



HAL
open science

Amélioration de l'estimation de la distribution spatiale des aérosols atmosphériques à l'aide des observations satellites et de l'apprentissage automatique

Farouk Lemmouchi

► **To cite this version:**

Farouk Lemmouchi. Amélioration de l'estimation de la distribution spatiale des aérosols atmosphériques à l'aide des observations satellites et de l'apprentissage automatique. Ingénierie de l'environnement. Université Paris-Est Créteil Val-de-Marne - Paris 12, 2022. Français. NNT : 2022PA120077 . tel-04892257

HAL Id: tel-04892257

<https://theses.hal.science/tel-04892257v1>

Submitted on 16 Jan 2025

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

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



Thèse pour l'obtention du titre de docteur de l'Université Paris-Est Créteil

École doctorale Sciences Ingénierie et Environnement SIE-531

Faculté de sciences et de technologies

Laboratoire Interuniversitaire des Systèmes Atmosphériques

LEMMOUCHI Farouk

**Amélioration de l'estimation de la distribution spatiale des
aérosols atmosphériques à l'aide des observations satellitaires et de
l'apprentissage automatique**

*Improvement of the spatial distribution estimation of atmospheric aerosols using
satellite observations and machine learning*

Membres du jury :

Mme DUFOUR Gaëlle	Présidente	CNRS, LISA
M. LEBBAH Mustapha	Rapporteur	UVSQ
Mme LAW Kathy	Rapporteuse	CNRS, LATMOS
M. CEAMANOS Xavier	Examineur	Météo France, CNRM
M. BRAJARD Julien	Examineur	NERSC
M. DEROGNAT Claude	Invité	Aria Technologies
M. CUESTA Juan	Directeur de thèse	UPEC, LISA

Remerciements

Je tiens tout d'abord à exprimer ma gratitude envers les organismes qui ont financé ce travail : la Région Ile-de-France dans le cadre du programme Paris Region PhD (PRPHD), à travers le Domaine d'Intérêt Majeur DIMQi2, en partenariat avec l'entreprise ARIA Technologies.

Je souhaite remercier mon directeur de thèse, Juan Cuesta, d'avoir cru en ma capacité de réaliser le projet. J'ai apprécié ses qualités d'écoute, compréhension et sa patience tout au long de ma thèse. Ses multiples consignes et l'encouragement m'ont énormément aidé pour réaliser ce projet.

Je tiens à remercier les membres du jury qui ont évalué et jugé mon travail. Merci à Gaëlle Dufour d'avoir accepté de présider le jury de soutenance. Mes sincères remerciements vont à Kathy Law et à Mustapha Lebba pour la relecture critique et constructive de ce manuscrit en tant que rapporteurs. Je remercie Julien Brajard, et Xavier Ceamanos pour avoir accepté d'examiner et d'évaluer mon travail durant la soutenance, et pour leurs remarques et suggestions scientifiques pertinentes en étant membres de mon comité de suivi de thèse.

Durant ma thèse, j'ai bénéficié de la collaboration de plusieurs personnes du LISA dont je suis reconnaissant, Maxim Eremenko, Pasquale Sellito, Guillaume Siour, Matthias Beekmann, et Adriana Coman, en plus des chercheurs d'autres institutions qui ont été aussi coopératifs notamment Mathieu Lachatre et Claude Derognat d'Aria Technologies, Solène Turquety, Dung Tran, Xiong Liu, Peter Zoogman, Ronny Lutz, Diego Loyola, en plus de Maarten Sneep et Veeffkind Pepijn.

Un grand merci à mes collègues et amis avec qui j'ai passé des moments de convivialité inoubliables, à mes premières rencontres et mes sources de motivation Safa, Dilek, Henda, Anussanth, Thithuy, Rebecca, Youstra, Sachiko, et Durgesh. Merci à notre responsable du groupe Spectroscopie Atmosphérique, Pascale Chelin pour sa gentillesse, l'encouragement, et d'avoir été à la hauteur des responsabilités. Je salue la bienveillance des collègues du groupe, notamment Martin Schwell et Nguyen Lam.

Je souligne les bonnes conditions de travail aux sein du LISA, merci aux personnels : Fazia Radi, Isabelle Rosetzky, Mustapha El-Akhel, Pascale Large pour la serviabilité et la bonne humeur. Je voudrais également remercier les personnels de l'Université Paris-Est Tiana Ramarosan, Fatima de Freitas, et Christelle Berol pour leurs réactivité, merci également à Gilles Forêt et Florence Brouillaud pour la bienveillance au sein de la DIMQi2, et à Subu Ramachandran pour le support.

Ma dernière année n'aurait pas été aussi amusante sans mes nouveaux collègues de bureau, je remercie chaleureusement Prem, Redha, Kaiwen, et Haoyue pour l'amitié et le soutien apporté. Enfin, j'ai aimé les quelques moments partagés avec les jeunes chercheurs que j'ai pu croiser et socialiser, merci pour l'ouverture d'esprit et la cordialité de tous.

Enfin, je remercie vivement l'ensemble de ma famille et mes amis, toujours présents aux moments difficiles.

Avant-propos

Les deux premiers chapitres constituent des introductions scientifiques au travail présenté au Chapitre III et IV. Ces deux derniers sont des versions étendues des articles rédigés et publiés dans le journal *Remote Sensing* de l'éditeur *Multidisciplinary Digital Publishing Institute* (MDPI) dont les liens sont : <https://doi.org/10.3390/rs14112582> et <https://doi.org/10.3390/rs15061510> respectivement. Par commodité, j'ai gardé la langue anglaise pour ces deux chapitres en incluant un résumé en français au début de chacun. Des figures d'illustrations annotées en anglais sont aussi présentes dans ce manuscrit, nous prions d'avance le lecteur d'excuser l'éventuelle gêne induite par le passage d'une langue à l'autre au cours de la lecture du document.

Résumé

Les particules de matière en suspension dans l'air appelées aérosols atmosphériques sont connues pour leurs effets nocifs sur la santé et l'environnement en plus de la perturbation du système climatique. Afin de connaître le cycle de vie des aérosols il faut étudier leurs origines, leurs réactivités, et leurs transports par le vent. Différentes méthodes plus ou moins précises existent pour estimer l'abondance des particules de matière. Les approches satellitaires fournissent classiquement des observations en deux dimension (2D), c'est-à-dire la distribution horizontale ou en lignes de profils verticaux (transects). Les modèles de chimie-transport simulent la distribution tridimensionnelle (3D) des particules, mais une grande incertitude est liée à ces simulations, surtout dans les régions qui manquent d'instruments de mesures in situ pour contraindre la modélisation.

L'objectif de ma thèse est d'améliorer les estimations existantes de la distribution spatiale des aérosols en exploitant les données satellitaires, les modèles de chimie-transport et l'apprentissage automatique. La première partie de mes travaux a permis le développement de la première méthode capable d'observer la distribution 3D des particules fines. L'algorithme appelé AEROS5P utilise les parties du spectre de réflectance terrestre visible et proche infrarouge mesuré par l'instrument satellitaire TROPOMI afin de déduire des profils d'extinction des aérosols. Ces derniers sont comparables avec des mesures indépendantes de haute résolution verticale issues de la télédétection active de lidar.

Dans une deuxième partie, j'ai développé une méthode combinant les avantages du modèle de chimie-transport CHIMERE et des observations satellitaires de l'instrument MODIS pour réduire les biais des champs simulés de l'épaisseur optique des aérosols. Cette correction est appliquée a posteriori à l'aide de modèles d'apprentissage automatique supervisé. Les nouvelles estimations de l'épaisseur optique ont moins de biais et elles sont comparables aux mesures des stations d'observations au sols AERONET.

