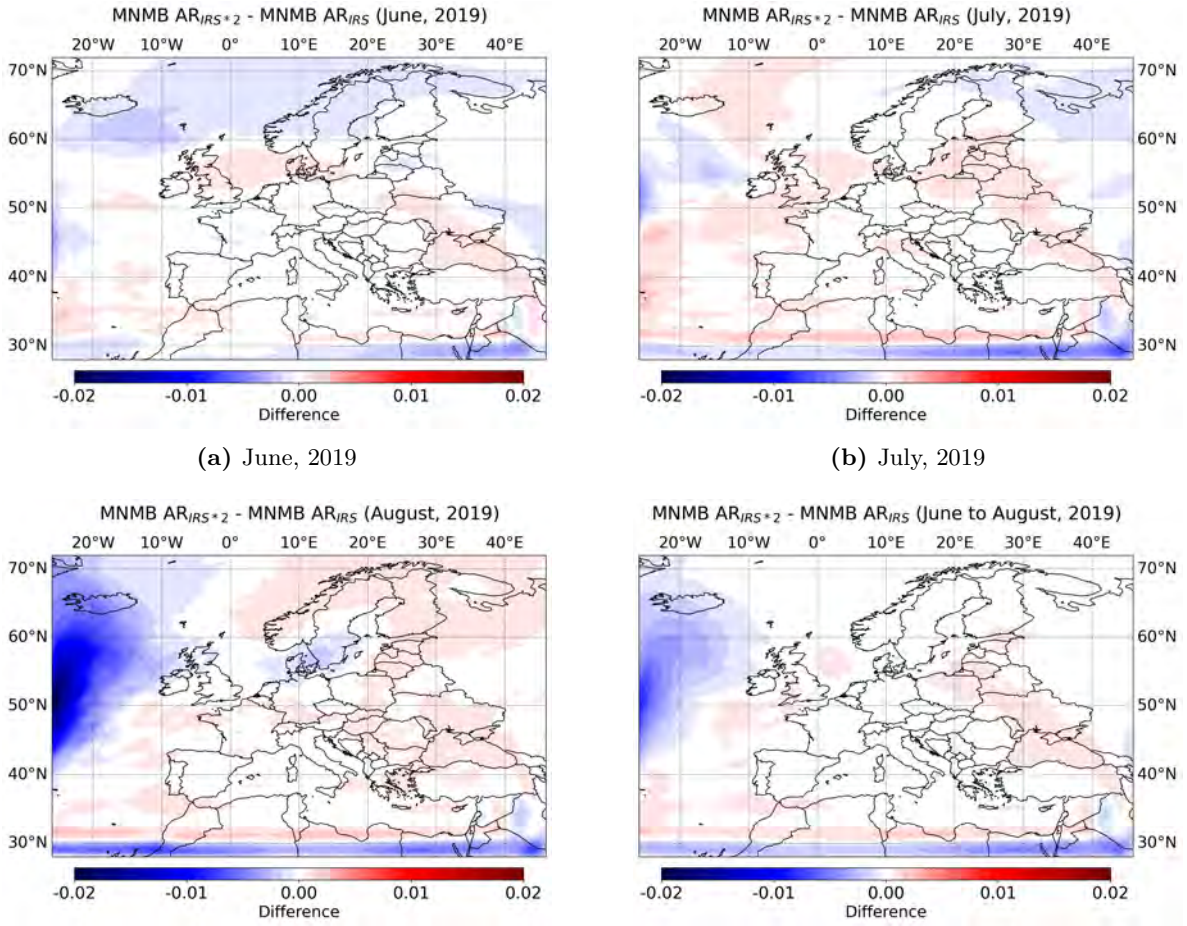
(c) August, 2019

(d) June to August, 2019

Figure 7.17: Relative differences between Fractional Gross Error (FGE) computed between AR_{IRS} and NR, and FGE between AR_{IRS*2} and NR. The results are shown for averages over the month of June **(a)**, July **(b)**, August 2019 **(c)** and over the 3 months **(d)**.

from the NR compared to the corresponding value obtained for the AR of IRS.

Figure 7.17 depicts the relative difference, in percentage, of the FGE obtained between AR of IRS*2 versus the NR and that of AR IRS versus NR (this second case is used as the term of comparison for the relative difference). The 3-month average, Panel 7.17d shows rather smooth values in the centre of the domain. The differences are of the order of a few percents. at the lower edge there are again several variations of the order of -20% between the FGEs for the two ARs. Positive values suggest higher FGEs for IRS*2 than for IRS with respect to the NR, negative values the opposite. These edge variations are less pronounced in the first month evaluated, i.e. June (7.17a), while they are more accentuated in July (7.17b) and August (7.17c). To go more into the specifics of what takes place over the different geographical areas showing non-zero percentages, each case should be examined individually (not done in this work). However, a few more cases can be mentioned more in detail. In June, values are found in the Atlantic Ocean off the Strait of Gibraltar that are up to 30% positive.



(a) June, 2019

(b) July, 2019

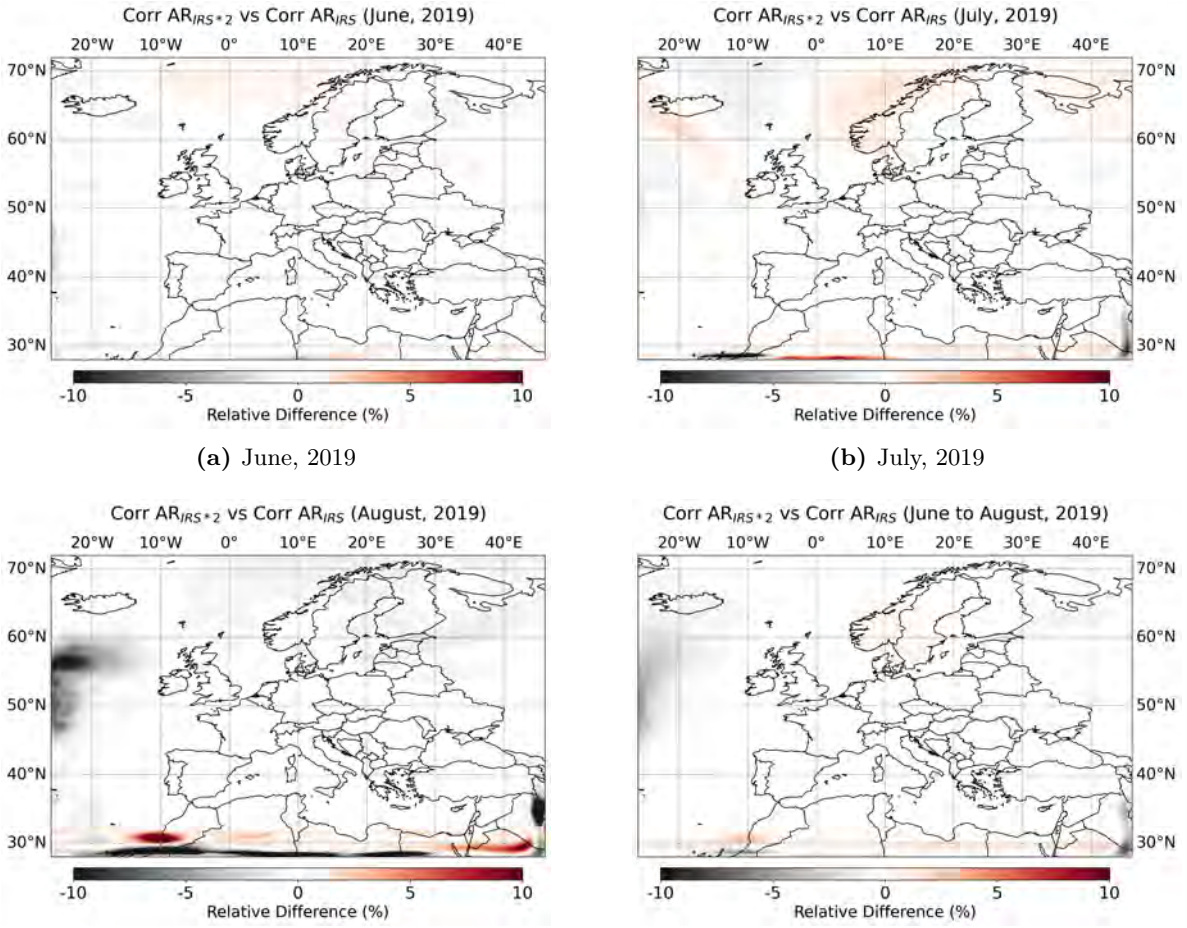
(c) August, 2019

(d) June to August, 2019

Figure 7.18: Differences between Modified Normalised Mean Bias (MNMB) occurring between AR_{IRS*2} and NR, and MNMB between AR_{IRS} and NR. The results are shown for averages over the month of June **(a)**, July **(b)**, August 2019 **(c)** and over the 3 months **(d)**.

Over the same area in July, values close to zero, or slightly positive, are found. For the month of August, both positive and negative values are found, but of a slightly lower order than for June (around $10 - 15\%$). Above Greece, in August, IRS*2 AR seems to have a larger error with respect to the NR than IRS does. These variations, as well as the ones that have not been pointed out, are most likely driven by the different density of observations involving the month and the case in question (see Figure 6.14 for a better comparison). A large area on the Atlantic Ocean (around 50°N and 50°W), in the month of August, shows positive values of the order above 30%. However this is in an area where few or no observations are assimilated and therefore, it does not enter into the evaluation of the impact of ARs.

Differences in terms of MNMB (of each AR vs the NR) are shown in Figure7.18 (blue stronger values of MNMB of AR IRS*2, red stronger values for MNMB of AR IRS). The differences in bias increase with time, but they are mostly found outside the area where both IRS and IRS*2 observations are located (please remember that the same number of equally



(a) June, 2019

(b) July, 2019

(c) August, 2019

(d) June to August, 2019

Figure 7.19: Relative differences between Correlations occurring between AR_{IRS} and NR, and Correlations between AR_{IRS*2} and NR. The results are shown for averages over the month of June **(a)**, July **(b)**, August 2019 **(c)** and over the 3 months **(d)**.

geo-localised observations are assimilated for both IRS versions), as well as, again, at the lower edge of the domain. The maximum variation values are reached in August with peaks of -0.02 above the Atlantic Ocean at 50°N . However, considering that the MNMB can range between -2 and 2, these extreme values are still quite small (a difference of 0.02 corresponds to 1% of variation). On average over the 3 months, values are very low, around zero, over the entire Mediterranean basin and continental Europe.

Looking at the differences between the Correlations, shown in Figure 7.19, these are, again, located around zero in the central part of the domain. Furthermore, as for the previous indicators, the highest values are encountered in August at the bottom border of the domain (both positive and negative, but of 10% magnitude), and over the Atlantic between 50 and 60°N.

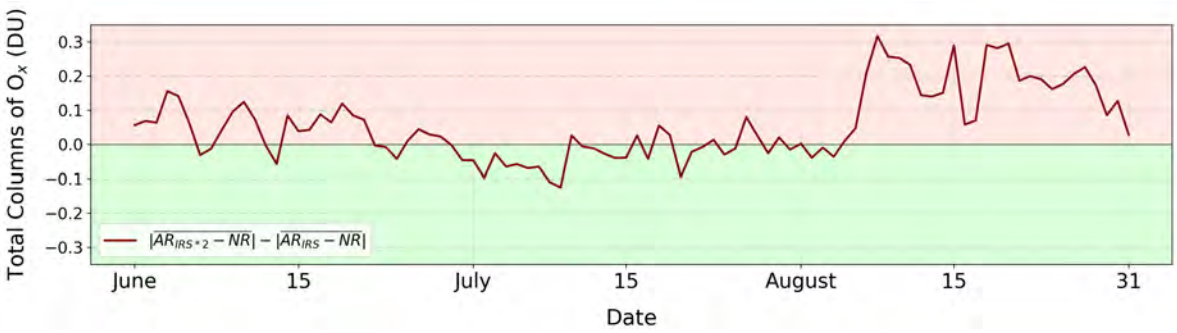
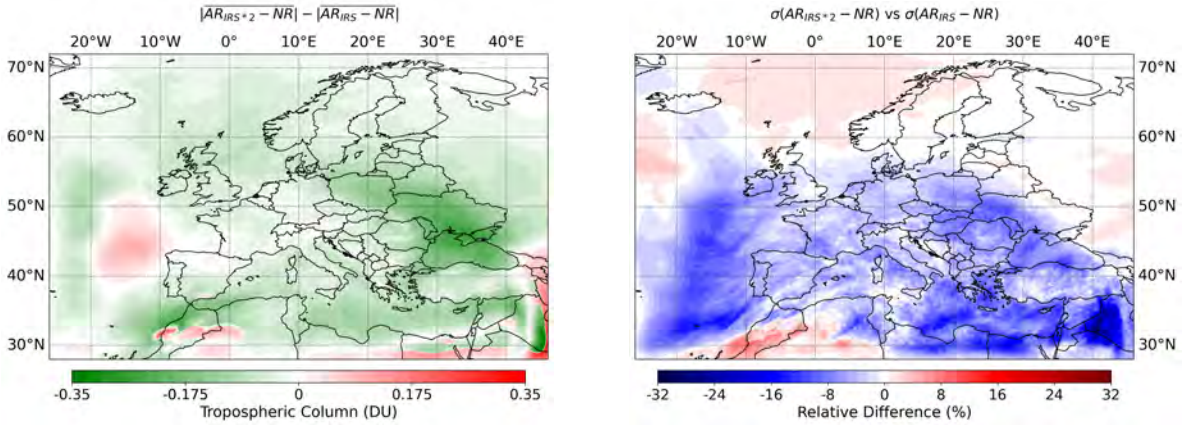


Figure 7.20: Difference between the biases of each AR with respect to the NR O_x Total Columns. Values of the biases are taken on average over the MACC01 domain, per day, and in absolute value: $|\overline{\mathrm{AR_{IRS*2} - NR}}| - |\overline{\mathrm{AR_{IRS} - NR}}|$. The green area correspond to values of AR IRS*2 closer to the NR than the values of AR IRS. The background area in red indicates the opposite.

A measure of which AR provides O_x total columns that most closely resemble those of the NR, can be assessed in terms of time series of differences between the absolute values of the biases with respect to the NR, averaged over the entire domain. Results are shown in Figure 7.20. The green area, in the background of the diagram, suggests that negative values correspond to values of AR of IRS*2 closer to the NR than the values of AR IRS. The background area in red, in contrast, corresponds to values of AR IRS closer to NR than AR IRS*2. A majority of positive values is observed during the months of June and August. The stronger values observed during the month of August are coherent with what had been found when examining the MNMB (Figure 7.18). In that case, stronger differences in August, and on the left edge of the domain, had already been observed. When averaging over the whole domain, those values inevitably lead to daily values of bias differences, that are stronger than in previous months. It can be concluded that, for the months of June and August, the Ox total columns obtained by AR of IRS observations are closer to the NR than AR IRS*2. For the month of July the trend is reversed, but the value are smaller compared to those of the other two months.

7.3.2.2 Tropospheric Columns

A comparison between the O_x tropospheric columns provided by AR of IRS and AR of IRS*2 radiations is also carried out at this stage. Similarly to what has been shown for the total columns in Figure 7.16, the differences between the absolute values of the average biases compared to the NR are shown in Panel 7.21a. As it is immediately evident when observing the chart, the tropospheric O_x columns obtained when assimilating IRS*2 radiances into MOCAGE are closer to the NR than the AR of IRS (green shades). Areas where the reverse trend is observed are however present, and they are localised over the Atlantic Ocean (between 40 and 50°N) and at the South of the domain, i.e. a spot over Morocco, plus the usual influence of coupling at the bottom right corner of the domain.



(a) **(b)**

Figure 7.21: Difference between the absolute values of the variations of the AR IRS*2 from the NR Tropospheric Columns, and those concerning AR IRS, over the period ranging from the 1st of June to the 31st of August, 2019 **(a)**. In **(b)** is the relative difference between the standard deviation of the differences occurring between each AR and the NR (Tropospheric Columns).

The relative difference between the standard deviations of each AR from the NR, are, instead, depicted in Panel 7.21b. These also show a better performance (i.e. closer to the NR) of the AR of IRS*2. Blue-shaded values indicate that the error associated with the bias of the IRS*2 AR compared to the NR are lower, in percentage terms, than those associated with the bias of the IRS AR. Variations of the order of 15% are found over the Atlantic Ocean, and, even stronger, over the Middle East Area, up to 20%. Still negative values, somewhat less strong, are distributed over the entire area of assimilation of simulated radiances. Over Marocco and Algeria, however, there areas where the sigma associated with AR IRS are smaller than those of IRS*2 (red values) are encountered too.