Mots clés : particules de matière, aérosols, télédétection, TROPOMI, épaisseur optique des aérosols, feux de biomasse, restitution d'aérosols, profil vertical d'extinction, feu australien, hauteur d'injection des fumées, poussière saharienne, dépression de Bodélé, correction de biais, apprentissage supervisé, modèle de chimie-transport.

Abstract

Particulate matter suspended in the air called atmospheric aerosols are known for their harmful effects on health and the environment in addition to the disruption of the climate system. In order to know the life cycle of aerosols, it is necessary to study their origins, their reactivity, and the way in which they are transported by wind. Different methods exist to estimate the abundance of particulate matter. Satellite approaches typically provide observations in two dimensions (2D), i.e. horizontal distribution or vertical profiles transects. Chemistry-transport models simulate the three dimensional (3D) distribution of particles, but large uncertainty may be associated with it, especially in regions lacking ground stations monitoring instruments to constrain the simulation.

The objective of my thesis is to improve existing estimates of the spatial distribution of aerosols by leveraging satellite data, chemistry-transport models and artificial intelligence. The first part of the work is the development of the first method capable of observing the 3D distribution of fine particles. The algorithm, called AEROS5P, uses the visible and near infrared parts of the terrestrial reflectance spectra measured by the satellite instrument TROPOMI in order to retrieve aerosol extinction profiles. These estimations are comparable with independent measurements of high vertical resolution from active lidar remote sensing.

In a second part, I developed a method combining the advantages of the CHIMERE chemistry-transport model and the satellite observations of the MODIS instrument to reduce the biases of the simulated aerosol optical depth fields. This correction is applied a posteriori using supervised machine learning models. The new aerosol optical depths estimates have less bias and are comparable to measurements from AERONET ground observing stations.

Keywords: Particulate matter, aerosols, remote sensing, TROPOMI, aerosol optical depth, biomass burning, aerosol retrieval, vertical extinction profile, Australian fire, smoke injection height, Saharan dust, Bodélé Depression, bias correction, supervised learning, chemistry-transport model.

Table des matières

Table des figures	13
Liste des tableaux.....	19
Acronymes	21
Introduction de la thèse	25
I. Chapitre 1 : Aérosols dans l’atmosphère	29
1. Généralité sur l’atmosphère terrestre	31
1.1. La troposphère	32
1.2. La stratosphère	33
1.3. La mésosphère	33
1.4. Dynamique de l’atmosphère	33
2. Aérosols atmosphériques	36
2.1. Classification des aérosols	38
2.2. Mécanismes de formation des aérosols atmosphériques	39
2.2.1. Aérosols issus de combustion	39
2.2.2. Poussière minérale	40
2.2.3. Aérosols de sels marins.....	40
2.2.4. Aérosols secondaires.....	40
2.3. Durée de vie des principaux aérosols atmosphériques	42
2.4. Impacts des aérosols	43
2.4.1. Effets sur la santé	43
2.4.2. Effets sur l’environnement.....	44
3. Interaction des particules avec le rayonnement	46
3.1. Le rayonnement électromagnétique	47
3.2. Absorption.....	49
3.3. Diffusion simple.....	50
3.3.1. Diffusion Rayleigh.....	51
3.3.2. Diffusion de Mie	52
3.3.3. Albédo de diffusion simple.....	54
4. Atténuation du rayonnement dans l’atmosphère.....	55
4.1. Épaisseur optique	57
5. Conclusions.....	57
II. Chapitre 2 : Quantification des aérosols dans l’atmosphère	59
1. Mesures in situ	61
2. Télédétection.....	62
2.1. Technique active	62
3.2. Technique passive	64
3.2.1. Capteur au sol	64
3.2.2. Capteurs embarqués	65
4. Observation des aérosols depuis l’espace	71
4.1. Méthode d’inversion de mesures passives de luminance	72
4.2. Équation de transfert radiatif	74
4.2.1. Modélisation direct du transfert radiatif	75
4.2.2. Inversion	78
4.2.3. Évaluation qualitative de la qualité de l’inversion.....	79

4.2.4.	Analyse des erreurs de l'inversion	80
4.2.5.	Régularisation Tikhonov-Phillips	81
5.	Modélisation régionale de la qualité de l'air.....	81
5.1.	Le modèle de chimie-transport CHIMERE	82
5.2.	Émission.....	83
5.3.	Advection.....	83
5.4.	Dépôt sec et humide.....	83
5.5.	Chimie.....	84
5.6.	Propriétés optiques.....	84
6.	Conclusion	85
III.	Chapitre 3 : Télédétection tridimensionnelle des aérosols fins	87
	Résumé.....	89
1.	Three-dimensional distribution of biomass burning aerosols:.....	91
2.	Materials and methods	94
2.1.	The AEROS5P method.....	94
2.2.	Inputs.....	96
2.3.	Forward calculations of reflectance spectra.....	103
2.4.	Inversion procedure	105
2.5.	Other aerosol observations used for comparisons with AEROS5P	107
3.	Comparison of AEROS5P retrievals with other aerosol products.....	109
3.1.	Aerosol optical depth	109
3.2.	Aerosol extinction profiles.....	114
3.3.	Daily evolution of the 3D distribution of biomass burning aerosol plumes	115
4.	Conclusion	120
IV.	Chapitre 4 : Synergie des observations satellitaires et la simulation de chimie-transport	123
	Résumé.....	125
1.	Synergy of satellite observations and chemistry-transport simulation	129
2.	AOD correction approach.....	130
3.	Inputs.....	131
3.1.	MODIS satellite observations	131
3.2.	CHIMERE simulations	132
3.3.	Dataset preparation	136
4.	Bias correction ML models construction.....	138
4.1.	Multiple linear regression	139
4.2.	Feed-forward neural networks	140
4.3.	Random forest.....	142
4.4.	Gradient boosting.....	143
5.	Results and discussion	143
5.1.	Comparison against independent MODIS observations	144
5.2.	Comparison with AERONET ground based measurements.....	154
6.	Conclusions.....	156
V.	Conclusion générale et perspectives	159
	Annexe A	163
	Annexe B	165
	Annexe C	173
	Références.....	177

Table des figures

Figure I-1 Profil vertical de température de l'atmosphère terrestre montrant la stratification des couches atmosphériques [11].....	31
Figure I-2 Développement de la couche limite atmosphérique pendant une journée [13,16].	33
Figure I-3 Moyenne zonale annuelle du flux radiatif solaire incident et émis par la terre au sommet de l'atmosphère tiré des mesures satellitaires [18].....	34
Figure I-4 Estimation de la différence entre l'insolation et l'émission radiatif terrestre entre 1985 et 1989 par imagerie satellitaire [22].	35
Figure I-5 Image montrant les différents phénomènes mise en jeu dans la circulation atmosphériques [19]......	35
Figure I-6 Distribution de taille des différents aérosols atmosphériques [24].	37
Figure I-7 Particule composée de sel marin et de sulfates [8].....	38
Figure I-8 Les types de mélange d'aérosols atmosphériques. Adaptée depuis la référence [26]......	39
Figure I-9 Processus de formation des aérosols sulfatés.[39].	41
Figure I-10 Nucléation des particules dans une forêt boréale pendant une journée, concentration mesuré à 2 m de hauteur à la forêt de Hyytiälä en 19-05-1999. D_p est le rayon de la particule et N le nombre de particule par unité de surface [40].	42
Figure I-11 Dégradation de la visibilité à Paris lors de l'épisode de pollution du 14 mars 2014 causée par les PMs. © AFP. www.lexpress.fr	44
Figure I-12 Estimation de la contribution au forçage radiatif effectif de la planète pour quelques émissions entre 1970 et 2019 [61]......	46
Figure I-13 Distribution d'énergie en fonction de la longueur d'onde pour quelques corps de température différentes [64]......	48
Figure I-14 Le spectre électromagnétique solaire	48
Figure I-15 Spectre d'absorption de quelques espèces atmosphériques [65]......	49
Figure I-16 Paramètre de taille x en fonction du rayon de la particule et de la longueur d'onde. Le type du rayonnement et de la particule sont montrés en haut et à droite respectivement [66]......	50
Figure I-17 Fonction de phase à 865 nm pour un aérosol grossier C, et trois aérosols fins S1, S2 et S3 [67]......	51