If, as already done for the total columns in Figure 7.20, the deviation of each AR run from the NR, averaged over the entire domain and for each day of the time-period examined, are computed, the time series in Figure 7.22 is obtained. The results are very interesting because, compared to what was seen when taking into account the total columns, when only tropospheric layers are taken into account, the AR of IRS*2 is the one that comes the closest

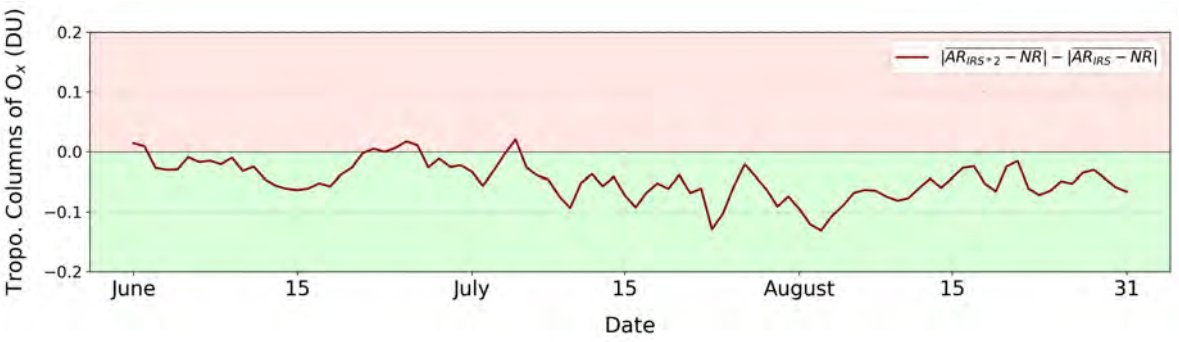


Figure 7.22: Difference between the biases of each AR with respect to the NR O_x Tropospheric Columns. Values of the biases are taken on average over the MACC01 domain, per day, and in absolute value: $|\overline{AR_{IRS*2} - NR}| - |\overline{AR_{IRS} - NR}|$. The green area correspond to values of AR IRS*2 closer to the NR than the values of AR IRS. The background area in red indicates the opposite.

to the NR during the whole period of study. Only a few days in June and July show a better performance of AR IRS (red highlighted area), although with values still very close to zero. During the rest of time-frame values remain always negative (green area) showing a better performance of the IRS*2 assimilation than the IRS one.

7.3.2.3 Vertical Variations

When examining the ozone concentrations by vertical level, instead of atmospheric columns, other considerations can be drawn on the different restitutions of the two ARs.

Figure 7.23 shows the differences between the variations (in absolute values) of the O_x mole fraction from the AR using IRS*2 compared to the NR, and the variations of the AR for IRS and the NR, averaged over longitudes. Panels 7.23c 7.23e are respectively the zoom on the levels 40 to 60 (i.e. approximately the tropospheric column), and from 45 to 60, using an appropriate colour scale for the smaller variations occurring in these layers. Please notice that green colours mean AR of IRS*2 closer to the NR than AR IRS, and red vice versa. Up to 36° latitude is where the greatest values of variation are found in the layer form 30 to 5, with the two ARs providing better results in turn. However, when looking at the rest of the domain (up to almost 60°N, that is where no observations are anymore present), larger green areas are encountered. In detail, at levels around 25 (lat. 40° to 56°N) and 15 (from 52°N) there is a better estimation of the O_x mole fraction through AR IRS*2 then through AR IRS. In the lower levels this is even more evident. Up to model level 55, and from level 50 to 41, there is a predominance of green all along the latitude of interest. While from 54 to 46 IRS assimilation seems to bring a closest result to the reality of the NR. Again, latitudes from 60° do not enter in the discussion since no observations are assimilated there.

The standard deviations of the biases between each assimilation run and the NR are shown in the Panels 7.23b, 7.23d and 7.23f. If excluding from the evaluation the latitudes where there coupling effects between the global and regional domains occur (up to 36°) and those where no observations are simulated, and then assimilated, lower error values of the difference with

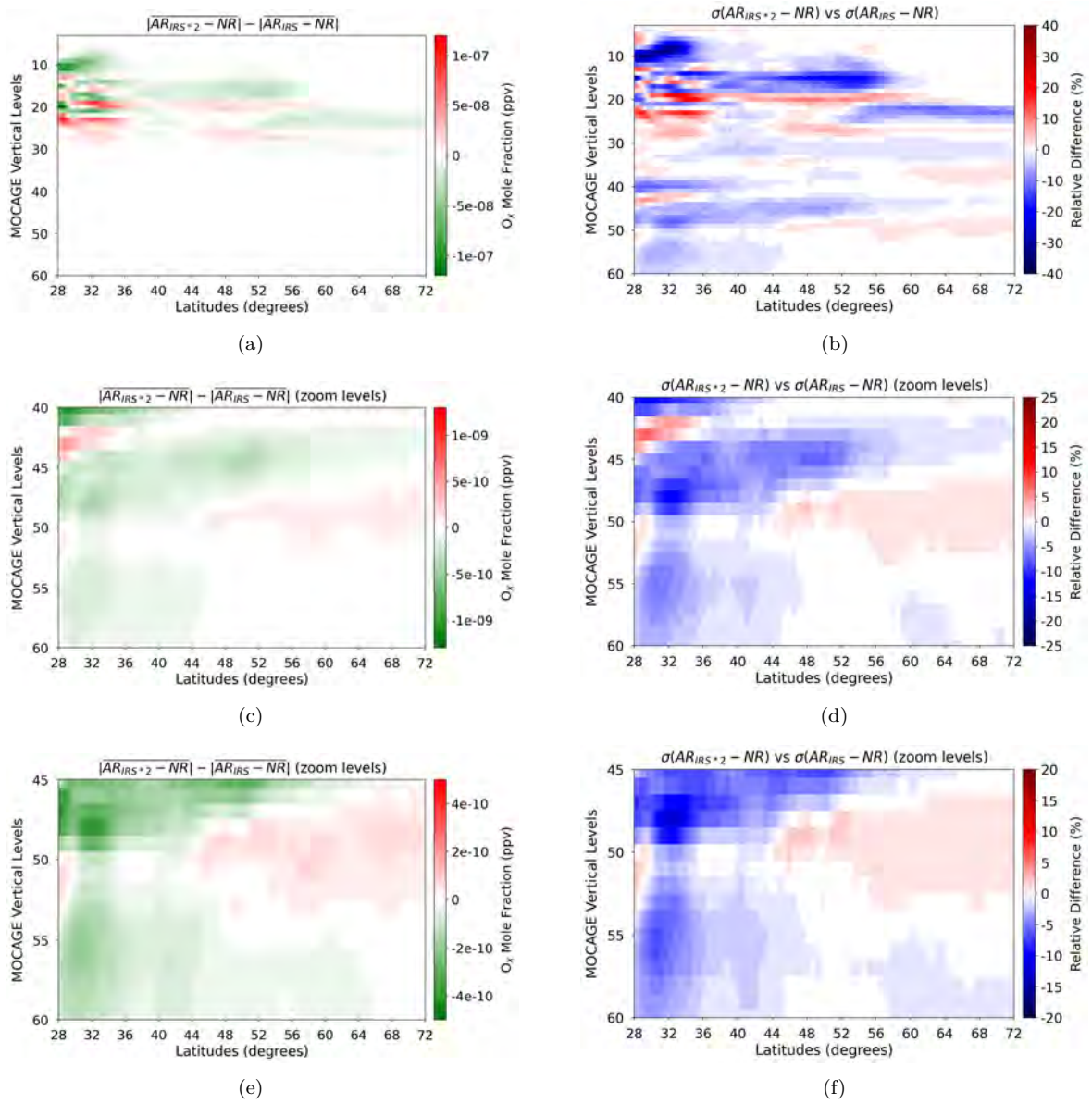
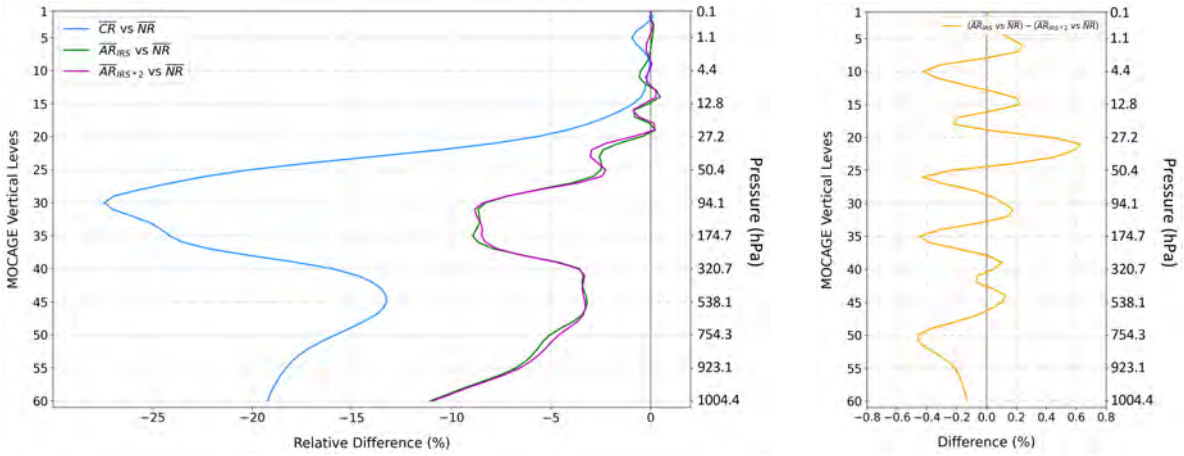


Figure 7.23: Difference between the absolute values of the variation of the AR IRS*2 from the NR O_x Mole Fraction (averaged over longitudes of the full domain), and those concerning AR IRS (over the period ranging from the 1st of June to the 31st of August, 2019) **(a)**. The corresponding zooms from model level 40 to 60 and from 45 to 60 are in Panels **(c)** and **(e)**. The relative difference between the errors of the differences (AR_{IRS*2} - NR) and (AR_{IRS} - NR) are in **(b)**, **(d)** and **(f)**. For equivalence between model and pressure levels, see Figure 7.24 or refer to Appendix A.

respect to the NR are observed for what concerns the AR of IRS*2 (predominance of blue colours). In the lower levels, similarly, that is where AR IRS*2 was the closest to the NR in terms of averages, is where the lowest relative difference error of the bias compared to the NR are found in the IRS case.

By averaging the ozone mole fractions also over both latitudes and longitudes, further conclusions can be drawn. The relative differences in percentage between CR and NR (blue),



(a) **(b)**

Figure 7.24: In Panel **(a)** are the relative differences between the averages of O_x Mole Fraction from CR, AR for IRS (AR_{IRS}), AR for IRS*2 (AR_{IRS*2}) vs the NR. The averages are computed, for each MOCAGE vertical level, over the 3-month period of study (1st of June up to 31st of August, 2019) and the geographical MACC01 domain. In **(b)** is an highlight of the deviation between the relative differences of each $\overline{AR}$ from the $\overline{NR}$.

AR for IRS and NR (green), AR for IRS*2 and NR (magenta) are depicted in Figure 7.24a. The first thing one observes is that the three runs being evaluated almost always provide values of O_x that are weaker than NR ones. This is not the case, however, for a few model levels, like 20 or 14, where assimilating IRS or IRS*2 simulated radiances brings to positive, even if very low, relative difference percentages. The deviation of the CR with respect to the NR averaged values is noticeably stronger than when assimilating IRS or IRS*2 simulated radiances into MOCAGE. More than 15% deviation around level 30 is found when no assimilation is performed compared to when it is. This means that the assimilation has an overall good impact over the model restitution, since it brings the difference compared to the NR reality closer to zero than when not performing any assimilation. The assimilation runs reveal much smaller variations.

In order to better assess the variations occurring between the ARs, a difference between the two relative difference curves is shown in Figure 7.24b. Non-zero differences between the contributions of the two sets of radiances to the model are retrieved. The divergences depicted show both positive and negative values, depending on the vertical level. More intense values, are, in general, found in the negative range. This means that IRS*2 AR brings to values that are closer to the NR than IRS does. This is especially true at levels like number 50, 35, 26 and 10. On the other hand, IRS seems to go closer to the NR values at the levels around 21, where it the computed difference reaches a maximum of 0.7%.

7.4 Highlights of this Chapter

- The assimilation of the two sets of synthetic observations, one for IRS and the other for IRS*2, has been carried out separately within the CR. In the two cases the same methodology and settings have been used. Background errors were derived as a percentage of the ozone profile (2% over the entire column). Observation errors, instead, have been obtained by means of the so-called *Desroziers diagnostics*.

  Observations minus Background (O-B), i.e. innovations, and *Observations minus Analysis* (O-A), i.e. residuals, have been computed as averages for each hour in the day (to highlight the diurnal cycle), as time-series and averaged over the whole period of study for each simulated wavelength. An assessment of these results shows analysis that always gets closer to the observations than the background.

- At first, the contribution of IRS synthetic radiance assimilation has been assessed.

  → A study on the total columns of ozone averaged over the whole 3-month period considered for the evaluation (June to August, 2019), shows that the assimilation of IRS radiances into the MOCAGE model always has a positive impact compared to no assimilation (CR). The evaluated total columns provided by the AR deviates, on average, very little (mostly $\sim 0\%$, few areas $\sim 3\%$) from the NR, where synthetic observations are more frequently assimilated. This is a significant improvement over the CR, which instead differed from the NR up to 13%. Errors are spatially consistent and of comparable magnitude. Standard deviation of the bias existing between AR of IRS with respect to the NR, has its smallest values in the area where IRS radiances are assimilated ($\sim 1.8\,\mathrm{DU}$).

  → An examination of statistical indicators computed for the total columns over the different months, and over the entire period of study, shows an important error reduction when assimilating IRS radiances. The FGE, indeed, is, always very close to zero over the entire assimilation area and over the time period. Recalling that the FGE values between CR and NR had reached peaks of about 0.2, there is, with the assimilation, an error reduction.

  → Extreme values were encountered, during the evaluation, at the edges of the domain, due to the coupling effect between the geographical domains global and regional.

  → Evaluating tropospheric columns, slightly stronger variations compared to the case of the total columns have been estimated, with values closer to 3% than to 0%. When computing averages per day, and over the whole domain, peaks of 8% are encountered on a few days. However, these are far from the the 20% variation encountered when evaluating the same parameter between CR and NR.

→ Assessing, instead of atmospheric columns, the mole fraction by vertical levels of the model, structures and intensities very similar to the considered NR reality are encountered. Variations of the order of $5-10\%$ (AR stronger values than the NR) are found in the model levels between 30 and 20, at latitudes from 28° to 36°. In the same location, the standard deviations of the bias present their maxima. This area, however, corresponds to the edge of the domain where anomalous values occur, produced by the coupling effect. If this area is excluded from the evaluation, the variation occurring is, on average, around 8% (NR values stronger than AR), with minima very close to zero on many vertical levels, and maxima approaching 15% in the lower-most troposphere between 48° and 60° of latitude.

Comparing with what had been obtained for the CR, where vertical variations were much more evident (up to the 25%), the impact of assimilation is considerable on the whole vertical column.

- In a second time the different impact of the two IRS versions was examined.

  → When evaluating the total columns averaged over 3 months, the differences of each AR with respect to the NR are very close to each other in the disc in which observations are assimilated (difference close to zero). In the same area, the associated standard deviation values are comparable, with differences of only a few percentage units. Coupling effects at the edges of the domain are also visible in the bottom border. On the left edge over the Atlantic Ocean, instead, between 45° and 60°N, the bias compared to the NR is stronger for the AR of IRS*2.

  → The differences between the statistical parameters FGE, MNMB and Correlation, which have been calculated between each AR and the NR, are very close to zero over the entire assimilation area. Consequently, these parameters are comparable for the two runs when studying the total columns.

  → In terms of time series, the values of total columns the closest to the NR are those produced by the AR of IRS for the months of June and August. During the month of July the total columns issued by the AR of IRS*2 seems to be closer to the reality compared to IRS. Differences between the biases remain rather small when compared to typical values of total columns (-0.1 to 0.3 DU for total columns of the order of ~ 300 DU). The slightly stronger values observed in August are coherent with what was found for MNMB, which shows stronger differences in August at the left edge of the domain. When averaging over the whole domain, those values lead to daily values of bias differences that are stronger than in the two previous months.

  → An investigation of the tropospheric column, shows different results. The AR of IRS*2 observations is, among the two, the closest to the NR. In this case, hence, IRS*2 shows a better contribution in assimilation compared to IRS. The values of

the bias differences remain small (maximum 0.35 DU), even considering the small tropospheric ozone concentration.

→ Assessing the tropospheric values averaged over the domain, day by day, a better contribution of IRS*2 radiances is found over almost the entire period. The difference are of the same order as those found for the total columns (maximum 0.15 DU), despite of the fact that the tropospheric concentration is lower.

→ An examination of the mole fraction, level by level, reveals that the difference in the impact of the two IRSs depends on the atmospheric level. From level 10 to 25, at very low latitudes, the widest variations between the bias of each AR compared to the NR are found. It is known, however, that these are caused by the coupling effects between the domains and, therefore, they will not be included in the discussion. Between latitudes of about 36° up to about 60° (beyond which no more observations are simulated), different behaviours of the two ARs are observed. At very high levels (from about 15 up to 25), the two IRSs alternate in giving the best performance (variations between biases ($\sim 10^{-7}$ ppv magnitude)), with a weak predominance of IRS*2 values that are closer to the NR than IRS. Below level 30 (i.e. ~ 90 hPa) IRS*2 turns out to be almost always the closest to the NR, except for an area between levels 47 and 53 (around 600 and 860 hPa). At these levels variations reach maxima of 10^{-9} ppv. In the lower-most troposphere, IRS*2 is always closer to the NR than IRS (at these levels variations reach maxima of 10^{-10} ppv.).

Conclusions and Perspectives

• Reminder of the Objectives

The present PhD project took place in the very actual context of preparing the arrival of the new Infrared Sounder (IRS), that will fly aboard the Meteosat Third Generation (MTG)-S satellites in the next few years. As already discussed in detail within the manuscript, such an instrument will show a considerable potential for application in the sphere of atmospheric chemistry monitoring and forecasting. It will cover two bands in the LWIR and MWIR, with a remarkable spectral sampling (about $0.603\,\mathrm{cm}^{-1}$ and $0.604\,\mathrm{cm}^{-1}$, respectively), combined with a high frequency of spectra acquisition (every 30 minutes over Europe) of an entire Earth disc from a geostationary platform.

The purpose of the work was, therefore, to assess the contribution that the assimilation of radiances that IRS will acquire, could bring to the characterisation of the atmospheric chemical composition (focus on the ozone) over Europe when assimilated into a chemistry transport model (CTM) such as MOCAGE.

At the same time, a version of the same IRS instrument, but with an almost doubled spectral sampling (about $0.325\,\mathrm{cm}^{-1}$), is already being considered. The second objective of the thesis was, then, to evaluate the possible different contributions of this IRS version, called, throughout the manuscript, IRS*2.

• OSSE Framework Creation and Evaluation

In order to fulfil these purposes, realistic data for both versions of the instrument had to be simulated. The method chosen was the Observing System Simulation Experiment (OSSE). A detailed study was carried out in order to build a robust OSSE framework over the European domain. For both the Nature Run (NR - i.e. the reference reality) and the Control Run (CR), the MOCAGE model was chosen. This provides a global (GLOB11) and regional (MACC01) domain configuration. To have the regional domain, i.e. the one of interest for the project, the global has to run too, since MOCAGE is a nested grid chemistry transport model. The configuration with 60 vertical levels, up to 0.1 hPa, has been exploited. The time-period chosen for the evaluation of the simulations ranged between the 1st of June till the 31st of August, 2019.

Planning for the construction of a realistic NR to be different enough from the CR, but

consistent with it, required a significant research effort. For both runs, the MOCAGE model, forced by the same meteorological input from the operational ARPEGE, has been used. Therefore, a strategy had to be chosen to differentiate the runs. First, different emission configurations, referring to different years, have been set for the NR and CR frameworks. Plus, another, more unconventional, choice was made: L1c radiances from the IASI instrument were assimilated within the global MOCAGE configuration in the NR framework. The MACC01 domain was then used as NR, which is forced at the edges by the global domain thus obtained, but within which, in turn, no assimilation was performed.

The topic of L1 data assimilation into a chemistry transport model, however, is a topic still being refined. A complementary study has, then, been performed with the aim of exploring the procedure of assimilation, to be exploited in a second stage also for IRS radiance assimilation, and to compare the NR framework to real observations (namely OMI and ozonesondes). This work provided interesting information about the possibility of improving the background errors evaluation and to, possibly, reduce the number of assimilated channels in future works. To support the literature, the results of this study have been submitted to the *Journal of Geophysical Research: Atmospheres*.

A study of inter-comparison has, then, been performed between NR and CR (which we remind to be the simulations on the regional MACC01 MOCAGE domain). The noticeable differences that emerged have been estimated good enough to avoid the identical twin problem. The O_x total columns averaged over the time-frame considered for the evaluation (from 1st of June to 31st of August, 2019), showed a CR producing always lower values than the NR, with a relative difference between 10.5 and 13.5%. The spatial distribution of the values was consistent for the two runs, as were the errors. A study on a time series of the same parameter found larger variations at the end of the period (up to 14% at the end of August). Assessing the tropospheric columns, even larger variations have been observed between the two evaluated runs. The maxima reached up to 20% on certain days (like in the end of July). Finally, distribution of ozone concentrations along the vertical levels was found to present a maximum variation of about 35%, located between approximately 25 and 180 hPa, and between 10 and 20% in the lower-most troposphere.

• Simulation and Assimilation of the Observations

From the NR, IRS observations have been simulated, using RTTOV v12, for the 195 contiguous wavelengths bounded in 982.464 and 1099.467 cm^{-1}, i.e. the range that had been confirmed by *a priori* sensitivity studies to be the most sensitive to the ozone species and, then, the most suitable for the purposes of this work. Similarly, for IRS*2, observations have been simulated in the corresponding band 982.343 − 1099.379 cm^{-1} (361 contiguous channels). Only clear-sky observations have been simulated (cloud filter applied through information issued from the meteorological forcing). A horizontal thinning of the observation was also applied in order to make the ensuing assimilation reliable and efficient in terms of computational time (one

pixel simulated per box of 0.4°). These actions have produced a good observation density and frequency along the time period of interest.

Perfect observations thus created, have been perturbed with the addition of an instrument noise specific for each version of the instrument.

The two sets of synthetic observations, one for each version of IRS being tested, have been separately assimilated within the CR. In the two cases the same methodology and settings have been exploited. The assimilation algorithm applied was a 3D-Var with an hourly assimilation window. Background errors have been derived as a percentage of the ozone profile (2% over the entire column). Observation errors, instead, have been obtained by means of the so-called *Desroziers diagnostics*.

Statistics of residuals and innovations have been computed and verified, and found to be correct, i.e. analysis closer to observations than the background. Furthermore, higher values have been found at the edges of the domain than in the inner area, in link with the coupling information coming from the lateral boundary conditions.

• Impact of the Assimilation of IRS Radiances into MOCAGE

At first, the contribution of IRS synthetic radiance assimilation has been assessed.

A study on the total columns of ozone averaged over the whole 3 months, from June to August, showed that **the assimilation of IRS radiances into the MOCAGE model always has a positive impact compared to no assimilation (i.e. the CR)**. The evaluated total columns averages, obtained from the AR of IRS, deviated from NR very little where synthetic observations are more frequently assimilated. Values were mostly close to $\sim 0\%$, with just a few areas touching $\sim 3\%$. This has represented a significant improvement over the CR case, which, instead, deviated from the NR up to $\sim 13\%$. Errors resulted spatially consistent and of comparable magnitude. Standard deviation of the bias existing between AR of IRS with respect to the NR, had its smallest values in the area where IRS radiances have been assimilated ($\sim 1.8\,\mathrm{DU}$).
Statistical indicators computed for the total columns over the different months, and over the entire period of study, also depicted an error reduction when assimilating IRS radiances. The FGE between AR IRS and NR, indeed, was always approximately over the entire assimilation area and over the time period, while it took values around 0.2 when it was up to the CR.
High values were encountered, during the evaluation, at the edges of the domain, due to the coupling effect between the geographical domain global and regional.

When evaluating tropospheric columns, slightly stronger variations compared to the case of the total columns have been estimated, with values closer to 3% than to zero. The temporal series of the averages per day over the whole MACC01 domain showed peaks of 8% on a few days. However, these were far from the 20% variation encountered when evaluating the same parameter between CR and NR.

Assessing the concentrations by vertical levels of the model, structures and intensities very

similar to the considered NR reality have been witnessed. Variations of the order of $5-10\%$, with AR stronger than the NR, have been found at the levels between 30 and 20 (almost 25 to 36 hPa), at latitudes from 28° to 36°. In the same location, the standard deviations of the bias presented their maxima. This area corresponds to the edge of the domain where extreme values occur, produced by the effect of coupling GLOB11 and MACC01. If this area was excluded from the evaluation, the variation occurring would be, on average, around 8% (NR stronger than AR), with minima very close to zero on many vertical levels, and maxima approaching 15% in the lower-most troposphere between 48° and 60° of latitude. Values obtained for latitudes above 60°N should not be taken into account either in the evaluation, since no observation has been simulated (and neither assimilated) at those latitudes.
Comparing with what obtained for the CR, where vertical variations were much more evident (up to the 25%), the impact of assimilation is considerable on the whole vertical column.

• Different Contribution of the Two Versions of IRS

A study on the impact of IRS*2 assimilation on the O_x field produced by MOCAGE has been carried out against the NR and, then, compared to what was obtained for IRS.

When evaluating the total columns averaged over 3 months, the differences of each AR with respect to the NR have been found to be very close to each other in the disc in which observations were assimilated (difference close to zero). In the same area, the associated standard deviation values were comparable, with differences of only a few percentage units. The effect due to the coupling of the global and regional MOCAGE domains has been detected again at the South and East borders. In this second case, indeed, between 45° and 60°N, the bias to the NR was stronger for the AR of IRS*2.
The differences between the statistical parameters FGE, MNMB and Correlation, which have been calculated between each ARs and the NR, resulted very close to zero over the entire assimilation area. Consequently, these parameters have been considered comparable for the two runs when studying the total columns.
In terms of time series, the values of total columns the closest to the NR belonged to the AR of IRS radiances during the months of June and August, while during the month of July to the AR of IRS*2. Differences between the biases resulted rather small when compared to typical values of total columns (-0.1 to 0.3 DU for total columns of the order of ~ 300 DU). The stronger values observed during the month of August have been considered consistent with what was observed for the MNMB (between each AR and the NR). This parameter, whose differences between the two IRS versions have been evaluated on average over each month, had shown stronger differences in August at the left edge of the domain, above the Atlantic Ocean. When averaging over the whole domain, those values lead to stronger daily values of bias differences compared to the other two months.

An investigation of the tropospheric column, showed different results. The AR of IRS*2 observations was, among the two, the closest to the NR. In this case, hence, IRS*2 showed

a better contribution in assimilation compared to IRS. The values of the bias differences remained low (maximum 0.35 DU), even considering the low tropospheric ozone concentration. Assessing the tropospheric values averaged over the domain, day by day, a better contribution of IRS*2 radiances was found over almost the entire period. The difference values are of the same order as those found for the total columns (maximum 0.15 DU), despite of the fact that the tropospheric concentration is lower.

An assessment of the mole fraction, by MOCAGE vertical levels, revealed that the difference in the impact of the two IRSs depends on the atmospheric level. IRS*2, however, even if with very small values of variation between the biases compared to the NR, showed a general better performance compared to IRS, mostly in the lower-most atmospheric levels. From level 10 to 25, at very low latitudes, the widest variations occur, but they were caused by the coupling effects between the domains and, therefore, they have not been included in the discussion. Between latitudes of about 36° up to about 60° (beyond which no more observations were simulated), different behaviours of the two ARs have been observed. At very high levels (from about 15 up to 25), the two IRSs alternated in giving the best performance (variations between biases ($\sim 10^{-7}$ ppv magnitude)), with a slightly better contribution of IRS*2 compared to IRS. Below level 30 (i.e. ~ 90 hPa) IRS*2 had proven to be almost always the closest to the NR, except for an area between levels 47 and 53 (around 600 and 860 hPa). At these levels variations reached maxima of 10^{-9} ppv. In the lower-most troposphere, IRS*2 was always closer to the NR than IRS (maximum variation of 10^{-10} ppv).

Perspectives

From the results obtained and examined, many perspectives opened up for the present work.

▶ Improving the Assimilation Assessment

The first thing to do in the very short term is to improve the assessment of the assimilation. It has been stated several times that the impact of coupling between MOCAGE domains produces more variable values of innovations and, thus, of ozone field at the southern edge of the treated MACC01 regional domain. To ensure that the averages over the domain are not impacted by these values, the impact of assimilation should also be evaluated on a smaller area, that cuts off these edges. Similarly, when evaluating values as geographical averages, one could exclude from the evaluation those areas that are located outside the simulated, and thus assimilated, observation disc.

In second place, the specific cases of ozone pollution, which led to the choice of the period of study, have not been studied into detail due to a lack of time, but only through the evaluation of temporal series with averages over the domain. A more in-depth check on these cases is, hence, required. Specific scores over the most intensely affected area should be verified.

Finally, the study carried out around the NR preparation framework showed the advan-

tages of using a background error matrix computed by the NMC method instead of the matrix classically reported in the literature (percentage of ozone profile). This kind of approach can be beneficial for the assimilation of IRS radiances, both simulated and, when available, real. A three-month period may not be enough to apply such a method on a regional model.

▶ Improving the IRS Observation Processing

As previously mentioned, it is official that Principal Components (PCs) only will be distributed for IRS. The present study, however, did not take into account this subject. A short study evaluating the different impact of PCs on the work illustrated so far, should be carried out. Of course, this would be possible, at this stage, for the official version of IRS only (and not for IRS*2).

Another point that arose from the study on the preparation of the NR framework has been the possibility to perform a channel selection, while maintaining a good quality in the assimilation results. This is also a commonly used procedure in NWP for many instruments and that I myself worked on in the context of another project on the future instrument IASI-NG [Vittorioso et al. (2021)]. Since the work carried out here on IRS was a first analysis, we wanted to investigate the behaviour of all the consecutive wavelengths considered. Moreover, using the full spectral band was a fair way of comparing the impact of IRS and IRS*2. At a later stage, however, it will be of interest to carry out a selection of a smaller group of channels too.

▶ Impact of Synergy with Other Instruments

On board the MTG-S satellites, together with IRS, will also be placed Sentinel-4, an Ultraviolet Visible Near-Infrared (UVN) imaging spectrometer. This is designed to monitor some key air quality trace gases and aerosols over Europe with a high spatial resolution and fast revisit time, in support of CAMS. Given the potential of the joint acquisition of the two instruments, their synergy in the characterisation of atmospheric pollution over Europe should definitely be tested.

If we consider assimilating IRS in the global domain of MOCAGE, the joint assimilation of IRS with GIIRS (hyperspectral IR sounder on board chinese geostationary satellites) could lead to interesting validation.

At the same time, other spectrometers working in a similar way, though on polar-orbiting platforms, would also be interesting to study. First on the list are certainly IASI and the new generation, IASI-NG (as soon as its data are exploitable), on board European satellites. Other IR sounders of the US and Chinese polar systems (CrIS and HIRAS) could be added also.

The joint assimilation of IRS radiances and surface measurements could also be considered in the future. Since IRS provides good vertical information in stratosphere, UTLS and free troposphere, surface data can bring additional details at the lower-most atmospheric layers.

▶ Towards the Treatment of Other Species

Certainly to be addressed as a continuation of this work, there will be the evaluation of the assimilation of radiances in the bands sensitive to other chemical species and aerosols. The OSSE framework created for this study was designed to provide, with minor modifications the basics for the extension to other species than ozone.

The sulphur dioxide (SO_2) could be investigated since IRS has the potential to sense this species, with a sensitivity towards the end of MWIR band [Coopmann et al. (2022)]. More specific case study, however, should be taken into account for this compound, since it is associated to specific and punctual natural events.

The carbon monoxide (CO) is, however, going to be the first target, given the results encountered during the study performed to assess the sensitivity of IRS in its two versions.

During an internship completed by two Master1 SOAC[1] students (Julie Wendelin and Louis Ramu), and that I co-supervised, carbon monoxide began to be evaluated. The students focused on only one day in the period of study (29th of June, 2019). The study, has been carried out on subpart of the MACC01 domain divided in a regular grid of 1428 points. They considered as reality the NR produced in this PhD OSSE framework. The background state was obtained by adding to the real state a Gaussian noise centred on a standard deviation of the 10% its value. As observations, they did not use those simulated in the OSSE framework, but they needed to simulate standard observations for wavenumbers which are sensitive to CO, through the use of RTTOV, taking into account the instrumental noise of both IRS and IRS*2.

For each grid point, they evaluated the information content in terms of degrees of freedom for signal (DFS) (see Appendix C). From an evaluation of the results (Figure I), they found that IRS*2 seems to bring more information than IRS. This difference is more pronounced over land and regions with higher surface temperatures, i.e. where the emission of the surface is stronger. Anyhow the difference remains small.