Figure I-18 Fonction de phase de la diffusion Rayleigh pour un rayonnement polarisé et non polarisé [69].	52
Figure I-19 Fonction de phase calculée avec la théorie de Mie pour trois paramètres de tailles différents [72].	53
Figure I-20 Facteur d'asymétrie des aérosols observés par la station Birdsville en Australie pour quatre heures de la journée et pour quatre longueur d'onde [73].	54
Figure I-21 Quelques valeurs d'albédo de diffusion simple de quelques types d'aérosols atmosphériques [74].	55
Figure I-22 Spectre d'indice de réfraction complexe de quelques composants minéraux de poussière dans la gamme spectrale de 2.5 à 25 μm , partie réelle tracée dans le panel du haut et la partie imaginaire en bas [36].	56
Figure II-1 Image de synthèse du satellite CALIPSO [78].	63
Figure II-2 Transects des profils de (a) rétrodiffusion atténuée des aérosols observée par CALIOP embarqué sur CALIPSO et (b) types d'aérosols déduits des mesures CALIOP, le 26 juin 2019. [échange privé J. Cuesta, 2022].	64
Figure II-3 (a) Carte de la distribution globale des stations du réseau AERONET [80], (b) Image d'un photomètre solaire du réseau AERONET [81].	65
Figure II-4 Illustration d'un instrument en orbite géostationnaire (à gauche) et en orbite polaire (à droite). Adaptée depuis la référence [83].	66
Figure II-5 Différents systèmes de d'observation d'aérosols depuis l'espace [84].	67
Figure II-6 Spectre de l'irradiance solaire au sommet de l'atmosphère (en jaune) et au niveau du sol (en rouge) montrant les fenêtres spectrales atmosphériques [85].	67
Figure II-7 Image de synthèse de la plateforme Aqua (crédit :NASA).	68
Figure II-8 AOD moyenne pour le mois d'aout 2021 vue par MODIS Terra [89].	69
Figure II-9 Image d'un modèle du satellite S5P [Wikipedia].	69
Figure II-10 Carte de la colonne totale moyenne de NO_2 pour le mois d'avril 2018 [95].	71
Figure II-11 Illustration de l'angle d'incidence et l'angle de diffusion dans les coordonnées polaires. Adaptée depuis la référence [98].	74
Figure II-12 Schéma de modélisation du transfert radiatif par le modèle direct.	77
Figure III-1 Scheme of the AEROS5P algorithm for retrieving the 3D distribution of fine BB aerosols. r : aerosol size distribution modal radius; s : aerosol size distribution standard deviation; h_{mean} : aerosol layer mean height; α_{550} : aerosol extinction profile at 550 nm; h_{top} : aerosol layer top height; AOD_{550} : aerosol optical depth at 550 nm.	95

Figure III-2 Scheme describing the aerosol altitude retrieval principle. (a) radiation path length traversing the atmosphere in the presence and absence of aerosols. (b) Oxygen A-band absorption band intensity in different aerosol altitude situations.	96
Figure III-3 Irradiance (bleu) and radiance (green) spectrum measurement by TROPOMI and their ratio corresponding to reflectance (red) spectrum with microwindows highlighted in grey that are used in AEROS5P algorithm.	98
Figure III-4 TROPOMI minimum (diffuse) surface LER reflectivity map of the month of December over Australia shown for 4 discrete wavelengths; a) 402 nm; b) 495 nm; c) 678; and d) 754 nm which are close to that used in AEROS5P.	99
Figure III-5 VIIRS AOD ₅₅₀ nm per aerosol type from 20 to 24 December 2019.	100
Figure III-6 Distribution de taille d'aérosols observée par la station Birdsville en Australie du réseau AERONET pour la journée 29 décembre 2019 [80].	101
Figure III-7 (a) A random example of a pixel from TROPOMI showing climatological surface LER data that are provided discontinuously at 8 wavelengths. (b): Interpolated surface LER data with high spectral resolution used in AEROS5P (c) High resolution 6 principal components of surface albedo database used for the interpolation, that are developed by [192].	105
Figure III-8 Reflectance spectral fit of a typical cloud free pixel observation.....	107
Figure III-9 Horizontal distribution of AOD at 550 nm from (a,c) AEROS5P and (b,d) MODIS using the MAIAC algorithm on (a,b) 21 and (c,d) 22 December 2019 over Australia.	110
Figure III-10 Comparison of AOD at 550 nm measured at 7 AERONET stations in Australia by sun photometer with respect to three satellite products: AEROS5P, MODIS/Aqua and VIIRS on 20 to 24 December 2019 (one panel per day from the upper to lower one).....	111
Figure III-11 AEROS5P AOD 550 nm from 20 to 24 December 2019.....	112
Figure III-12 Scatterplot comparison of AOD at 550 nm derived from AEROS5P with respect to co-located retrievals from (a) MODIS/AQUA and (b) VIIRS for the period 20–24 December 2019. Colors in panel (a) represent the number of points for each sector of 0.1×0.1 of AOD. For (a,b), purple lines are linear regression lines of the scatter of points, dotted lines are the 1:1 curve and blue shade is the $\pm 20\%$ interval. (c) Frequency histogram of differences of AOD at 550 nm of MODIS/AQUA and VIIRS with respect to AEROS5P. Symbols and colors in panel (b) indicate the aerosol type derived from the VIIRS approach.....	113
Figure III-13 Comparison of transects of aerosol extinction profiles (in km^{-1}) derived from (a,b) the CALIOP spaceborne lidar and (c,d) the AEROS5P passive approach on (a,c) 21	

December and (b,d) 22 December 2019. Dashed lines indicate the mean altitude of the aerosol layers defined as the ones with half of the AOD above and below. The corresponding AOD at 550 nm from MODIS MAIAC and AEROS5P, and CALIOP at 532 nm is shown in panels (e,f)..... 115

Figure III-14 (a) Three-dimensional distribution of biomass burning aerosols on 20 December 2019 derived from TROPOMI measurements using the AEROS5P method, depicted as iso-aerosol extinction surface of 0.1 km^{-1} . (b) Aerosol layer top derived from AEROS5P as the aerosol highest layer with aerosol extinction larger than 0.1 km^{-1} . (c) Maximum fire radiative power [megawatt] derived from MODIS sensors onboard Aqua and Terra within each $0.2 \times 0.2^\circ$ grid pixel. (d) Static atmospheric stability depicted by the Brunt–Vaisala frequency [rad/s] derived from ERA5 reanalysis. Winds [$\text{m}\cdot\text{s}^{-1}$] from ERA5 reanalysis are overlaid on panels (a,b) at 850 hPa and (d) at 600 hPa. For clarity, wind speed scales are shown in the upper left corners and wind vectors are displayed for each $1.5 \times 1.5^\circ$ grid point. 116

Figure III-15 TROPOMI Carbon monoxide total column from 20 to 24 December 2019 117

Figure III-16 Same as Figure III-14 but for 21 December 2019. 117

Figure III-17 Same as Figure III-14 but for 22 December 2019. 118

Figure III-18 Same as Figure III-14 but for 23 December 2019. 119

Figure III-19 Land cover type map for Australia [206]. 120

Figure IV-1 Flowchart of the pixel-by-pixel bias correction approach of the CHIMERE AOD estimation. 131

Figure IV-2 Scatter plot of 966 collocated data points comparing the AOD at 550 nm measured by 8 AERONET sun photometers for the year 2021 with (a) MODIS satellite observations and (b) simulated AODs by the CHIMERE chemistry transport model. The black line is the $y = x$ line with the gray shaded areas representing the $\pm 20\%$ interval of its slope. 132

Figure IV-3 Annual median of the difference between CHIMERE and MODIS AOD at 550 nm for the year 2021 over the simulation domain of North Africa and the Arabian Peninsula. The stars indicate the 8 AERONET stations used for the comparisons in Figure IV-2. 132

Figure IV-4 (a) MODIS true color showing the mineral dust emitted from Bodélé depression region on 23 March 2021. (b) Average MODIS AOD₅₅₀ values for 2021. (c) Number of observations used in the computation of the average AOD₅₅₀ of panel (b)..... 135

Figure IV-5 Scatterplot showing the linear relationship between CHIMERE AOD₅₅₀ bias toward MODIS observations for different AOD₅₅₀ levels represented by the CHIMERE

AOD ₅₅₀ . The colorbar represents the occurrence in the year 2021. The green line represents the linear fit with the correlation coefficient a and the intercept b	136
Figure IV-6 Histogram showing the logarithmic distribution of the AODs for both CHIMERE and MODIS over North Africa for the year 2021.	138
Figure IV-7 A Neural network graph representaion.	141
Figure IV-8 Illustration of a decision tree (left) and the random forest regressor (right panel). Boxes represent split nodes, and circles represent the prediction, adapted from [254].	142
Figure IV-9 Horizontal distribution of AOD at 550 nm on 30 September 2021 (not used for training) from (a) MODIS satellite observations, (b) raw CHIMERE simulation, and bias-corrected AODs using the (c) MLR, (d) NN, (e) RF, and (f) XGB models. The AOD predictions of the remaining days are available at the link provided in the data availability clause.....	145
Figure IV-10 Mean absolute Shapley values for the input variables of the ML models for the 30 September 2021.....	147
Figure IV-11 Shapley values for (a) raw CHIMERE AOD, (b) specific humidity, (c) PM ₁₀ , and (d) nitrate particles for the RF model for the 30 September 2021.	148
Figure IV-12 Shapley values for (a) hydroperoxide (ROOH) and (b) non-methane hydrocarbons (NMHC) for MLR model for the 30 September 2021.	149
Figure IV-13 Histogram showing the frequency per 0.02 bin size bias, which is the difference between modeled (raw CHIMERE) or corrected (ML models) AOD ₅₅₀ and satellite observed AOD ₅₅₀ from MODIS for D_{Test}	150
Figure IV-14 Scatterplots showing the AOD biases amplitudes corrected by the following ML models: (a) MLR, (b) NN, (c) RF, and (d) MLR, with respect to the a priori AODs (raw CHIMERE AOD ₅₅₀ estimates) for D_{Test} . The green line represents the linear fit with a slope a and the intercept b . Colormap represents the occurrence.	151
Figure IV-15 Scatterplots of corrected AOD ₅₅₀ inferred by the (a) MLR, (b) NN, (c) RF, and (d) MLR models, with respect to AOD measurements from MODIS for D_{Test} . The green line represents the linear fit with a slope a and the intercept b . The gray lines represent the lines $y=x$. Colormap represents the occurrence. The apparent limit of the MODIS AOD at the value 3.5 is an arbitrary upper limit fixed by the original Deep Blue retrieval algorithm.	152
Figure IV-16 Median values of the difference between CHIMERE bias corrected AOD ₅₅₀ and MODIS AOD ₅₅₀ for the dates of D_{Test} for the (a) MLR, (b) NN, (c) RF, and (d) XGB ML models.	153

Figure IV-17 Time series of AOD₅₅₀ in Cairo during the year 2021 from AERONET ground-based sun photometer measurements, raw CHIMERE simulations, ML-corrected AODs, and MODIS. The gray areas represent dates used for training the models. 156

Liste des tableaux

Tableau I-1 Bilan global moyen des années 1990, 1996 et 1997 de différents types d'aérosols [45].	43
Tableau II-1 Les 8 bandes spectrale de TROPOMI	70
Table III-1. AEROS5P inputs and their corresponding sources.	102
Table III-2 Comparison of AOD from AEROS5P and VIIRS for each aerosol type detected by VIIRS. r: Pearson correlation coefficient; RMSE: Root Mean Squared Error; MAE: Mean Absolute Error.	113
Table IV-1 Average performance of each AOD bias corrector and raw CHIMERE on D_{Test} with respect to MODIS AOD. t: daily mean inference time cost; r: Pearson correlation coefficient; RMSE: root mean squared error; MAE: mean absolute error; Skp: Pearson's coefficient of skewness; and μ : mean; percentages show the percentiles of the error. The metrics are calculated for the 737129 pixels making D_{Test} .	139
Table IV-2 The network architecture search space.	142
Table IV-3 The best architecture for the NN regressor.	142
Table IV-4 Correlation and residuals of AOD estimations for the year 2021 against MODIS Terra observation. The raw AODs are CHIMERE outputs at 10:00 LT and for the corrected AOD with the RF model (trained with MODIS AQUA observations at 13:30 LT). Number of data points is 3.7 million.	154
Table IV-5 Correlation and residuals of 330 AOD estimations on testing dates that are collocated with the 8 AERONET stations.	155

Acronymes

AERONET AErosol RObotic NETwork
AOD Aerosol optical depth
BB Biomass Burning
BC Black Carbon
CALIOP Cloud-Aerosol Lidar with Orthogonal Polarization
CALIPSO Cloud Aerosol Lidar and Infrared Pathfinder Satellite Observations
CNES Centre National des Etudes Spatiale
CNRS Centre National de la Recherche Scientifique
COV Composés Organiques Volatils
CTM Chemistry-Transport Model
DIM-Qi2 Domaine d'Intérêt Majeur Réseau de recherche Qualité de l'air en Ile-de-France
DISORT Discrete Ordinates Radiative Transfer
DLER Directionally dependent Lambertian-equivalent reflectivity
DLR German Aerospace Center
ECMWF European Centre for Medium-Range Weather Forecasts
EOF Empirical Orthogonal Functions
EOSDIS Earth Observing System Data and Information System
ESA European Space Agency
ETR Équation de transfert radiative
FASCODE Fast Atmospheric Signature Code
FIRMS Fire Information for Resource Management System
GEMS Geostationary Environment Monitoring Spectrometer
GOME Global Ozone Monitoring Experiment
IASI Infrared Atmospheric Sounding Interferometer
INSU Institute National de Sciences de l'Univers
KNMI Royal Netherlands Meteorological Institute
LBLRTM Line-By-Line Radiative Transfer Model
LIDAR Light Detection And Ranging
LISA Laboratoire Interuniversitaire des Systèmes Atmosphériques
LMD Laboratoire de Météorologie Dynamique
INCA Interaction with Chemistry and Aerosols
MAE Mean Absolute Error
MAIAC Multi-Angle Implementation of Atmospheric Correction
MAPGA Make Our Planet Great Again