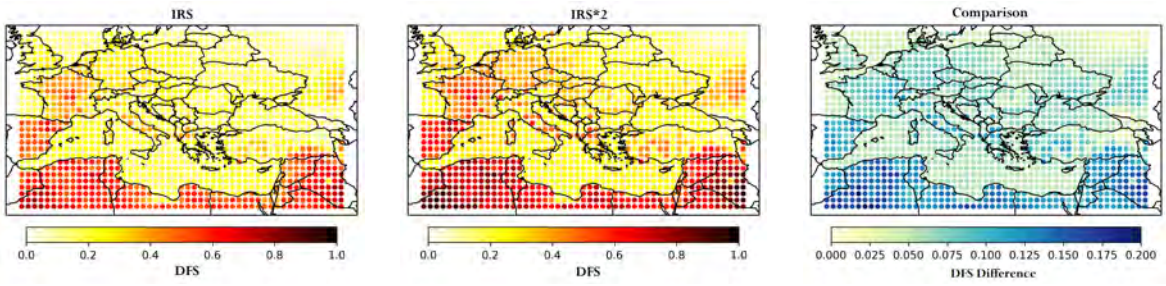


Figure I: DFS charts for IRS (left) and IRS*2 (middle) and comparison of DFS_{IRS*2} - DFS_{IRS} instruments (right).

[1]Sciences de l'Océan, de l'Atmosphère et du Climat (SOAC)

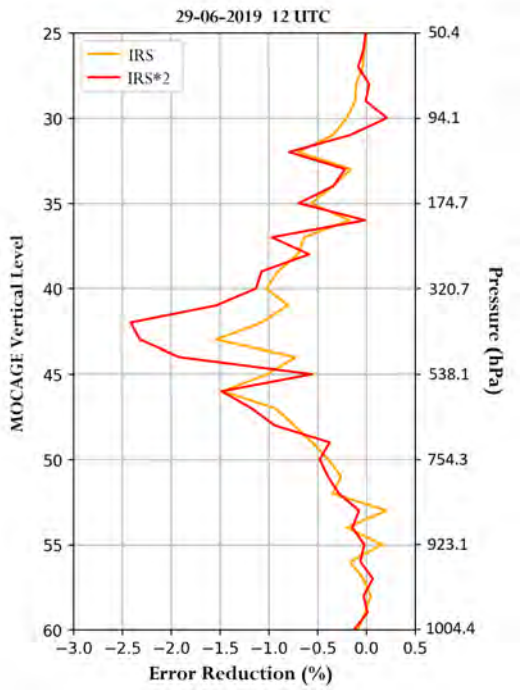


Figure II: Zoom on the standard deviations of the differences between background and analysis errors, for IRS at IRS*2.

Thanks to one-dimensional retrievals, they determined the analysis for each grid point, obtaining their retrieved CO profiles. They computed a statistical evaluation of the percentage relative difference in the standard deviation of the analysis error versus the standard deviation of the background error[2]. The result is shown in Figure II, for both IRS and IRS*2, as a zoom on the vertical column from model level 25 ($\sim 50\,\mathrm{hPa}$) to the surface. It is evident that at the levels between 40 and 45 (i.e. between $\sim$ 320 and 530 hPa), there is a reduction in the error which is 1.5% stronger for IRS*2 than for IRS.

These results are very promising and encourage to undertake as soon as possible the same kind of evaluation that has been done in this thesis for ozone, also for carbon monoxide.

[2] $[\sigma(\mathbf{x}_\mathrm{a} - \mathbf{x}_\mathrm{t}) - \sigma(\mathbf{x}_\mathrm{b} - \mathbf{x}_\mathrm{t})]/\sigma(\mathbf{x}_\mathrm{b} - \mathbf{x}_\mathrm{t})$

Conclusions et perspectives

• *Rappel des objectifs*

Ce projet de thèse s'est déroulé dans le contexte très actuel de la préparation de l'arrivée du nouveau sondeur infrarouge (IRS), qui volera à bord des satellites Meteosat Troisième Génération (MTG)-S dans les prochaines années. Comme déjà discuté en détail dans le manuscrit, cet instrument présentera un potentiel considérable d'application dans le domaine de la surveillance et de la prévision de la composition atmosphérique. Il couvrira deux bandes dans le LWIR et le MWIR, avec un échantillonnage spectral considérable (environ $0,603\ \mathrm{cm}^{-1}$ *et* $0,604\,\mathrm{cm}^{-1}$*, respectivement), combiné à une fréquence élevée d'acquisition de spectres (toutes les 30 minutes sur l'Europe) d'un disque terrestre entier depuis une plateforme géostationnaire.*

Le but de ce travail a donc été d'évaluer la contribution que l'assimilation des luminances que IRS va acquérir pourrait apporter à la caractérisation de la composition chimique de l'atmosphère (focus sur l'ozone) sur l'Europe une fois assimilées dans un modèle de chimie-transport tel que MOCAGE.

En même temps, une version du même instrument IRS, mais avec un échantillonnage spectral presque doublé (environ $0,325\ \mathrm{cm}^{-1}$*), est déjà envisagée. Le second objectif de la thèse a donc été d'évaluer les différentes contributions possibles de cette version d'IRS, appelée, tout au long du manuscrit, IRS*2.*

• *Création et évaluation du cadre d'OSSE*

Pour atteindre ces objectifs, il a fallu simuler des données réalistes pour les deux versions de l'instrument. La méthode choisie a été l'expérience de simulation du système d'observation (OSSE en anglais). Une étude détaillée a été réalisée afin de construire un cadre OSSE robuste sur le domaine européen. Pour le Nature Run (NR - c'est-à-dire la réalité de référence) et le Control Run (CR), le modèle choisi a été MOCAGE. Celui-ci fournit une configuration de domaine global (GLOB11) et régional (MACC01). Pour avoir le domaine régional, c'est-à-dire celui qui nous intéresse dans le cadre de ce projet, le global doit également doit être exécuté, puisque MOCAGE est un modèle de transport chimique imbriqué. La configuration avec 60 niveaux verticaux, montant jusqu'à $0,1\,\mathrm{hPa}$*, a été exploitée. La période choisie pour l'évaluation des simulations était comprise entre le 1er juin et le 31 août 2019.*

La construction d'un NR réaliste et suffisamment différent du CR, mais cohérent avec

celui-ci, a nécessité un effort de recherche important. Le modèle MOCAGE, forcé par les mêmes données météorologiques d'ARPEGE opérationnel, a été utilisé pour les deux séries de mesures. Par conséquent, une stratégie a dû être choisie pour différencier les séries. Tout d'abord, des configurations d'émissions se référant à des années différentes, ont été définies pour les cadres NR et CR. De plus, un autre choix, non conventionnel, a été fait : les luminances L1c de l'instrument IASI ont été assimilées dans la configuration globale MOCAGE dans le cadre NR. Le domaine MACC01 a alors été utilisé comme NR, qui est forcé aux limites par le domaine global ainsi obtenu, mais au sein duquel, à son tour, aucune assimilation n'a été effectuée.

Le sujet de l'assimilation des données L1 dans un modèle de chimie-transport est, cependant, un sujet encore en cours d'amélioration. Une étude complémentaire a donc été réalisée dans le but d'explorer la procédure d'assimilation des données IASI, à exploiter dans un second temps également pour l'assimilation des observations IRS, et de comparer le cadre NR aux observations réelles (à savoir OMI et ozonesondes). Ce travail a fourni des informations intéressantes sur la possibilité d'améliorer l'évaluation des erreurs d'ebauche et, éventuellement, de réduire le nombre de canaux assimilés dans les travaux futurs. Les résultats de cette étude ont été soumis au Journal of Geophysical Research : Atmospheres.

Une étude d'inter-comparaison a, ensuite, été réalisée entre NR et CR (que nous rappelons être les simulations sur le domaine régional MACC01 de MOCAGE). Les différences notables qui sont apparues ont été estimées suffisamment bonnes pour éviter le problème de l'identical twin. Les colonnes totales de O_x, moyennées sur la période considérée pour l'évaluation (du 1er juin au 31 août 2019), ont montré que le CR produit des valeurs toujours plus faibles que le NR, avec une différence relative comprise entre $10,5$ *et* $13,5$*. La distribution spatiale des valeurs était cohérente pour les deux run, de même que les erreurs. Une étude sur une série temporelle du même paramètre a trouvé des variations plus importantes en fin de période (jusqu'à* 14 *à la fin du mois d'août). En évaluant les colonnes troposphériques, des variations encore plus importantes ont été observées entre les deux séries évaluées. Les maxima ont atteint jusqu'à* 20% *certains jours (comme à la fin du mois de juillet). Enfin, on a constaté que la distribution des concentrations d'ozone le long des niveaux verticaux présentait une variation maximale d'environ* 35%*, située entre environ* 25 *et* 180 hPa*, et entre* 10 *et* 20% *dans la basse troposphère.*

- ***Simulation et assimilation des observations***

Á partir du NR, les observations d'IRS ont été simulées, à l'aide de RTTOV v12, pour les 195 *longueurs d'onde contiguës délimitées entre* $982,464$ *et* $1099,467\ \mathrm{cm}^{-1}$*, c'est-à-dire la gamme qui avait été confirmée par les études de sensibilité* a priori *comme étant la plus sensible aux espèces d'ozone et, donc, la plus appropriée pour les besoins de ce travail. De même, pour IRS*2, les observations ont été simulées dans la bande correspondante* $982,343 - 1099,379\ \mathrm{cm}^{-1}$ *(*361 *canaux contigus). Seules les observations en ciel clair ont été simulées*

(filtre de nuages appliqué grâce aux informations issues du forçage météorologique). Un "écrémage horizontal" des observations a également été appliqué afin de rendre l'assimilation qui en découle fiable et efficace en termes de temps de calcul (un pixel simulé par boîte de $0,4°$*). Ces actions ont permis d'obtenir une bonne densité et fréquence d'observation sur la période d'intérêt.*

Les observations parfaites ainsi créées ont été perturbées par l'ajout d'un bruit instrumental spécifique à chaque version de l'instrument.

Les deux ensembles d'observations synthétiques, un pour chaque version d'IRS testée, ont été assimilés séparément dans le CR. Dans les deux cas la même méthodologie et les mêmes paramètres ont été exploités. L'algorithme d'assimilation appliqué était un 3D-Var avec une fenêtre d'assimilation horaire. Les erreurs d'ébauche ont été dérivées en pourcentage du profil d'ozone (2% sur toute la colonne). Les erreurs d'observation, quant à elles, ont été obtenues au moyen de ce que l'on appelle le diagnostic de Desroziers*.*

Les statistiques des résidus et des innovations ont été calculées et vérifiées, et se sont avérées correctes, c'est-à-dire que l'analyse est toujours plus proche des observations que l'ébauche. Par ailleurs, des valeurs plus élevées ont été trouvées sur les bords du domaine que dans la zone intérieure, en lien avec les informations de couplage provenant des conditions aux limites latérales.

• *Impact de l'assimilation des luminances spectrales IRS dans MOCAGE*

Dans un premier temps, la contribution de l'assimilation des luminances synthétiques IRS a été évaluée.

Une étude sur les colonnes totales d'ozone, moyennées sur l'ensemble des trois mois de juin à août, a montré que ***l'assimilation des luminances IRS dans le modèle MOCAGE a toujours un impact positif par rapport à l'absence d'assimilation (c'est-à-dire le CR)****. Les moyennes totales des colonnes évaluées, obtenues à partir de l'AR d'IRS, s'écartaient très peu du NR dans les zones où les observations synthétiques sont plus fréquemment assimilées. Les valeurs étaient pour la plupart proches de* $\sim 0\%$*, avec seulement quelques zones touchant* $\sim 3\%$*. Cela a représenté une amélioration significative par rapport au cas CR, qui, au contraire, s'écartait du NR jusqu'à* $\sim 13\%$*. Les erreurs résultantes étaient spatialement cohérentes et d'une ampleur comparable. L'écart-type du biais existant entre l'AR d'IRS par rapport au NR, avait ses plus petites valeurs dans la zone où les luminances d'IRS ont été assimilées (*$\sim 1,8$ DU*).*

Les indicateurs statistiques calculés pour l'ensemble des colonnes totales sur les différents mois, et sur toute la période d'étude, ont également mis en évidence une réduction de l'erreur lors de l'assimilation des luminances IRS. Le Fractional Gross Error (FGE) entre l'AR IRS et le NR, en effet, valait toujours environ zéro sur toute la zone d'assimilation et sur toute la période, alors qu'il prenait des valeurs autour de $0,2$ *quand il s'agissait du CR.*

Des valeurs anormales ont été rencontrées, lors de l'évaluation, à la marge inférieure du

domaine, en raison de l'effet de couplage entre le domaine géographique global et régional.

Lors de l'évaluation de colonnes troposphériques, des variations légèrement plus fortes que dans le cas des colonnes totales ont été estimées, avec des valeurs plus proches de 3% *que de zéro. La série temporelle des moyennes par jour sur l'ensemble du domaine MACC01 a montré des pics de* 8% *sur quelques jours. Cependant, ces valeurs étaient loin des* 20% *de variation rencontrés lors de l'évaluation du même paramètre entre CR et NR.*

En évaluant les concentrations par niveaux verticaux du modèle, on a constaté des structures et des intensités très similaires à la réalité du NR considéré. Des variations de l'ordre de $5 - 10\%$*, avec un AR plus fort que le NR, ont été trouvées aux niveaux entre* 30 *et* 20 *(presque* 25 *à* 36 hPa*), à des latitudes de* 28° *à* 36°*. Au même endroit, les écarts types du biais ont présenté leurs maxima. Cette zone correspond au bord du domaine où se produisent des valeurs extrêmes, produites par l'effet du couplage de GLOB11 et MACC01. Si cette zone était exclue de l'évaluation, la variation observée serait, en moyenne, de l'ordre de* 8% *(NR plus fort que AR), avec des minima très proches de zéro sur de nombreux niveaux verticaux, et des maxima approchant* 15% *dans la basse troposphère entre* 48° *et* 60° *de latitude. Les valeurs obtenues pour les latitudes supérieures à* $60°N$ *ne doivent pas non plus être prises en compte dans l'évaluation, car aucune observation n'a été simulée (ni assimilée) à ces latitudes. Si l'on compare avec ce qui a été obtenu pour le CR, où les variations verticales étaient beaucoup plus marquées (jusqu'aux* 25%*), l'impact de l'assimilation est considérable sur toute la colonne verticale.*

• *Comparaison de la contribution des deux versions d'IRS*

*Une étude de l'impact de l'assimilation IRS*2 sur le champ* O_x *produit par MOCAGE a été réalisée par rapport au NR et, ensuite, comparée à ce qui avait été obtenu pour IRS.*

Lors de l'évaluation des colonnes totales moyennées sur trois mois, les différences de chaque AR par rapport au NR se sont avérées très proches les unes des autres dans la partir du disque où les observations ont été assimilées (différence proche de zéro). Dans la même zone, les valeurs d'écart-type associées étaient comparables, avec des différences de quelques unités de pourcentage seulement. L'effet dû au couplage des domaines global et régional de MOCAGE a été détecté à nouveau au bord Sud-Est. Dans ce second cas, en effet, entre 45° *et* $60°N$*, le biais par rapport au NR était plus fort pour l'AR de IRS*2.*

Les différences entre les paramètres statistiques FGE, MNMB et Corrélation, qui ont été calculés entre chaque AR et le NR, ont été très proches de zéro sur toute la zone d'assimilation. Par conséquent, ces paramètres ont été considérés comme comparables pour les deux séries lors de l'étude des colonnes totales.

*En termes de séries temporelles, les valeurs des colonnes totales qui étaient les plus proches du NR appartenaient à l'AR des luminances IRS pendant les mois de juin et août, tandis que pendant le mois de juillet elles étaient issues de l'AR d'IRS*2. Les différences entre les biais ont été plutôt faibles par rapport aux valeurs typiques des colonnes totales (*$-0,1$ *à* $0,3$ DU

pour des colonnes totales de l'ordre de $\sim 300\,\mathrm{DU}$*). Les valeurs plus fortes observées au cours du mois d'août ont été considérées comme cohérentes avec ce qui a été observé pour le MNMB (entre chaque AR et le NR). Ce paramètre, dont les différences entre les deux versions d'IRS ont été évaluées en moyenne sur chaque mois, avait montré des différences plus fortes en août au bord gauche du domaine, au-dessus de l'Océan Atlantique. Lorsque l'on fait la moyenne sur l'ensemble du domaine, ces valeurs conduisent à des valeurs quotidiennes plus fortes des différences de biais par rapport aux deux autres mois.*

*Une étude de la colonne troposphérique, a montré des résultats différents. L'AR des observations IRS*2 était, parmi les deux, le plus proche du NR. Dans ce cas, donc, IRS*2 a montré une meilleure contribution dans l'assimilation par rapport à IRS. Les valeurs des différences de biais sont restées faibles (maximum* $0,35\,\mathrm{DU}$*), même en considérant la faible concentration d'ozone troposphérique.*
*En évaluant les valeurs troposphériques moyennées sur le domaine, jour par jour, une meilleure contribution des luminances IRS*2 a été trouvée sur presque toute la période. Les valeurs de différence sont du même ordre que celles trouvées pour les colonnes totales (maximum* $0,15\,\mathrm{DU}$*), malgré le fait que la concentration troposphérique soit plus faible.*

*Une évaluation de la fraction molaire, par MOCAGE niveaux verticaux, a révélé que la différence d'impact des deux IRS dépend du niveau atmosphérique. IRS*2, cependant, même avec de très petites valeurs de variation entre les biais par rapport au NR, a montré une meilleure performance générale par rapport à IRS, principalement dans les niveaux atmosphériques les plus bas. Du niveau* 10 *à* 25*, aux très basses latitudes, les variations les plus importantes se produisent, mais elles ont été causées par les effets de couplage entre les domaines et, par conséquent, elles n'ont pas été incluses dans la discussion. Entre des latitudes d'environ* $36°$ *jusqu'à environ* $60°$ *(au-delà desquelles plus aucune observation n'a été simulée), des comportements différents des deux AR ont été observés. Aux niveaux très élevés (d'environ* 15 *jusqu'à* 25*), les deux IRS ont alterné en donnant la meilleure performance (variations entre les biais (*10^{-7} *ppv magnitude)), avec une contribution légèrement meilleure d'IRS*2 par rapport à IRS. En dessous du niveau* 30 *(c'est-à-dire* $\sim 90\,\mathrm{hPa}$*), IRS*2 s'était avéré être presque toujours le plus proche du NR, à l'exception d'une zone située entre les niveaux* 47 *et* 53 *(autour de* 600 *et* 860 hPa*). À ces niveaux, les variations ont atteint des maxima de* 10^{-9} *ppv. Dans la basse troposphère, IRS*2 était toujours plus proche du NR que IRS (variation maximale de* 10^{-10} *ppv).*

Perspectives

A partir des résultats obtenus et examinés, de nombreuses perspectives s'ouvrent pour ces travaux.

▶ *Amélioration de l'évaluation de l'assimilation*

La première action à mener à très court terme est d'améliorer l'évaluation de l'assimilation. Il a été indiqué à plusieurs reprises que l'impact du couplage entre les domaines MOCAGE produit des valeurs plus variables des innovations et, donc, du champ d'ozone, à la limite sud du domaine régional MACC01 traité. Pour s'assurer que les moyennes sur le domaine ne sont pas affectées par ces valeurs, l'impact de l'assimilation devrait être aussi évalué sur une zone plus petite, qui coupe ces bords. De même, lorsqu'on évalue les valeurs en tant que moyennes géographiques, on pourrait exclure de l'évaluation les zones qui sont situées en dehors du disque d'observation simulé, et donc assimilé.

En second lieu, les cas spécifiques de pollution par l'ozone, qui ont conduit au choix de la période d'étude, n'ont pas été étudiés en détail par manque de temps, mais seulement à travers l'évaluation de séries temporelles avec des moyennes sur le domaine. Une vérification plus approfondie de ces cas est donc nécessaire. Des scores spécifiques sur la zone la plus intensément touchée devraient être calculés et étudiés..

Enfin, l'étude menée autour du cadre de préparation du NR a montré l'intérêt d'utiliser une matrice d'erreur d'ébauche calculée par la méthode NMC, au lieu de la matrice classiquement utilisée dans la littérature (pourcentage du profil d'ozone). Ce type d'approche peut être bénéfique pour l'assimilation des radiances IRS, à la fois pour les données simulées et, lorsque les observations seront disponibles, une fois le satellite lancé. Une période de trois mois peut ne pas être suffisante pour appliquer une telle méthode sur un modèle régional.

▶ *Amélioration du traitement des observations IRS*

*Comme mentionné précédemment, seules les composantes principales (en anglais Principal Components - PCs) seront distribuées pour IRS. Cette étude n'a, cependant, pas pris en compte ce sujet. Une courte étude évaluant les différents impacts des PCs sur le travail illustré jusqu'à présent, devrait être réalisée. Bien entendu, cela ne serait possible, à ce stade, que pour la version officielle d'IRS (et non pour IRS*2).*

Un autre point qui est ressorti de l'étude sur la préparation du cadre NR a été la possibilité d'effectuer une sélection de canaux, tout en maintenant une bonne qualité dans les résultats de l'assimilation. Il s'agit également d'une procédure couramment utilisée en NWP pour de nombreux instruments et sur laquelle j'ai moi-même travaillé dans le cadre d'un autre projet sur le futur instrument IASI-NG [Vittorioso et al. (2021)]. Le travail effectué ici sur IRS

*étant une première analyse, nous avons voulu étudier le comportement de toutes les longueurs d'onde consécutives considérées. De plus, l'utilisation de la bande spectrale complète était un moyen équitable de comparer l'impact d'IRS et d'IRS*2. Il sera toutefois intéressant, à un stade ultérieur, d'effectuer également une sélection d'un groupe plus restreint de canaux.*

▶ *Impact of Synergy with Other Instruments*

À bord des satellites MTG-S, avec IRS, sera également placé Sentinel-4, un spectromètre imageur UVN. Il est conçu pour surveiller certains gaz à l'état de traces et aérosols essentiels à la qualité de l'air en Europe, avec une haute résolution spatiale et un temps de revisite rapide, à l'appui du CAMS. Étant donné le potentiel de l'acquisition conjointe des deux instruments, leur synergie dans la caractérisation de la pollution atmosphérique au-dessus de l'Europe doit absolument être testée.

Si l'on considère l'assimilation d'IRS dans le domaine global de MOCAGE, l'assimilation conjointe d'IRS avec GIIRS (sondeur IR hyperspectral à bord de satellites géostationnaires chinois) pourrait conduire à une validation intéressante.

Parallèlement, d'autres spectromètres fonctionnant de manière similaire, mais sur des plateformes en orbite polaire, seraient également intéressants à étudier. Les premiers sur la liste sont certainement IASI et la nouvelle génération, IASI-NG (dès que ses données seront exploitables), à bord de satellites européens. D'autres sondeurs IR des systèmes polaires américains et chinois (CrIS et HIRAS) pourraient également être ajoutés.

L'assimilation conjointe des radiances IRS et des mesures de surface pourrait aussi être envisagée dans le futur. Comme IRS fournit de bonnes informations verticales dans la stratosphère, l'UTLS et la troposphère libre, les données de surface peuvent apporter des détails supplémentaires dans les couches atmosphériques les plus basses.

▶ *Vers le traitement d'autres espèces*

Dans la suite de ce travail, il y aura certainement l'évaluation de l'assimilation des radiances dans les bandes sensibles à d'autres espèces chimiques et aux aérosols. Le cadre OSSE créé pour cette étude a été pensé pour fournir, avec des modifications mineures, les bases de l'extension à d'autres espèces que l'ozone.

Le dioxyde de soufre (SO_2) pourrait être étudié puisque IRS a le potentiel de détecter cette espèce, avec une sensibilité vers la fin de la bande MWIR [Coopmann et al. (2022)]. Une étude de cas plus spécifique devrait cependant être prise en compte pour ce composé, car il est associé à des événements naturels spécifiques et ponctuels.

Le monoxyde de carbone va pourtant être la première cible, au vu des résultats rencontrés lors de l'étude suivante réalisée pour évaluer la sensibilité de l'IRS dans ses deux versions.

*Lors d'un stage effectué par deux étudiants de Master1 SOAC[3] (Julie Wendelin et Louis Ramu), et que j'ai co-encadré, le monoxyde de carbone a commencé à être évalué. Les étudiants se sont concentrés sur une seule journée dans la période d'étude (29 juin 2019). Ce jour-là, de fortes concentrations de CO ($\sim 3 \cdot 10^{-7}$ ppv) ont été relevées sur deux villes du domaine (à savoir Dunkerque et Bagdad). L'étude, a été réalisée sur une sous-partie du domaine MACC01 divisé en une grille régulière de 1428 points. Ils ont considéré comme réalité le NR produit dans le cadre de l'OSSE de cette these. L'ébauche a été obtenue en ajoutant à l'état réel un bruit gaussien centré sur un écart-type de la valeur de 10%. Comme observations, ils n'ont pas utilisé celles simulées dans le cadre de l'OSSE, mais ils ont dû simuler des observations pour les longueurs d'ondes sensibles au CO avec RTTOV, en tenant compte du bruit instrumental de l'IRS et de l'IRS*2.*

*Pour chaque point de grille, ils ont évalué le contenu en information en termes de degrés de liberté DFS (voir l'Annexe C). A partir d'une évaluation des résultats (Figure I), ils ont constaté que IRS*2 semble apporter plus d'information qu'IRS. Cette différence est plus prononcée sur les continents et les régions où la température de surface est plus élevée, c'est-à-dire où l'émission de la surface est plus forte. Quoi qu'il en soit, la différence entre les deux versions de l'instrument reste faible.*

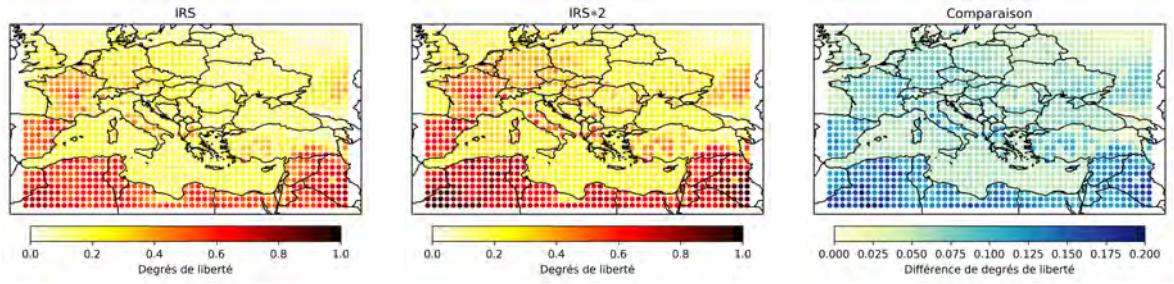


Figure I: Cartes des DFS pour IRS (à gauche) et IRS*2 (au milieu) ; et comparaison des instruments $\mathrm{DFS}_{\mathrm{IRS}*2}$ - $\mathrm{DFS}_{\mathrm{IRS}}$ (à droite).

*Grâce aux restitutions unidimensionnelles, ils ont déterminé l'analyse pour chaque point de grille, en obtenant les profils CO. Ils ont effectué une évaluation statistique de la différence relative en pourcentage de l'écart-type de l'erreur d'analyse par rapport à l'écart-type de l'erreur d'ebauche[4]. Le résultat est illustré dans la Figure II, pour IRS et IRS*2, sous forme de zoom sur la colonne verticale allant du niveau 25 du modèle ($\sim$ 50 hPa) à la surface. Il est évident qu'aux niveaux entre 40 et 45 (c'est-à-dire entre $\sim$ 320 et 530 hPa), il y a une réduction de l'erreur qui est 1.5% plus forte pour IRS*2 que pour IRS.*

[3] Sciences de l'Océan, de l'Atmosphère et du Climat (SOAC)
[4] $[\sigma(\mathbf{x}_\mathrm{a} - \mathbf{x}_\mathrm{t}) - \sigma(\mathbf{x}_\mathrm{b} - \mathbf{x}_\mathrm{t})]/\sigma(\mathbf{x}_\mathrm{b} - \mathbf{x}_\mathrm{t})$

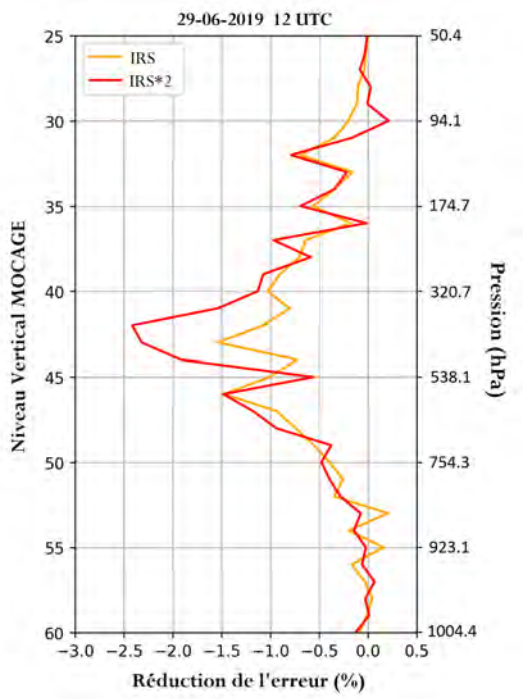


Figure II: Zoom sur la différence normalisée des écarts-types des erreurs d'ébauche et d'analyse pour IRS at IRS*2.

Ces résultats sont très prometteurs et encouragent à entreprendre, dès que possible, aussi pour le monoxyde de carbone le même type d'évaluation qui a été fait dans cette thèse pour l'ozone.

Acronyms

EUMETSAT EUropean Organization for the Exploitation of METeorological SATellites . . . 2

FCI Flexible Combined Imager . . . 72

FFT Fast Fourier Transform . . . 79

FGE fractional gross error . . . 218

FOV field of view . . . 75

FRP fire radiative power . . . 53

FTS Fourier-Transform Spectrometer . . . 74

GCMs General Circulation Models . . . 49

GEIA Global Emissions Inventory Activity . . . 53

GEMS Geostationary Environmental Monitoring Spectrometer . . . 71

GEO geostationary orbit . . . 70

GFAS Global Fire Assimilation System . . . 53

GIIRS Geostationary Interferometric Infrared Sounder . . . 72

GhG greenhouse gases . . . 14

HIRAS Hyperspectral Infrared Atmospheric Sounder . . . 72

IAGOS In-service Aircraft for a Global Observing System . . . 67

IASI Infrared Atmospheric Sounding Interferometer . . . 45

IKFS-2 Infrared Fourier Spectrometer - 2 . . . 72

IFS Integrated Forecast System . . . 50

IR infrared . . . 14

IREMIS IR EMISsivity . . . 45

IRS Infrared Sounder . . . 65

ISEM Infrared Surface Emissivity Model . . . 45

ISFR instrument spectral response function . . . 44

ISS International Space Station . . . 12

JAXA Japan Aerospace Exploration Agency . . . 71

KCARTA KCompressed Atmospheric Radiative Transfer Algorithm . . . 43

KNMI Koninklijk Nederlands Meteorologisch Instituut . . . 49

LAC Local Area Coverage . . . 77

LBL line-by-line . . . 43

LBLRTM Line-By-Line Radiative Transfer Model . . . 43

Bibliography

Andersson, E., Haseler, J., Undén, P., Courtier, P., Kelly, G., Vasiljevic, D., Brankovic, C., Gaffard, C., Hollingsworth, A., Jakob, C., et al. (1998). The ecmwf implementation of three-dimensional variational assimilation (3d-var). iii: Experimental results. *Quarterly Journal of the Royal Meteorological Society*, 124(550):1831–1860.

Arnold Jr, C. P. and Dey, C. H. (1986). Observing-systems simulation experiments: Past, present, and future. *Bulletin of the American Meteorological Society*, 67(6):687–695.

Atlas, R. (1997). Atmospheric observations and experiments to assess their usefulness in data assimilation (gtspecial issueltdata assimilation in meteology and oceanography: theory and practice). *Journal of the Meteorological Society of Japan. Ser. II*, 75(1B):111–130.

Aumont, B. (2005). Production d'ozone troposphérique et régimes chimiques. *Cours Master Recherche" SGE" spécialité" Atmosphère et qualité de l'air", Université Paris VII.*

Baek, K., Kim, J. H., Bak, J., Haffner, D. P., Kang, M., and Hong, H. (2022). Evaluation of total ozone measurements from geostationary environmental monitoring satellite (gems). *EGUsphere*, pages 1–23.

Barré, J., Peuch, V.-H., Lahoz, W., Attié, J.-L., Josse, B., Piacentini, A., Eremenko, M., Dufour, G., Nedelec, P., von Clarmann, T., et al. (2014). Combined data assimilation of ozone tropospheric columns and stratospheric profiles in a high-resolution ctm. *Quarterly Journal of the Royal Meteorological Society*, 140(680):966–981.

Beer, R. (2006). Tes on the aura mission: Scientific objectives, measurements, and analysis overview. *IEEE Transactions on Geoscience and remote sensing*, 44(5):1102–1105.

Berk, A., Conforti, P., Kennett, R., Perkins, T., Hawes, F., and van den Bosch, J. (2014). Modtran® 6: A major upgrade of the modtran® radiative transfer code. In *2014 6th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, pages 1–4.