MDPI Multidisciplinary Digital Publishing Institute
MELCHIOR Modèle Lagrangien de Chimie de l'Ozone à l'échelle Régionale
MISR Multi Angle Imaging Spectroradiometer
ML Machine Learning
MLR Multiple Linear Regression
MODIS Moderate resolution Imaging Spectrometer
MODTRAN MODerate resolution atmosphere TRANsmission
MTG Meteosat Troisième Génération
NASA National Aeronautics and Space Administration
NCEP National Centers for Environmental Prediction
NIR Near Infra Red
NMHC Non Methane Hydrocarbons
NN Neural Network
NPP Suomi National Polar-orbiting Partnership
OC Organic Carbon
OMI Ozone Monitoring Instrument
PHOTON PHOtométrie pour le Traitement Opérationnel de Normalisation Satellitaire
PM Particulate Matter
PNTS Programme National de Télédétection Spatiale
PRPHD Programme Paris Region PhD
RF Random Forest
RMS Root Mean Squares
RMSE Root Mean Squared error
RT Radiative Transfer
RTTOV Radiative Transfer for TIROS Operational Vertical Sounder
S5P Sentinel-5 precursor
SCIAMACHY SCanning Imaging Absorption spectroMeter for Atmospheric CHartographY
SMART-G Speed-Up Monte Carlo Advanced Radiative Transfer using GPU
SOA Secondary Organic Aerosols
SWIR Short Wave InfraRed
TEMPO Tropospheric Emissions: Monitoring of Pollution
TOSCA Terre Ocean Surface Continental et Atmosphère
TIROS Television infrared observation satellites
TROPOMI TROPospheric Ozone Monitoring Instrument
UPEC Université Paris Est Créteil
USGS United States Geological Survey
UV Ultra Violet

UVIS Ultraviolet et Visible

VIIRS Visible Infrared Imaging Radiometer Suite

VLIDORT Vector Discrete Ordinate Radiative Transfer

WRF Weather Research and Forecasting Model

XGB Extreme Gradient Boosting

Introduction de la thèse

Contexte scientifique :

Les aérosols atmosphériques sont des particules de matière (PM) minérale ou organique en suspension dans l'air. Les PM se forment soit par réaction chimiques de gaz précurseurs ou par émission directe depuis la source. Elles ont des temps de résidence variables dans l'atmosphère, pouvant rester en suspension de quelques heures jusqu'à quelques jours [1], et elles peuvent être transportées à des milliers de kilomètres de leur lieu d'origine [2]. Les impacts négatifs de la présence d'aérosols dans l'atmosphère varient selon leur altitude. En effet, les PM en basse altitude dégradent la qualité de vie de la population locale. D'abord, ils diminuent la visibilité et détériorent esthétiquement la ville en se déposant sur les surfaces. En plus de ces effets indésirables immédiates sur la vie quotidienne, l'inhalation des aérosols provoque sur le long terme différentes maladies, des allergies, jusqu'aux cancers de poumons [3–6]. Les aérosols atteignant des hautes altitudes dans l'atmosphère comme les émissions de certains volcans, le carbone de suie des feux de forêts, et les poussières désertiques ont une durée de vie plus longue. Ils modifient le bilan radiatif terrestre [7,8] de façon difficile à quantifier, ce qui constitue une source d'incertitude dans l'évaluation et le suivi du réchauffement climatique de la planète.

Certains panaches de particules comme ceux des émissions de feux se développent rapidement. L'altitude à laquelle la matière est injectée est souvent mal connue. Cette altitude d'injection des fumées est une information clé, car elle est utilisée dans les modèles de simulation de panaches d'aérosols. Une estimation précise de cette altitude constitue un grand pas vers une connaissance meilleure des cycles de vie des aérosols. La connaissance de la distribution spatiotemporelle des aérosols nous permet d'étudier le devenir des aérosols et de concevoir des stratégies pour limiter l'exposition à la pollution. Il existe quelques méthodes pour estimer l'abondance des PM dans l'atmosphère. Certaines méthodes reposent sur l'échantillonnage in situ, d'autres estiment l'abondance des aérosols à distance via la télédétection. Enfin, la modélisation des processus physiques et chimiques de l'atmosphère et du transport des constituants atmosphériques fournit également des estimations utiles des concentrations d'aérosols dans l'air. Chacune de ces trois méthodes est dotée de points forts mais elles ont aussi des limitations.

La mesure *in situ* est très localisée dans l'espace, il faudrait soit échantillonner beaucoup pour couvrir une large superficie, voir équiper des avions pour mesurer la variabilité de la concentration en altitude. Bien que cette technique soit précise, elle est parfois inadaptée pour étudier la pollution à grande échelle, surtout à cause de son important coût financier de déploiement.

La télédétection satellitaire des aérosols échappe à cet inconvénient. Le principe est de déployer un équipement doté de capteurs au bord d'un satellite orbitant la terre, et d'observer constamment la surface de la terre. L'instrument enregistre des spectres de radiance qui seront utilisés pour déduire une information sur l'abondance d'aérosols dans l'atmosphère en utilisant des méthodes de modélisation inverse. Même si l'observation satellitaire est occultée par les nuages, elle reste pratique et plus rentable au vue de la couverture spatiale. Néanmoins, à l'heure actuelle, l'abondance des PM fines est estimée par télédétection uniquement en 2D, soit en termes de l'épaisseur optique (AOD pour *Aerosol Optical Depth*) intégrale de la colonne atmosphérique, avec une mesure par jour, ce qui donne un champ d'abondance horizontale des aérosols, soit en transects latitudinaux de bonne résolution verticale en utilisant la technologie lidar (*Light Detection and Ranging*), ces transects ne sont pas suffisants pour avoir une estimation spatialement continue puisqu'ils sont espacés de plusieurs milliers de kilomètres à cause du faible angle de vue.

Afin de garantir la couverture des régions polaires, les études d'aérosols exploitent les mesures satellitaires en orbites polaires. Néanmoins, la fréquence de ces observations est journalière, ce qui est insuffisant pour une bonne connaissance de l'évolution parfois rapide des panaches de PM. Ainsi, pour avoir une meilleure résolution temporelle, on se sert des observations issues de plateformes géostationnaires, mais ces derniers n'observent pas les régions polaires. Afin d'avoir une bonne couverture à la fois spatiale et temporelle, on utilise la modélisation de chimie-transport. Un modèle de ce genre dit Eulérien implémente des équations basées sur la physique et la chimie, et parfois empiriques. Ces derniers simulent les phénomènes influençant la concentration des aérosols : l'émission, la chimie, le transport, etc. Les particules sont modélisées dans une grille tridimensionnelle qui constitue le domaine de simulation. L'estimation de ces modèles est fournie à des échelles de temps courtes généralement de l'ordre de l'heure. Par contre, une plus grande incertitude est liée à la simulation comparée à l'observation satellitaire. En générale, l'incertitude de la modélisation de chimie-transport vient de la paramétrisation physique, les données d'entrées, et

l'approximation numérique. Ces facteurs contribuent jusqu'à 17% dans le cas de la simulation de l'ozone [9]. Quant à la simulation des aérosols, différents facteurs engendrent des erreurs pouvant atteindre 50% [10]. Ces derniers dépendent de l'altitude et elles sont probablement liées aux incertitudes des taux d'émissions des polluants, et des caractéristiques des aérosols [10]. Par exemple, les flux de soulèvement de sable du désert nord-africain sont mal estimés. En effet, ce sont des phénomènes assez complexes qui dépendent non seulement de la vitesse du vent à la surface, mais aussi de la vitesse de friction et des caractéristiques de la surface comme l'humidité, la granularité du sable, etc. En outre, l'incertitude est d'autant plus grande dans les régions qui ne possèdent pas des stations de mesure à la surface pour contraindre la déviation du modèle de la réalité.