Bermudo, F., Rousseau, S., Pequignot, E., and Bernard, F. (2014). Iasi-ng program: A new generation of infrared atmospheric sounding interferometer. In *2014 IEEE Geoscience and Remote Sensing Symposium*, pages 1373–1376. IEEE.

Bessagnet, B., Terrenoire, E., Tognet, F., Rouïl, L., Colette, A., Letinois, L., and Malherbe, L. (2012). The chimere atmospheric model. *EC4MACS Modelling Methodology, Report*, 10.

Bloom, H. J. (2001). The cross-track infrared sounder (cris): a sensor for operational meteorological remote sensing. In *IGARSS 2001. Scanning the Present and Resolving the Future. Proceedings. IEEE 2001 International Geoscience and Remote Sensing Symposium (Cat. No. 01CH37217)*, volume 3, pages 1341–1343. IEEE.

Blumstein, D., Chalon, G., Carlier, T., Buil, C., Hebert, P., Maciaszek, T., Ponce, G., Phulpin, T., Tournier, B., Simeoni, D., et al. (2004). Iasi instrument: Technical overview and measured performances. *Infrared Spaceborne Remote Sensing XII*, 5543:196–207.

Borbas, E. E., Hulley, G., Feltz, M., Knuteson, R., and Hook, S. (2018). The combined aster modis emissivity over land (camel) part 1: Methodology and high spectral resolution application. *Remote Sensing*, 10(4):643.

Borbas, E. E. and Ruston, B. C. (2011). *The RTTOV UWiremis IR land surface emissivity module*. EUMETSAT Darmstadt.

Boukabara, S.-A., Moradi, I., Atlas, R., Casey, S. P., Cucurull, L., Hoffman, R. N., Ide, K., Krishna Kumar, V., Li, R., Li, Z., et al. (2016). Community global observing system simulation experiment (osse) package (cgop): description and usage. *Journal of Atmospheric and Oceanic Technology*, 33(8):1759–1777.

Bouttier, F. and Courtier, P. (2002). Data assimilation concepts and methods march 1999. *Meteorological training course lecture series. ECMWF*, 718:59.

Bouyssel, F., Berre, L., Bénichou, H., Chambon, P., Girardot, N., Guidard, V., Loo, C., Mahfouf, J.-F., Moll, P., Payan, C., et al. (2022). The 2020 global operational nwp data assimilation system at météo-france. In *Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications (Vol. IV)*, pages 645–664. Springer.

Boxe, C., Worden, J., Bowman, K., Kulawik, S., Neu, J., Ford, W., Osterman, G., Herman, R., Eldering, A., Tarasick, D., et al. (2010). Validation of northern latitude tropospheric emission spectrometer stare ozone profiles with arc-ions sondes during arctas: sensitivity, bias and error analysis. *Atmospheric Chemistry and Physics*, 10(20):9901–9914.

Brasseur, G., Hauglustaine, D., Walters, S., Rasch, P., Müller, J.-F., Granier, C., and Tie, X. (1998). Mozart, a global chemical transport model for ozone and related chemical tracers: 1. model description. *Journal of Geophysical Research: Atmospheres*, 103(D21):28265–28289.

Brousseau, P., Seity, Y., Ricard, D., and Léger, J. (2016). Improvement of the forecast of convective activity from the arome-france system. *Quarterly Journal of the Royal Meteorological Society*, 142(699):2231–2243.

Burkholder, J. B., Abbatt, J. P., Barnes, I., Roberts, J. M., Melamed, M. L., Ammann, M., Bertram, A. K., Cappa, C. D., Carlton, A. G., Carpenter, L. J., et al. (2017). The essential role for laboratory studies in atmospheric chemistry.

Butchart, N. (2014). The brewer-dobson circulation. *Reviews of geophysics*, 52(2):157–184.

Carnevale, C., Finzi, G., Pisoni, E., Volta, M., Kishcha, P., Curci, G., and Alpert, P. (2011). Improved ctm boundary conditions using dream desert dust forecasts: a case study over the po valley. In *Air Pollution Modeling and its Application XXI*, pages 599–605. Springer.

Cayla, F.-R. (2001). L'interféromètre iasi-un nouveau sondeur satellitaire haute résolution. *La météorologie*.

Chailleux, Y. (2018). *Observation satellitaire de la pollution à l'ozone par synergie multispectrale à trois bandes Ultraviolet+Visible+ Infrarouge*. PhD thesis, Université Paris-Est.

Chapman, S. (1930). Xxxv. on ozone and atomic oxygen in the upper atmosphere. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 10(64):369–383.

Claeyman, M. (2010). *Etude par modélisation et assimilation de données d'un capteur infrarouge géostationnaire pour la qualité de l'air*. PhD thesis, Toulouse 3.

Claeyman, M., Attié, J.-L., Peuch, V.-H., El Amraoui, L., Lahoz, W. A., Josse, B., Joly, M., Barré, J., Ricaud, P., Massart, S., et al. (2011). A thermal infrared instrument onboard a geostationary platform for co and o 3 measurements in the lowermost troposphere: Observing system simulation experiments (osse). *Atmospheric Measurement Techniques*, 4(8):1637–1661.

Clarisse, L., Van Damme, M., Hurtmans, D., Franco, B., Clerbaux, C., and Coheur, P.-f. (2021). The diel cycle of nh3 observed from the fy-4a geostationary interferometric infrared sounder (giirs). *Geophysical Research Letters*, 48(14):e2021GL093010.

Clerbaux, C., Boynard, A., Clarisse, L., George, M., Hadji-Lazaro, J., Herbin, H., Hurtmans, D., Pommier, M., Razavi, A., Turquety, S., et al. (2009). Monitoring of atmospheric composition using the thermal infrared iasi/metop sounder. *Atmospheric Chemistry and Physics*, 9(16):6041–6054.

Clough, S., Shephard, M., Mlawer, E., Delamere, J., Iacono, M., Cady-Pereira, K., Boukabara, S., and Brown, P. (2005). Atmospheric radiative transfer modeling: A summary of the aer codes. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 91(2):233–244.

Clough, S. A., Iacono, M. J., and Moncet, J.-L. (1992). Line-by-line calculations of atmospheric fluxes and cooling rates: Application to water vapor. *Journal of Geophysical Research: Atmospheres*, 97(D14):15761–15785.

Coopmann, O., Fourrié, N., Chambon, P., Vidot, J., Brousseau, P., Martet, M., and Birman, C. (2023). Preparing the assimilation of the future mtg-irs sounder into the mesoscale nwp arome model [manuscript submitted for publication]. *Quarterly Journal of the Royal Meteorological Society*.

Coopmann, O., Fourrié, N., and Guidard, V. (2022). Analysis of mtg-irs observations and general channel selection for numerical weather prediction models. *Quarterly Journal of the Royal Meteorological Society*.

Courtier, P., Freydier, C., Geleyn, J., Rabier, F., and Rochas, M. (1991). The arpege project at météo-france. In *Proceedings of the 1991 ECMWF Seminar*, pages 193–231.

Crevoisier, C., Clerbaux, C., Guidard, V., Phulpin, T., Armante, R., Barret, B., Camy-Peyret, C., Chaboureau, J.-P., Coheur, P.-F., Crépeau, L., et al. (2014). Towards iasi-new generation (iasi-ng): impact of improved spectral resolution and radiometric noise on the retrieval of thermodynamic, chemistry and climate variables. *Atmospheric Measurement Techniques*, 7:4367–4385.

Cussac, M., Marécal, V., Thouret, V., Josse, B., and Sauvage, B. (2020). The impact of biomass burning on upper tropospheric carbon monoxide: a study using mocage global model and iagos airborne data. *Atmospheric Chemistry and Physics*, 20(15):9393–9417.

Delmas, R., Mégie, G., and Peuch, V.-H. (2005). *Physique et chimie de l'atmosphère*. Belin.

Descheemaecker, M., Plu, M., Marécal, V., Claeyman, M., Olivier, F., Aoun, Y., Blanc, P., Wald, L., Guth, J., Sič, B., et al. (2019). Monitoring aerosols over europe: an assessment of the potential benefit of assimilating the vis04 measurements from the future mtg/fci geostationary imager. *Atmospheric Measurement Techniques*, 12(2):1251–1275.

DeSouza-Machado, S., Strow, L. L., and Hannon, S. E. (1997). kcompressed atmospheric radiative transfer algorithm (kcarta). In *Satellite Remote Sensing of Clouds and the Atmosphere II*, volume 3220, pages 156–162. SPIE.

Desroziers, G., Berre, L., Chapnik, B., and Poli, P. (2005). Diagnosis of observation, background and analysis-error statistics in observation space. *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography*, 131(613):3385–3396.

Diehl, T., Heil, A., Chin, M., Pan, X., Streets, D., Schultz, M., and Kinne, S. (2012). Anthropogenic, biomass burning, and volcanic emissions of black carbon, organic carbon, and so 2 from 1980 to 2010 for hindcast model experiments. *Atmospheric Chemistry and Physics Discussions*, 12(9):24895–24954.

Ding, S., Yang, P., Weng, F., Liu, Q., Han, Y., Van Delst, P., Li, J., and Baum, B. (2011). Validation of the community radiative transfer model. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 112(6):1050–1064.

Dufour, G., Eremenko, M., Griesfeller, A., Barret, B., LeFlochmoën, E., Clerbaux, C., Hadji-Lazaro, J., Coheur, P.-F., and Hurtmans, D. (2012). Validation of three different scientific ozone products retrieved from iasi spectra using ozonesondes. *Atmospheric Measurement Techniques*, 5(3):611–630.

Duruisseau, F., Chambon, P., Guedj, S., Guidard, V., Fourrié, N., Taillefer, F., Brousseau, P., Mahfouf, J.-F., and Roca, R. (2017). Investigating the potential benefit to a mesoscale nwp model of a microwave sounder on board a geostationary satellite. *Quarterly Journal of the Royal Meteorological Society*, 143(706):2104–2115.

Eckermann, S. (2009). Hybrid σ–p coordinate choices for a global model. *Monthly Weather Review*, 137(1):224–245.

El Aabaribaoune, M. (2022). *Assimilation des luminances IASI dans un modèle de chimie transport pour la surveillance de l'ozone et des poussières désertiques*. PhD thesis, Université Toulouse III-Paul Sabatier.

El Aabaribaoune, M., Emili, E., and Guidard, V. (2021). Estimation of the error covariance matrix for iasi radiances and its impact on the assimilation of ozone in a chemistry transport model. *Atmospheric Measurement Techniques*, 14(4):2841–2856.

El Amraoui, L., Attié, J.-L., Semane, N., Claeyman, M., Peuch, V.-H., Warner, J., Ricaud, P., Cammas, J.-P., Piacentini, A., Josse, B., et al. (2010). Midlatitude stratosphere–troposphere exchange as diagnosed by mls o 3 and mopitt co assimilated fields. *Atmospheric Chemistry and Physics*, 10(5):2175–2194.

El Amraoui, L., Plu, M., Guidard, V., Cornut, F., and Bacles, M. (2022). A pre-operational system based on the assimilation of modis aerosol optical depth in the mocage chemical transport model. *Remote Sensing*, 14(8):1949.

Emili, E., Barret, B., Le Flochmoën, E., and Cariolle, D. (2019). Comparison between the assimilation of iasi level 2 ozone retrievals and level 1 radiances in a chemical transport model. *Atmospheric Measurement Techniques*, 12(7):3963–3984.

Emmons, L. K., Walters, S., Hess, P. G., Lamarque, J.-F., Pfister, G. G., Fillmore, D., Granier, C., Guenther, A., Kinnison, D., Laepple, T., et al. (2010). Description and evaluation of the model for ozone and related chemical tracers, version 4 (mozart-4). *Geoscientific Model Development*, 3(1):43–67.

Eresmaa, R. and McNally, A. P. (2016). Diverse profile datasets based on the cams atmospheric composition forecasting system. *ECMWF Rep.*

Errico, R. M., Yang, R., Masutani, M., and Woollen, J. (2007). The use of an osse to estimate characteristics of analysis error. *Meteorologische Zeitschrift*.

Filler, A. (1964). Apodization and interpolation in fourier-transform spectroscopy. *JOSA*, 54(6):762–767.

Finlayson-Pitts, B. J. and Pitts Jr, J. N. (1997). Tropospheric air pollution: ozone, airborne toxics, polycyclic aromatic hydrocarbons, and particles. *Science*, 276(5315):1045–1051.

Fioletov, V. (2008). Ozone climatology, trends, and substances that control ozone. *Atmosphere-Ocean*, 46(1):39–67.

Fujino, J., Nair, R., Kainuma, M., Masui, T., and Matsuoka, Y. (2006). Multi-gas mitigation analysis on stabilization scenarios using aim global model. *The Energy Journal*, (Special Issue# 3).

Gambacorta, A. and Barnet, C. D. (2012). Methodology and information content of the noaa nesdis operational channel selection for the cross-track infrared sounder (cris). *IEEE transactions on geoscience and remote sensing*, 51(6):3207–3216.

Geer, A., Lahoz, W., Bekki, S., Bormann, N., Errera, Q., Eskes, H., Fonteyn, D., Jackson, D., Juckes, M., Massart, S., et al. (2006). The asset intercomparison of ozone analyses: method and first results. *Atmospheric Chemistry and Physics*, 6(12):5445–5474.

Gettelman, A., Hoor, P., Pan, L., Randel, W., Hegglin, M. I., and Birner, T. (2011). The extratropical upper troposphere and lower stratosphere. *Reviews of Geophysics*, 49(3).

Granier, C., Bessagnet, B., Bond, T., D'Angiola, A., Denier van der Gon, H., Frost, G. J., Heil, A., Kaiser, J. W., Kinne, S., Klimont, Z., et al. (2011). Evolution of anthropogenic and biomass burning emissions of air pollutants at global and regional scales during the 1980–2010 period. *Climatic change*, 109(1):163–190.

Granier, C., Darras, S., van der Gon, H. D., Jana, D., Elguindi, N., Bo, G., Michael, G., Marc, G., Jalkanen, J.-P., Kuenen, J., et al. (2019). *The Copernicus atmosphere monitoring service global and regional emissions (April 2019 version)*. PhD thesis, Copernicus Atmosphere Monitoring Service.

Guedj, S., Guidard, V., Ménétrier, B., Mahfouf, J.-F., and Rabier, F. (2014). Future benefits of high-density radiance data from mtg-irs in the arome fine-scale forecast model, final report. *Météo-France and CNRS/CNRM-GAME*.

Guevara, M., Petetin, H., Jorba, O., Denier van der Gon, H., Kuenen, J., Super, I., Jalkanen, J.-P., Majamäki, E., Johansson, L., Peuch, V.-H., et al. (2022). European primary emissions of criteria pollutants and greenhouse gases in 2020 modulated by the covid-19 pandemic disruptions. *Earth System Science Data*, 14(6):2521–2552.

Guth, J., Josse, B., Marécal, V., Joly, M., and Hamer, P. (2016). First implementation of secondary inorganic aerosols in the mocage version r2. 15.0 chemistry transport model. *Geoscientific Model Development*, 9(1):137–160.

Han, Y., Revercomb, H., Cromp, M., Gu, D., Johnson, D., Mooney, D., Scott, D., Strow, L., Bingham, G., Borg, L., et al. (2013). Suomi npp cris measurements, sensor data record algorithm, calibration and validation activities, and record data quality. *Journal of Geophysical Research: Atmospheres*, 118(22):12–734.

Han, Y., Suwinski, L., Tobin, D., and Chen, Y. (2015). Effect of self-apodization correction on cross-track infrared sounder radiance noise. *Applied Optics*, 54(34):10114–10122.

Hilton, F., Armante, R., August, T., Barnet, C., Bouchard, A., Camy-Peyret, C., Capelle, V., Clarisse, L., Clerbaux, C., Coheur, P.-F., et al. (2012). Hyperspectral earth observation from iasi: Five years of accomplishments. *Bulletin of the American Meteorological Society*, 93(3):347–370.

Hocking, J., Rayer, P., Rundle, D., Saunders, R., Matricardi, M., Geer, A., Brunel, P., and Vidot, J. (2019). Rttov v12 users guide. *EUMETSAT*.

Holton, J. R. (1982). The role of gravity wave induced drag and diffusion in the momentum budget of the mesosphere. *Journal of Atmospheric Sciences*, 39(4):791–799.

Holton, J. R., Haynes, P. H., McIntyre, M. E., Douglass, A. R., Rood, R. B., and Pfister, L. (1995). Stratosphere-troposphere exchange. *Reviews of geophysics*, 33(4):403–439.

Hoor, P., Fischer, H., Lange, L., Lelieveld, J., and Brunner, D. (2002). Seasonal variations of a mixing layer in the lowermost stratosphere as identified by the co-o3 correlation from in situ measurements. *Journal of Geophysical Research: Atmospheres*, 107(D5):ACL–1.