Objectifs et Méthodologie :

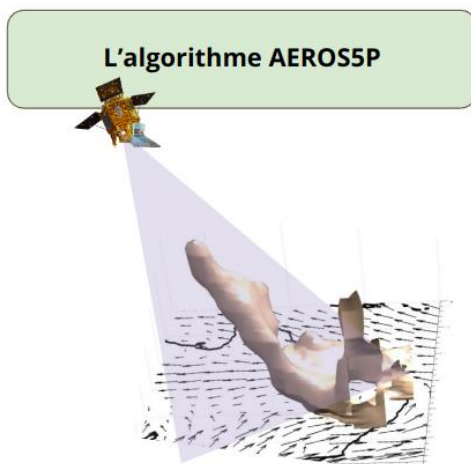
L'objectif de ma thèse est d'améliorer les estimations existantes de la distribution spatiale des aérosols en exploitant les données satellitaires et l'intelligence artificielle. Ces améliorations ont ciblé les deux méthodes d'estimations classiques : la télédétection depuis l'espace et la simulation de chimie-transport. La première partie de mon travail va combler le manque d'information tridimensionnelle (3D) sur les panaches d'aérosols fins des méthodes d'observation depuis l'espace. La deuxième partie présente une méthode novatrice pour améliorer la précision de la simulation régionale de chimie-transport de l'AOD. Cette dernière méthode sera extrapolée dans le futur pour améliorer le profil d'extinction vertical également en utilisant le produit développé dans la première partie. Pour réaliser ce projet, j'ai procédé aux étapes suivantes :

Premièrement, j'ai adapté un algorithme d'observation des aérosols grossiers pour observer la distribution 3D des particules fines. Ensuite, j'ai utilisé l'algorithme développé dans un cas d'étude d'un grand épisode de feux en Australie, puis j'ai analysé la distribution tridimensionnelle des aérosols issue de ces feux. Les observations 3D développées ici constituent une source d'information complémentaire aux autres méthodes existantes d'estimation de l'altitude des aérosols. La deuxième étape de mes travaux a consisté à utiliser les observations satellitaires pour améliorer la précision des simulations de la distribution spatiale des aérosols par un modèle de chimie-transport. Pour cela, j'ai développé à l'aide d'outils d'apprentissage automatique un correcteur des biais de la sortie de simulation d'un modèle de chimie-transport en se basant sur les observations satellitaires. Cette méthode de synergie satellite-modèle a permis d'améliorer la précision de l'estimation de la distribution

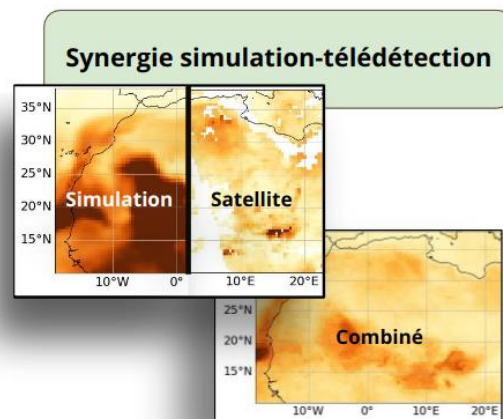
spatiale des aérosols en termes d'AOD. Elle fournit une distribution spatiale qui couvre entièrement le domaine géographique de la simulation, plusieurs fois par jour, même dans les situations où des mesures satellitaires ne sont pas disponibles (à cause des nuages par exemple).

Cette dernière approche est appliquée aux champs horizontaux de la distribution 2D des aérosols. L'entraînement du modèle d'apprentissage automatique est fait à l'aide des mesures satellitaires de l'AOD du capteur MODIS (largement validées sur un grand nombre de conditions atmosphériques). Son application avec des observations de la distribution 3D des aérosols (comme celles développées dans la première partie de la thèse) sera effectuée en perspectives de ce travail. Cela nécessitera la consolidation préalable de la nouvelle méthode de télédétection satellitaire pour être applicable à un grand nombre de situations (différents types d'aérosols et de régions).

1 Observation 3D des aérosols fins par télédétection depuis l'espace



2 Amélioration de la simulation de l'AOD par apprentissage supervisé



I. Chapitre 1 : Aérosols dans l'atmosphère

Nous allons décrire dans ce chapitre les caractéristiques des aérosols, et leurs différents types, comment ils se forment, et le rôle qu'ils jouent dans l'environnement. Ce chapitre constitue une base scientifique pour comprendre l'activité optique de ces composés et leurs enjeux climatiques et de santé.

1.	Généralité sur l'atmosphère terrestre	31
1.1.	La troposphère	32
1.2.	La stratosphère	33
1.3.	La mésosphère	33
1.4.	Dynamique de l'atmosphère	33
2.	Aérosols atmosphériques	36
2.1.	Classification des aérosols	38
2.2.	Mécanismes de formation des aérosols atmosphériques	39
2.2.1.	Aérosols issus de combustion	39
2.2.2.	Poussière minérale	40
2.2.3.	Aérosols de sels marins.....	40
2.2.4.	Aérosols secondaires.....	40
2.3.	Durée de vie des principaux aérosols atmosphériques	42
2.4.	Impacts des aérosols	43
2.4.1.	Effets sur la santé	43
2.4.2.	Effets sur l'environnement.....	44
3.	Interaction des particules avec le rayonnement	46
3.1.	Le rayonnement électromagnétique	47
3.2.	Absorption.....	49
3.3.	Diffusion simple.....	50
3.3.1.	Diffusion Rayleigh.....	51
3.3.2.	Diffusion de Mie	52
3.3.3.	Albédo de diffusion simple	54
4.	Atténuation du rayonnement dans l'atmosphère.....	55
4.1.	Épaisseur optique	57
5.	Conclusions.....	57

1. Généralité sur l’atmosphère terrestre

L’atmosphère terrestre est la couche de gaz qui enveloppe la terre. C’est un milieu semi-transparent qui s’étend depuis la surface jusqu’à l’altitude où la pression tend vers zéro, elle peut aller jusqu’à plusieurs milliers de kilomètres. On trouve différentes espèces gazeuses et particulaires dans l’atmosphère ; ces espèces se distribuent verticalement de manière différente suivant leurs densités.

L’atmosphère est une enveloppe gazeuse subdivisée en deux grandes couches : la première couche est l’homosphère qui s’étend depuis la surface jusqu’à environ 90 km. L’homosphère est constituée de la troposphère, la stratosphère, et la mésosphère (Figure I-1). Sous l’effet de la turbulence, sa composition en principaux gaz (azote et oxygène) est relativement constante. Cependant, les gaz minoritaires (vapeur d’eau, ozone, dioxyde de carbone, méthane, etc.) et les particules se distribuent de manière très hétérogène. La deuxième couche concentrique à la première est l’hétérosphère celle-ci complète la première jusqu’à l’espace vide, elle est constituée de la thermosphère (ionosphère) et l’exosphère. A l’opposition de l’homosphère, la composition de l’hétérosphère en principaux gaz est très variable selon l’altitude à cause de la stratification des composés sous l’effet du champ gravitationnel et du gradient de température (Figure I-1).

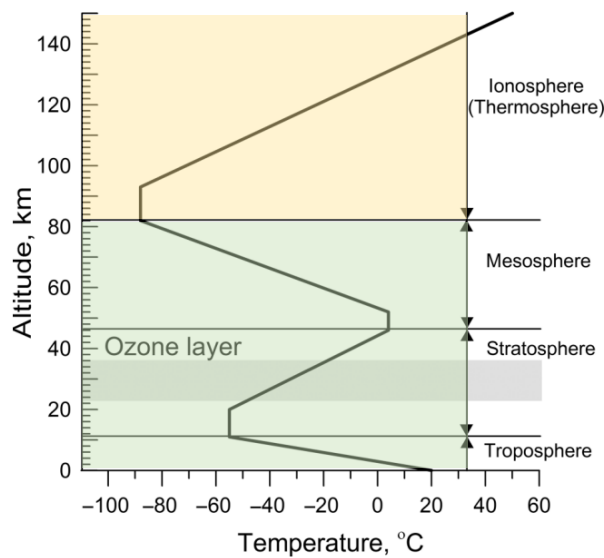


Figure I-1 Profil vertical de température de l’atmosphère terrestre montrant la stratification des couches atmosphériques [11].

Les différentes couches atmosphériques sont distinguées grâce à la variation de la température avec l'altitude. Si la température augmente avec l'altitude ($\frac{\partial T}{\partial z} > 0$) alors l'atmosphère est stable. À l'inverse, si la température diminue avec l'altitude ($\frac{\partial T}{\partial z} < 0$) alors l'atmosphère est instable. Ce gradient de température délimite les différentes couches atmosphériques. À la troposphère sa valeur est négative ce qui explique l'instabilité de la couche. Par contre, à la tropopause, en raison de la présence de l'Ozone (O_3) l'absorption du rayonnement ultraviolet réchauffe l'air ce qui se manifeste par une inversion du signe du gradient de température. La stabilité de la stratosphère crée une barrière dynamique qui réduit le transport de la matière depuis la troposphère vers la stratosphère [12].

1.1. La troposphère

La troposphère a une épaisseur qui varie selon la latitude, la saison, et les conditions météorologiques, elle peut s'étaler de 8 à 18 km. La température dans la troposphère diminue avec l'altitude avec un gradient de 5 à 6° C par kilomètre. La troposphère contient environ 80% de la masse totale de l'atmosphère [12]. Elle est constituée principalement d'azote (78,08%), 20,95% d'oxygène, 0,93% d'argon, de vapeur d'eau en proportions variables et 0,04% de dioxyde de carbone, en plus des traces d'autres gaz. Elle contient également des composés des particules de matière solide ou liquide comme la poussière, et les nuages.

Les phénomènes d'échange de chaleur, de matière et de quantité de mouvement entre la surface de la terre et l'atmosphère se produisent dans la partie inférieure de la troposphère qu'on appelle la couche limite atmosphérique. Celle-ci s'étend de quelques mètres jusqu'à 2 km aux latitudes moyennes [13] et jusqu'à 6 km dans les régions désertiques dans les tropiques [14]. L'épaisseur de cette couche varie selon l'intensité du rayonnement solaire reçu à la surface et la stratification de l'air dans la troposphère libre qui se trouve au-dessus de celle-ci. Grâce à l'échange de chaleur et de la friction issue du frottement de l'air avec les reliefs à la surface, la couche limite est brassée. Durant la journée, elle est bien mélangée, elle s'étend depuis la surface jusqu'au niveau de l'inversion du profil de température [15]. La nuit, le mélange turbulent induit par la flottabilité diminue rapidement et on se retrouve avec une couche limite plus fine (<30 m), laissant une couche résiduelle (Figure I-2), qui s'étend jusqu'à la troposphère libre [13,16].

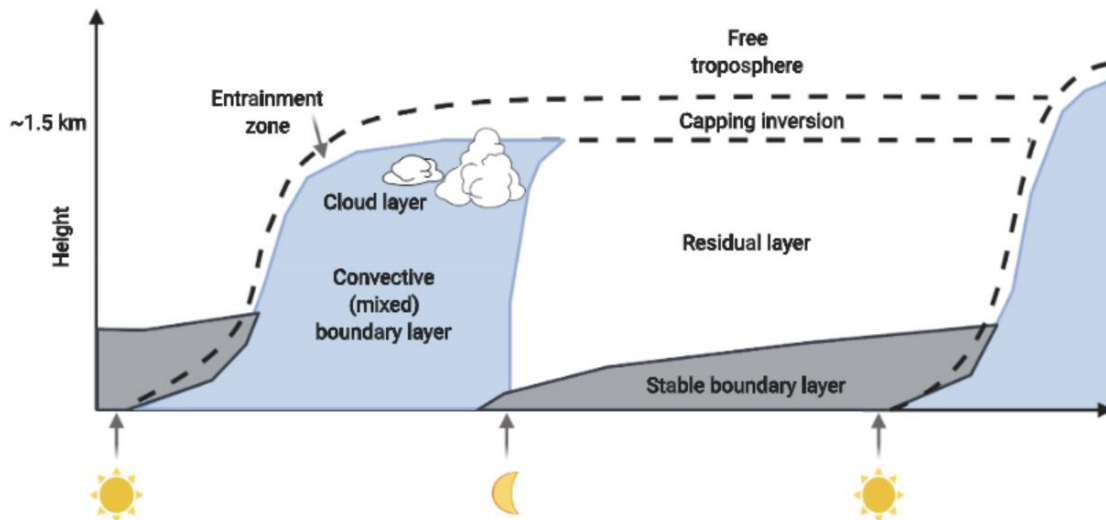


Figure I-2 Développement de la couche limite atmosphérique pendant une journée [13,16].

1.2. La stratosphère

La stratosphère commence de la tropopause qui est la couche d'interface troposphère-stratosphère de 8 à 18 km d'épaisseur et s'étend jusqu'à 50 km d'altitude (Figure I-1). L'air est stratifié dans la stratosphère à cause du fort gradient de température. Il est composé de très peu de vapeur d'eau, mais majoritairement d'azote et d'oxygène. On estime que 80 à 90% de l'ozone (O_3) atmosphérique se trouve dans la stratosphère, formant la couche d'ozone protectrice des rayonnements UV. Les particules qu'on trouve dans la stratosphère sont principalement des gouttelettes d'acide sulfurique en phase aqueuse, mais dans certaines situations, les composés soufrés atteignent la stratosphère par injection depuis les volcans. D'autres aérosols issus des feux de forêts intenses peuvent également atteindre la stratosphère à l'aide de l'effet *self-lofting* (voir page 39).

1.3. La mésosphère

La température décroît à la mésosphère atteignant le point le plus froid, c'est la mésopause, qui se situe vers 80 km d'altitude (Figure I-1). La mésosphère s'étend approximativement de la stratopause (vers 48 km) jusqu'à la mésopause.

1.4. Dynamique de l'atmosphère

L'étude du mouvement horizontal et vertical des masses d'air fait partie de la météorologie qui est l'étude de la troposphère et la stratosphère [17]. La circulation générale de l'atmosphère a été étudiée depuis des décennies, au point qu'on arrive à identifier l'origine

et à prévoir avec une certaine précision les phénomènes météorologiques. L'origine des mouvements de la matière dans l'atmosphère est en effet l'instabilité du système atmosphérique. Cette instabilité vient de la variation du profil vertical de température, du mouvement de rotation de la terre sur son axe, ainsi que de la distribution régionale inégale de l'énergie externe (solaire) [12].

La terre ne reçoit pas la même quantité d'énergie aux pôles qu'aux zones tropicales. Bien que le flux radiatif soit approximativement symétrique entre l'hémisphère nord et sud, la distribution de l'énergie reçue n'est pas pareille entre l'équateur et les pôles. D'après la Figure I-3, l'équateur reçoit plus de flux de courtes longueurs d'ondes, ceci est associé à la différence de l'angle zénithal et de l'albédo du sol de la région comparé aux pôles. L'émission infrarouge en revanche est quasi similaire entre les tropiques et les pôles. Donc le flux radiatif net reçu par les régions tropicales est positif alors qu'il est négatif pour les régions polaires.