Hoor, P., Wernli, H., Hegglin, M. I., and Bönisch, H. (2010). Transport timescales and tracer properties in the extratropical utls. *Atmospheric Chemistry and Physics*, 10(16):7929–7944.

Huijnen, V., Eskes, H., Poupkou, A., Elbern, H., Boersma, K., Foret, G., Sofiev, M., Valdebenito, A., Flemming, J., Stein, O., et al. (2010). Comparison of omi no 2 tropospheric columns with an ensemble of global and european regional air quality models. *Atmospheric Chemistry and Physics*, 10(7):3273–3296.

Huijnen, V., Pozzer, A., Arteta, J., Brasseur, G., Bouarar, I., Chabrillat, S., Christophe, Y., Doumbia, T., Flemming, J., Guth, J., et al. (2019). Quantifying uncertainties due to chemistry modelling–evaluation of tropospheric composition simulations in the cams model (cycle 43r1). *Geoscientific Model Development*, 12(4):1725–1752.

Inness, A., Flemming, J., Heue, K.-P., Lerot, C., Loyola, D., Ribas, R., Valks, P., Van Roozendael, M., Xu, J., and Zimmer, W. (2019). Monitoring and assimilation tests with tropomi

data in the cams system: near-real-time total column ozone. *Atmospheric Chemistry and Physics*, 19(6):3939–3962.

Josse, B., Simon, P., and Peuch, V.-H. (2004). Radon global simulations with the multiscale chemistry and transport model mocage. *Tellus B: Chemical and Physical Meteorology*, 56(4):339–356.

Kaiser, J., Heil, A., Andreae, M., Benedetti, A., Chubarova, N., Jones, L., Morcrette, J.-J., Razinger, M., Schultz, M., Suttie, M., et al. (2012). Biomass burning emissions estimated with a global fire assimilation system based on observed fire radiative power. *Biogeosciences*, 9(1):527–554.

Kuenen, J., Dellaert, S., Visschedijk, A., Jalkanen, J.-P., Super, I., and Denier van der Gon, H. (2022). Cams-reg-v4: a state-of-the-art high-resolution european emission inventory for air quality modelling. *Earth System Science Data*, 14(2):491–515.

Kuilman, M. S. and Karlsson, B. (2018). The role of the winter residual circulation in the summer mesopause regions in waccm. *Atmospheric Chemistry and Physics*, 18(6):4217–4228.

Lacressonnière, G., Peuch, V.-H., Arteta, J., Josse, B., Joly, M., Marécal, V., Saint Martin, D., Déqué, M., and Watson, L. (2012). How realistic are air quality hindcasts driven by forcings from climate model simulations? *Geoscientific Model Development*, 5(6):1565–1587.

Lamarque, J.-F., Bond, T. C., Eyring, V., Granier, C., Heil, A., Klimont, Z., Lee, D., Liousse, C., Mieville, A., Owen, B., et al. (2010). Historical (1850–2000) gridded anthropogenic and biomass burning emissions of reactive gases and aerosols: methodology and application. *Atmospheric Chemistry and Physics*, 10(15):7017–7039.

Lamarque, J.-F., Shindell, D. T., Josse, B., Young, P., Cionni, I., Eyring, V., Bergmann, D., Cameron-Smith, P., Collins, W. J., Doherty, R., et al. (2013). The atmospheric chemistry and climate model intercomparison project (accmip): Overview and description of models, simulations and climate diagnostics. *Geoscientific Model Development*, 6(1):179–206.

Lamotte, C., Guth, J., Marécal, V., Cussac, M., Hamer, P. D., Theys, N., and Schneider, P. (2021). Modeling study of the impact of so2 volcanic passive emissions on the tropospheric sulfur budget. *Atmospheric Chemistry and Physics*, 21(14):11379–11404.

Lefevre, F., Brasseur, G., Folkins, I., Smith, A., and Simon, P. (1994). Chemistry of the 1991–1992 stratospheric winter: Three-dimensional model simulations. *Journal of Geophysical Research: Atmospheres*, 99(D4):8183–8195.

Lelieveld, J., Berresheim, H., Borrmann, S., Crutzen, P. J., Dentener, F., Fischer, H., Feichter, J., Flatau, P., Heland, J., Holzinger, R., et al. (2002). Global air pollution crossroads over the mediterranean. *Science*, 298(5594):794–799.

Lelieveld, J. and Dentener, F. J. (2000). What controls tropospheric ozone? *Journal of Geophysical Research: Atmospheres*, 105(D3):3531–3551.

Levelt, P. F., Joiner, J., Tamminen, J., Veefkind, J. P., Bhartia, P. K., Stein Zweers, D. C., Duncan, B. N., Streets, D. G., Eskes, H., van der A, R., et al. (2018). The ozone monitoring instrument: overview of 14 years in space. *Atmospheric Chemistry and Physics*, 18(8):5699–5745.

Levelt, P. F., Van Den Oord, G. H., Dobber, M. R., Malkki, A., Visser, H., De Vries, J., Stammes, P., Lundell, J. O., and Saari, H. (2006). The ozone monitoring instrument. *IEEE Transactions on geoscience and remote sensing*, 44(5):1093–1101.

Lu, R., Turco, R. P., and Jacobson, M. Z. (1997). An integrated air pollution modeling system for urban and regional scales: 2. simulations for scaqs 1987. *Journal of Geophysical Research: Atmospheres*, 102(D5):6081–6098.

Lübken, F.-J., Von Zahn, U., Manson, A., Meek, C., Hoppe, U.-P., Schmidlin, F., Stegman, J., Murtagh, D., Rüster, R., Schmidt, G., et al. (1990). Mean state densities, temperatures and winds during the mac/sine and mac/epsilon campaigns. *Journal of atmospheric and terrestrial physics*, 52(10-11):955–970.

Marécal, V., Peuch, V.-H., Andersson, C., Andersson, S., Arteta, J., Beekmann, M., Benedictow, A., Bergström, R., Bessagnet, B., Cansado, A., et al. (2015). A regional air quality forecasting system over europe: the macc-ii daily ensemble production. *Geoscientific Model Development*, 8(9):2777–2813.

Martet, M., Peuch, V., Laurent, B., Marticorena, B., and Bergametti, G. (2009). Evaluation of long-range transport and deposition of desert dust with the ctm mocage. *Tellus B: Chemical and Physical Meteorology*, 61(2):449–463.

Martin, T. B., Prunet, S., and Drissen, L. (2016). Optimal fitting of gaussian-apodized or under-resolved emission lines in fourier transform spectra providing new insights on the velocity structure of ngc 6720. *Monthly Notices of the Royal Astronomical Society*, 463(4):4223–4238.

Martinet, P., Lavanant, L., Fourrié, N., Rabier, F., and Gambacorta, A. (2014). Evaluation of a revised iasi channel selection for cloudy retrievals with a focus on the mediterranean basin. *Quarterly Journal of the Royal Meteorological Society*, 140(682):1563–1577.

Masutani, M., Schlatter, T. W., Errico, R. M., Stoffelen, A., Andersson, E., Lahoz, W., Woollen, J. S., Emmitt, G. D., Riishøjgaard, L.-P., and Lord, S. J. (2010). Observing system simulation experiments. In *Data Assimilation*, pages 647–679. Springer.

McCarty, W., Errico, R. M., and Gelaro, R. (2012). Cloud coverage in the joint osse nature run. *Monthly weather review*, 140(6):1863–1871.

Menut, L., Bessagnet, B., Khvorostyanov, D., Beekmann, M., Blond, N., Colette, A., Coll, I., Curci, G., Foret, G., Hodzic, A., et al. (2013). Chimere 2013: a model for regional atmospheric composition modelling. *Geoscientific model development*, 6(4):981–1028.

Mircea, M., Ciancarella, L., Briganti, G., Calori, G., Cappelletti, A., Cionni, I., Costa, M., Cremona, G., D'Isidoro, M., Finardi, S., et al. (2014). Assessment of the ams-minni system capabilities to simulate air quality over italy for the calendar year 2005. *Atmospheric Environment*, 84:178–188.

Morgenstern, O., Hegglin, M. I., Rozanov, E., O'Connor, F. M., Abraham, N. L., Akiyoshi, H., Archibald, A. T., Bekki, S., Butchart, N., Chipperfield, M. P., et al. (2017). Review of the global models used within phase 1 of the chemistry–climate model initiative (ccmi). *Geoscientific Model Development*, 10(2):639–671.

Naylor, D. A. and Tahic, M. K. (2007). Apodizing functions for fourier transform spectroscopy. *JOSA A*, 24(11):3644–3648.

Norton, R. and Beer, R. (1977). Errata: New apodizing functions for fourier spectrometry. *JOSA*, 67(3):419–419.

Norton, R. H. and Beer, R. (1976). New apodizing functions for fourier spectrometry. *JOSA*, 66(3):259–264.

Orbe, C., Yang, H., Waugh, D. W., Zeng, G., Morgenstern, O., Kinnison, D. E., Lamarque, J.-F., Tilmes, S., Plummer, D. A., Scinocca, J. F., et al. (2018). Large-scale tropospheric transport in the chemistry–climate model initiative (ccmi) simulations. *Atmospheric Chemistry and Physics*, 18(10):7217–7235.

Parker, S. F., Patel, V., Tooke, P. B., and Williams, K. P. (1991). The effect of apodization and finite resolution on fourier transform raman spectra. *Spectrochimica Acta Part A: Molecular Spectroscopy*, 47(9-10):1171–1178.

Parrish, D. F. and Derber, J. C. (1992). The national meteorological center's spectral statistical-interpolation analysis system. *Monthly Weather Review*, 120(8):1747–1763.

Petzold, A., Thouret, V., Gerbig, C., Zahn, A., Brenninkmeijer, C. A., Gallagher, M., Hermann, M., Pontaud, M., Ziereis, H., Boulanger, D., et al. (2015). Global-scale atmosphere monitoring by in-service aircraft–current achievements and future prospects of the european research infrastructure iagos. *Tellus B: Chemical and Physical Meteorology*, 67(1):28452.

Peuch, V.-H., Engelen, R., Rixen, M., Dee, D., Flemming, J., Suttie, M., Ades, M., Agustí-Panareda, A., Ananasso, C., Andersson, E., et al. (2022). The copernicus atmosphere monitoring service: from research to operations. *Bulletin of the American Meteorological Society*, 103(12):E2650–E2668.

Phulpin, T., Cayla, F., Chalon, G., Diebel, D., and Schlüssel, P. (2002). Iasi on board metop: Project status and scientific preparation. In *12th International TOVS Study Conference, Lorne, Victoria, Australia*, volume 26.

Privé, N., Errico, R., and Tai, K.-S. (2013a). Validation of the forecast skill of the global modeling and assimilation office observing system simulation experiment. *Quarterly Journal of the Royal Meteorological Society*, 139(674):1354–1363.

Privé, N. C., Xie, Y., Woollen, J. S., Koch, S. E., Atlas, R., and Hood, R. E. (2013b). Evaluation of the earth systems research laboratory's global observing system simulation experiment system. *Tellus A: Dynamic Meteorology and Oceanography*, 65(1):19011.

Proedrou, E. and Hocke, K. (2016). Characterising the three-dimensional ozone distribution of a tidally locked earth-like planet. *Earth, Planets and Space*, 68(1):1–20.

Rabier, F., Fourrié, N., Chafäi, D., and Prunet, P. (2002). Channel selection methods for infrared atmospheric sounding interferometer radiances. *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography*, 128(581):1011–1027.

Ricaud, P., Attié, J.-L., Teyssedre, H., El Amraoui, L., Peuch, V.-H., Matricardi, M., and Schluessel, P. (2009). Equatorial total column of nitrous oxide as measured by iasi on metop-a: implications for transport processes. *Atmospheric Chemistry & Physics*, 9(12).

Rodgers, C. D. (1996). Information content and optimization of high-spectral-resolution measurements. In *Optical spectroscopic techniques and instrumentation for atmospheric and space research II*, volume 2830, pages 136–148. International Society for Optics and Photonics.

Rossby, C.-G. (1939). Relation between variations in the intensity of the zonal circulation of the atmosphere and the displacements of the semi-permanent centers of action. *J. mar. Res.*, 2:38–55.

Rossby, C.-G. (1949). On the dispersion of planetary waves in a barotropic: Atmosphere 1. *Tellus*, 1(1):54–58.

Rouil, L., Honore, C., Vautard, R., Beekmann, M., Bessagnet, B., Malherbe, L., Meleux, F., Dufour, A., Elichegaray, C., Flaud, J.-M., et al. (2009). Prev'air: an operational forecasting and mapping system for air quality in europe. *Bulletin of the American Meteorological Society*, 90(1):73–84.

Saunders, R., Hocking, J., Rundle, D., Rayer, P., Havemann, S., Matricardi, M., Geer, A., Lupu, C., Brunel, P., and Vidot, J. (2017). Rttov v12 science and validation report. *available at:* `https://www.nwpsaf.eu/site/download/documentation/rtm/docs_rttov12/rttov12_svr.pdf` *(last access: 12 Nov. 2022).*

Saunders, R., Hocking, J., Turner, E., Rayer, P., Rundle, D., Brunel, P., Vidot, J., Roquet, P., Matricardi, M., Geer, A., et al. (2018). An update on the rttov fast radiative transfer model (currently at version 12). *Geoscientific Model Development*, 11(7):2717–2737.

Saunders, R., Matricardi, M., and Brunel, P. (1999). An improved fast radiative transfer model for assimilation of satellite radiance observations. *Quarterly Journal of the Royal Meteorological Society*, 125(556):1407–1425.

Schulz, M., Tsyro, S., Gauss, M., Valdebenito, A., Mortier, A., Kranenburg, R., and Bennuona, Y. (2019a). Cams71 d3.2.1-2019-01, episode analysis report n°01 in 2019. *Copernicus Atmosphere Monitoring Service (CAMS) report, https://policy.atmosphere.copernicus.eu/reports/CAMSReportJun2019-episode.pdf*.

Schulz, M., Tsyro, S., Gauss, M., Valdebenito, A., Mortier, A., Kranenburg, R., and Bennuona, Y. (2019b). Cams71 d3.2.1-2019-02, episode analysis report n°02 in 2019. *Copernicus Atmosphere Monitoring Service (CAMS) report, https://policy.atmosphere.copernicus.eu/reports/CAMSReportJul2019-episode*.

Seity, Y., Brousseau, P., Malardel, S., Hello, G., Bénard, P., Bouttier, F., Lac, C., and Masson, V. (2011). The arome-france convective-scale operational model. *Monthly Weather Review*, 139(3):976–991.

Shephard, M. and Cady-Pereira, K. (2015). Cross-track infrared sounder (cris) satellite observations of tropospheric ammonia. *Atmospheric Measurement Techniques*, 8(3):1323–1336.

Sherlock, V. and Saunders, R. (1999). Isem-6: Infrared surface emissivity model for rttov-6. *NWP-SAF report*.

Sič, B., Amraoui, L. E., Piacentini, A., Marécal, V., Emili, E., Cariolle, D., Prather, M., and Attié, J.-L. (2016). Aerosol data assimilation in the chemical transport model mocage during the traqa/charmex campaign: aerosol optical depth. *Atmospheric Measurement Techniques*, 9(11):5535–5554.

Sič, B., El Amraoui, L., Marécal, V., Josse, B., Arteta, J., Guth, J., Joly, M., and Hamer, P. (2015). Modelling of primary aerosols in the chemical transport model mocage: development and evaluation of aerosol physical parameterizations. *Geoscientific Model Development*, 8(2):381–408.

Siméoni, D., Singer, C., and Chalon, G. (1997). Infrared atmospheric sounding interferometer. *Acta Astronautica*, 40(2-8):113–118.

Simmons, A. and Strüfing, R. (1981). *An Energy and Angular Momentum Conserving Finite-difference Scheme, Hybrid Co-ordinates and Medium-range Weather Prediction*. ECMWF.

Simpson, D. and Darras, S. (2021). Global soil no emissions for atmospheric chemical transport modelling: Cams-glob-soil v2. 2. *Earth System Science Data Discussions*, pages 1–35.

Sindelarova, K., Granier, C., Bouarar, I., Guenther, A., Tilmes, S., Stavrakou, T., Müller, J.-F., Kuhn, U., Stefani, P., and Knorr, W. (2014). Global data set of biogenic voc emissions calculated by the megan model over the last 30 years. *Atmospheric Chemistry and Physics*, 14(17):9317–9341.

Stockwell, W. R., Kirchner, F., Kuhn, M., and Seefeld, S. (1997). A new mechanism for regional atmospheric chemistry modeling. *Journal of Geophysical Research: Atmospheres*, 102(D22):25847–25879.

Strow, L. L., Hannon, S. E., De Souza-Machado, S., Motteler, H. E., and Tobin, D. (2003). An overview of the airs radiative transfer model. *IEEE Transactions on Geoscience and Remote Sensing*, 41(2):303–313.

Stull, R. B. (1988). *An introduction to boundary layer meteorology*, volume 13. Springer Science & Business Media.

Tarrasón, L., Hamer, P. D., Meleux, F., Colette, A., and Rouïl, L. (2020). Interim annual assessment report for 2019, european air quality in 2019. *ECMWF Copernicus Report,* `https://policy.atmosphere.copernicus.eu/reports/CAMS71_IAR_2019_final.pdf`.

Teyssèdre, H., Michou, M., Clark, H., Josse, B., Karcher, F., Olivié, D., Peuch, V.-H., Saint-Martin, D., Cariolle, D., Attié, J.-L., et al. (2007). A new tropospheric and stratospheric chemistry and transport model mocage-climat for multi-year studies: evaluation of the present-day climatology and sensitivity to surface processes. *Atmospheric chemistry and physics*, 7(22):5815–5860.

Timmermans, R. M., Lahoz, W., Attié, J.-L., Peuch, V.-H., Curier, R., Edwards, D., Eskes, H., and Builtjes, P. (2015). Observing system simulation experiments for air quality. *Atmospheric Environment*, 115:199–213.

Tkemaladze, G. S. and Makhashvili, K. (2016). Climate changes and photosynthesis. *Annals of agrarian science*, 14(2):119–126.

Van Vuuren, D. P., Edmonds, J. A., Kainuma, M., Riahi, K., and Weyant, J. (2011). A special issue on the rcps. *Climatic Change*, 109(1):1–4.

Vittorioso, F., Guidard, V., and Fourrié, N. (2021). An infrared atmospheric sounding interferometer–new generation (iasi-ng) channel selection for numerical weather prediction. *Quarterly Journal of the Royal Meteorological Society*, 147(739):3297–3317.

Wark, K. and Warner, C. F. (1981). Air pollution: its origin and control.

Watson, L., Lacressonnière, G., Gauss, M., Engardt, M., Andersson, C., Josse, B., Marécal, V., Nyiri, A., Sobolowski, S., Siour, G., et al. (2016). Impact of emissions and+ 2 c climate change upon future ozone and nitrogen dioxide over europe. *Atmospheric environment*, 142:271–285.

Weston, P., Bell, W., and Eyre, J. (2014). Accounting for correlated error in the assimilation of high-resolution sounder data. *Quarterly Journal of the Royal Meteorological Society*, 140(685):2420–2429.

Williams, J. E., Huijnen, V., Bouarar, I., Meziane, M., Schreurs, T., Pelletier, S., Marécal, V., Josse, B., and Flemming, J. (2022). Regional evaluation of the performance of the global cams chemical modeling system over the united states (ifs cycle 47r1). *Geoscientific Model Development*, 15(12):4657–4687.

WMO (2022). Scientific assessment of ozone depletion: 2022. executive summary. gaw report no. 278, 56 pp.; world meteorological organization (wmo): Geneva, 2022. `https://library.wmo.int/doc_num.php?explnum_id=11377`.

Worden, H. M., Bowman, K. W., Worden, J. R., Eldering, A., and Beer, R. (2008). Satellite measurements of the clear-sky greenhouse effect from tropospheric ozone. *Nature Geoscience*, 1(5):305–308.

Yang, J., Zhang, Z., Wei, C., Lu, F., and Guo, Q. (2017). Introducing the new generation of chinese geostationary weather satellites, fengyun-4. *Bulletin of the American Meteorological Society*, 98(8):1637–1658.

Zeng, X., Atlas, R., Birk, R. J., Carr, F. H., Carrier, M. J., Cucurull, L., Hooke, W. H., Kalnay, E., Murtugudde, R., Posselt, D. J., et al. (2020). Use of observing system simulation experiments in the united states. *Bulletin of the American Meteorological Society*, 101(8):E1427–E1438.

Zeng, Z.-C., Lee, L., and Qi, C. (2022). Diurnal carbon monoxide observed from a geostationary infrared hyperspectral sounder: First result from giirs onboard fy-4b. *Atmospheric Measurement Techniques Discussions*, pages 1–32.

Zoogman, P., Liu, X., Suleiman, R., Pennington, W., Flittner, D., Al-Saadi, J., Hilton, B., Nicks, D., Newchurch, M., Carr, J., et al. (2017). Tropospheric emissions: Monitoring of pollution (tempo). *Journal of Quantitative Spectroscopy and Radiative Transfer*, 186:17–39.

Appendix A: Pressure Equivalence for MOCAGE Vertical Levels

Level	Pressure Equivalence (hPa)		
	Min.	Median	Max.
1	0.1	0.1	0.1
2	0.3	0.3	0.3
3	0.5	0.5	0.5
4	0.8	0.8	0.8
5	1.1	1.1	1.1
6	1.6	1.6	1.6
7	2.1	2.1	2.1
8	2.7	2.7	2.7
9	3.4	3.4	3.4
10	4.4	4.4	4.4
11	5.6	5.6	5.6
12	7.0	7.0	7.0
13	8.6	8.7	8.7
14	10.5	10.6	10.7
15	12.5	12.8	12.8
16	14.8	15.2	15.3
17	17.2	17.9	18.0
18	19.7	20.8	20.9
19	22.5	23.9	24.1
20	25.5	27.2	27.5
21	28.6	30.9	31.2
22	31.9	34.9	35.3
23	35.5	39.4	39.9
24	39.6	44.6	45.2
25	44.3	50.4	51.2
26	49.4	57.1	58.1
27	55.2	64.7	65.9
28	61.6	73.3	74.8
29	68.7	83.0	84.9
30	76.6	94.1	96.3

Level	Pressure Equivalence (hPa)		
	Min.	Median	Max.
31	85.4	106.4	109.1
32	95.2	120.4	123.6
33	106.1	136.3	140.1
34	118.3	154.3	158.9
35	131.9	174.7	180.1
36	147.0	197.7	204.2
37	163.9	223.8	231.3
38	182.6	252.9	261.9
39	203.0	285.2	295.6
40	225.2	320.7	332.8
41	249.4	359.6	373.6
42	275.5	402.4	418.4
43	302.9	447.4	465.7
44	330.1	492.7	513.3
45	357.3	538.1	561.0
46	384.0	583.2	608.5
47	410.2	627.7	655.2
48	435.8	671.2	701.1
49	460.5	713.5	745.6
50	484.2	754.3	788.5
51	506.7	793.1	829.4
52	527.9	829.7	868.0
53	547.5	863.8	903.9
54	565.5	895.0	936.8
55	581.6	923.1	966.3
56	595.8	947.6	992.2
57	607.7	968.3	1014.0
58	617.3	984.9	1031.5
59	624.4	997.1	1044.3
60	628.9	1004.4	1052.0

Table A1: Minimum, median and maximum value of pressure encountered over the domain MACC01 (28°N, 26°W and 72°N, 46°E) for each of the 60 MOCAGE vertical levels.

Appendix B: Statistical Indices

This Appendix contains the definitions of the statistical indices used throughout the whole manuscript.

The relative difference of a with respect to b is computed as:

$$D_{rel} = \frac{a-b}{b} \tag{B1}$$

Given a population N, the **average** of the variable a can be computed as:

$$\overline{a} = \frac{1}{N}\sum_{i}^{N}(a_i) \tag{B2}$$

The **standard deviation** associated to the same variable and population is:

$$\sigma_a = \sqrt{var(a)} \tag{B3}$$

while the **variance**:

$$var(a) = \frac{1}{N}\sum_{i}^{N}(a_i - \overline{a}_i)^2 \tag{B4}$$

The **covariance** between two variables a and b is defined as:

$$cov(a,b) = \frac{1}{N}\sum_{i}^{N}(a_i - \overline{a}_i)(b_i - \overline{b}_i) \tag{B5}$$

The **mean bias (MB)** provides an indication of the difference between two data sets. It can range in $(-\infty, +\infty)$ and it is defined by the following formula:

$$\mathrm{MB} = \frac{1}{N}\sum_{i}^{N}(a_i - b_i) \tag{B6}$$

where a is the parameter to be evaluated through comparison to the reference b over a population N.

The **root mean square error (RMSE)** is the standard deviation of the errors occurring between a and the reference b. RMSE gives a measure of how these errors spread out. It is

defined by the formula

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i}^{N}(a_i - b_i)^2} \tag{B7}$$

and it ranges between $[0, +\infty)$. This indicator is strongly dominated by the largest values because of the square root operation. Accordingly, the interpretation of the RMSE and its reliability may be tricky whenever major outliers are present.

Both MB and RMSE need the knowledge of typical mean values of the involved quantities in order to be well interpreted. These two parameters can be made dimensionless and relative by scaling them to the population of reference. The resulting indicators are the MNMB and the fractional gross error (FGE), that are sometimes more reliable when interpreting scores of species with large divergence in order of magnitude.

The **MNMB** is the MB normalised by the mean of the two involved quantities (modelled a and reference b):

$$\mathrm{MNMB} = \frac{2}{N}\sum_{i}^{N}\left(\frac{a_i - b_i}{a_i + b_i}\right). \tag{B8}$$

It bounds the bias in between -2 and $+2$. The closer the value is to 0, the smaller the difference between the quantities.

The **fractional gross error (FGE)**, on the other hand, is defined as

$$\mathrm{FGE} = \frac{2}{N}\sum_{i}^{N}\left|\frac{a_i - b_i}{a_i + b_i}\right| \tag{B9}$$

and it ranges in $[0, 2]$. It is often associated to the RMSE since it doesn't emphasize the extremely out-of-range values.

Finally, the **Pearson's correlation coefficient** can be used to test the match in the patterns of the modelled populations. It is computed through the formula:

$$\rho = \frac{cov(a,b)}{\sigma_a \sigma_b} = \frac{\frac{1}{N}\sum_{i}^{N}(a_i - \bar{a})\left(b_i - \bar{b}\right)}{\sigma_a \sigma_b} \tag{B10}$$

with $\bar{a}$ and $\bar{b}$ mean values of the tested quantities and σ_a and σ_b the associated standard deviations. This coefficient, ranging in $[-1, 1]$, reaches its Max. value of 1 when $(a_i - \bar{a}) = c\left(b_i - \bar{b}\right)$, with c a positive constant. In that case, the data sets being compared present the same kind of variability, but they are not identical unless the constant c has a value of unity all over each site.

Appendix C: Degrees of Freedom for Signal

The *Degrees of Freedom for Signal* (DFS) [e.g. Rodgers (1996); Rabier et al. (2002); Vittorioso et al. (2021)], is defined as:

$$\text{DFS} := \text{Tr}\ (\mathbf{I}\ -\ \mathbf{A}\mathbf{B}^{-1}) \tag{C1}$$

where $\mathbf{I}$ is the identity matrix, while $\mathbf{A}$ and $\mathbf{B}$ the analysis and background-error covariance matrices, respectively.
Since the $\mathbf{A}$ matrix, in the linear case, can be expressed as:

$$\mathbf{A} = (\mathbf{B}^{-1} + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H})^{-1} \tag{C2}$$

with $\mathbf{H}$ the Jacobian matrix, i.e. the linearized version of the observation operator, and $\mathbf{H}^{\mathrm{T}}$ its transpose.
The $\mathbf{A}\mathbf{B}^{-1}$ term becomes:

$$\begin{aligned} \mathbf{A}\mathbf{B}^{-1} &= (\mathbf{B}^{-1} + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{B}^{-1} \\ &= [\mathbf{B}\ (\mathbf{B}^{-1} + \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H})]^{-1} \\ &= [\mathbf{I} + \mathbf{B}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}]^{-1} \end{aligned} \tag{C3}$$

Consequently, Equation 11 can also be written as follows:

$$\text{DFS} = \text{Tr}\ [\mathbf{I}\ -\ (\mathbf{I}\ +\ \mathbf{B}\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H})^{-1}]. \tag{C4}$$

Acknowledgements

Tout d'abord, je tiens à remercier ***Thales Alenia Space*** *et* ***Météo France*** *qui ont financé ce projet.*

Un grand merci va à tous les membres du jury de thèse qui ont pris le temps de lire le manuscrit. Merci en particulier aux ***rapporteurs*** *pour leurs corrections qui ont contribué à rendre la lecture du manuscrit plus claire et qui ont apporté des idées intéressantes pour d'éventuels futurs développements de ce projet.*

Je remercie les membres de mon ***comité de thèse****, qui ont fourni des conseils précieux pour le déroulement de cette thèse.*

Je remercie également les équipes PLASMA et COMETS qui m'ont accompagné dans la réalisation de ce travail. Tout le monde a été toujours prêt à répondre aux questions techniques et scientifiques, toujours prêt à aider. Chacun d'entre eux a partagé son expertise, m'aidant à me sentir accompagné dans ce parcour.

Je tiens à remercier ***Jérôme VIDOT*** *qui, d'abord co-directeur, puis encadrant, a été capable, dans les deux rôles, d'apporter un soutien important à la thèse et a toujours été prêt à partager sa précieuse expertise et à me conseiller au mieux.*

Je suis très reconnaissante à ***Marine CLAEYMAN****, ma co-directrice de thèse, et* ***Francis OLIVIER*** *de m'avoir confié ce projet. Ce fut une grande satisfaction professionnelle de travailler avec Thales. Malheureusement, en raison des périodes de pandémie, je n'ai pas pu travailler à leurs côtés en présenciel comme il était initialement prévu. Cependant, tous deux ne se sont pas découragés et ont toujours réussi à me suivre parfaitement même à distance.*

Et enfin, je tiens à remercier du fond du cœur ma directrice de thèse, ***Nadia FOURRIÉ****, et mon co-directeur,* ***Vincent GUIDARD****. Ils ont tous les deux cru en moi plus que je ne l'ai fait. Je travaille avec eux depuis plus de six ans et j'ai pu apprendre énormément à leur côté.*
Nadia*, malgré son emploi de temps, n'a jamais été absente dans la poursuite de cette thèse. Au contraire, elle s'est toujours organisée pour être présente au maximum, toujours disponible pour transmettre son expértise et à apporter une contribution précieuse à l'ensemble du projet. En même temps elle a toujours été à l'écoute de mes préoccupations, de mes doutes et de mes attentes. Nadia a été pour moi, au fil des ans, un modèle de professionnalisme et d'humanité, auquel j'aspirerai toujours en tant que femme et scientifique.*
Et puis ***Vincent****, qui est lui aussi mon encadrant depuis le jour de mon arrivée à Toulouse en 2016. Il a toujours été à mon côté pour m'apprendre et me guider. Je ne pense pas que*

quelqu'un puisse remplir ce rôle mieux que lui. Un scientifique extrêmement compétent, il sait se remettre continuellement en question pour acquérir de nouvelles connaissances. Aussi, il montre une énorme empathie. Vincent est la personne que vous souhaitez avoir à vos côtés en temps de crise :) Il sait comment gérer les situations jusqu'à ce que le calme revienne. Sans lui, ce projet n'aurait pas vu le jour. Et pendant qu'on travaille, à côté la vie suit son cours. Pendant ces six ans, j'ai eu des bons et des mauvais moments et Vincent a toujours été la pour m'ecouter, comme un chef, mais aussi comme un grand frère. Je ne pourrai jamais le remercier assez.

Et après voici quelques remerciements supplémentaires...

Merci à ***Flavien****, pour les séances de rédaction et pour avoir partagé les deadlines avec moi :) Merci à* ***Clément*** *et* ***Olivier****, pour la collaboration scientifique et leur compagnie de bureau.*

Merci à tous les amis qui, avec un peu chocolat, un SMS, un appel d'urgence ou un mot de soutien, ont réussi à me donner la force d'aller jusqu'au bout de ce doctorat.

Un merci spécial à ***Filipa*** *et* ***Robin****, qui m'ont accueilli chez eux alors que je n'avais pas de toit sur la tête... ou plutôt, alors que j'en avais un sur mon lit :)*
Et merci à ***Anne-Lise****, qui me fait sentir comme un membre de sa famille.*

Many thanks to all of **my other friends, in Toulouse and around the world**, that I will not list one by one, because another manuscript would be needed.

Grazie ai miei ***amici in Italia****. Sono stati preziosi, come sempre, nel supportarmi e nel credere in me in un periodo così faticoso e, spesso, frustrante. Grazie per la paziena.*

Ovviamente, grazie alla mia famiglia, sulla quale so di poter sempre contare. Vi voglio bene.

And last, but not least...:

«E quindi uscimmo a riveder le stelle[5]*.»*

Dante Alighieri, *INFERNO, Canto XXXIV*

[5] *Thence we came forth to rebehold the stars.*