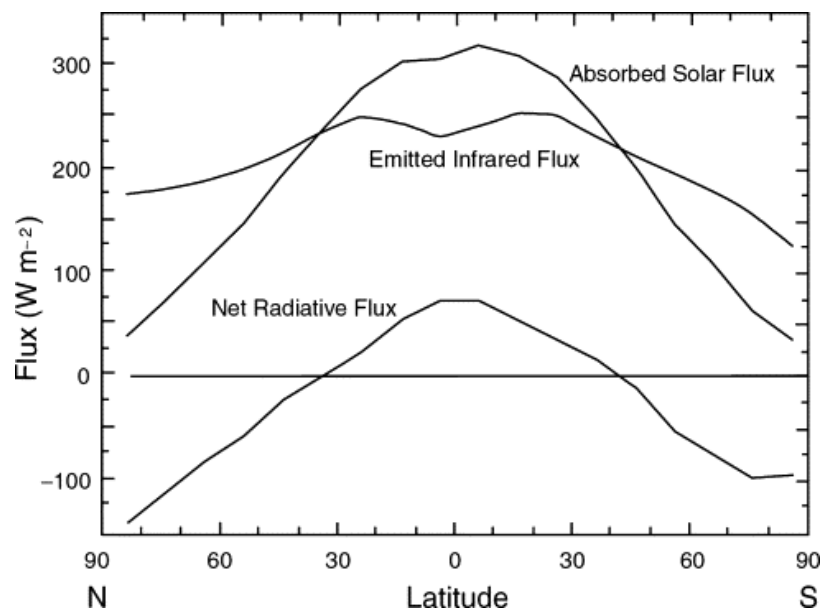


Figure I-3 Moyenne zonale annuelle du flux radiatif solaire incident et émis par la terre au sommet de l'atmosphère tiré des mesures satellitaires [18].

Le bilan énergétique régional (Figure I-4) montre que les déserts tropicaux et les régions polaires ont un déficit dans la puissance radiative reçue. Alors que les zones de moussons ont un excès d'énergie. La circulation zonale et verticale de Hadley-Walker assure la redistribution de l'énergie jusqu'à la zone $\pm 30^\circ$ latitude. Au-delà de ces régions d'autres processus assure le transport méridien de l'énergie vers de plus haute latitude [12] tel que les cellules dites de Ferrel et les cellules polaire Figure I-5. Des circulations horizontales en conséquence de ces cellules convectives sont régies par le mouvement de l'air des hautes pressions (anticyclones)

vers les basses pressions (cyclones). Ce déplacement se fait détourné vers la droite dans l'hémisphère nord, et vers la gauche à l'hémisphère sud sous l'effet de la force de Coriolis [19], cette dernière est une force apparente due à la rotation de la terre autour d'elle même. On constate les Alizés près de la surface en zone de subsidence des cellules d'Hadley, et le courant Jet ou le vent polaire à la tropopause. Ces vents contribuent au transport des aérosols également [20,21].

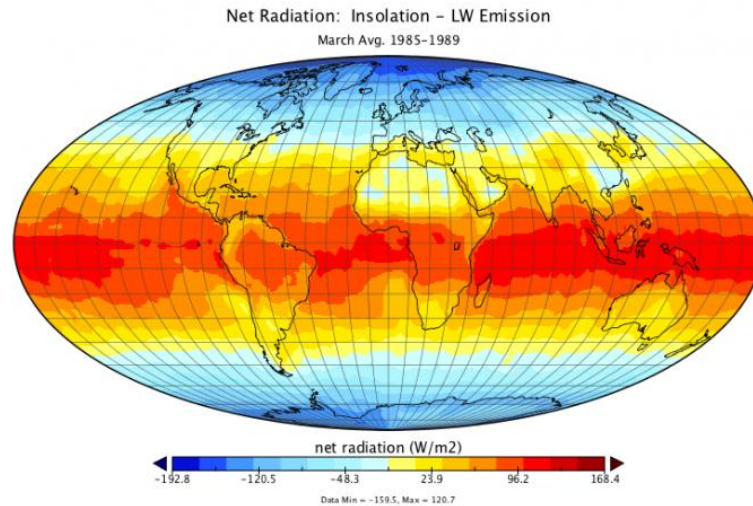


Figure I-4 Estimation de la différence entre l'insolation et l'émission radiatif terrestre entre 1985 et 1989 par imagerie satellitaire [22].

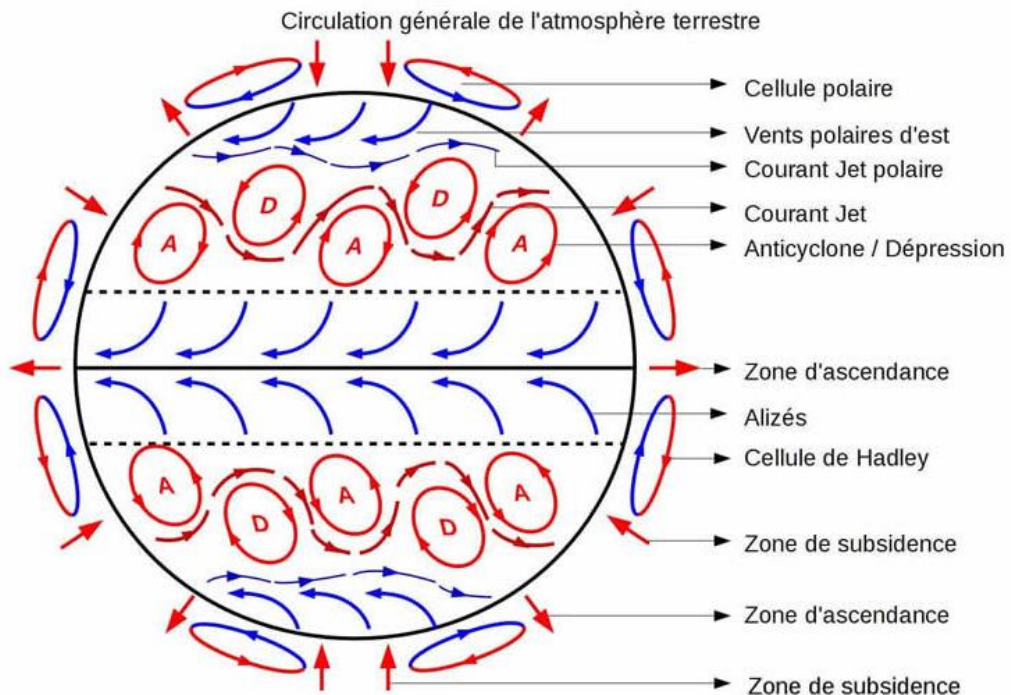


Figure I-5 Image montrant les différents phénomènes mise en jeu dans la circulation atmosphériques [19].

Le mélange d'air dans le sens zonal prend un à plusieurs mois selon la région, au sens méridien il dure environ une année [12]. La durée d'échange caractéristique entre la troposphère et la stratosphère est de plusieurs années typiquement entre 2 et 10 ans [12]. La circulation dans la stratosphère est dominée par des mouvement d'échelle planétaire (mouvement horizontaux de 10 000 km comme les anticyclones subtropicaux...) alors que dans la troposphère on trouve toutes les gammes d'échelles de circulation : planétaire, synoptique (comme les mouvements zonaux de ~1000 km), méso-échelle (mouvement horizontale de 10 à 100 km), échelle convective (de 1 à 10 km comme les orages de moyennes altitudes, cumulonimbus...) et à la micro-échelle (< 1 km comme les tourbillons de la couche limite) [12].

2. Aérosols atmosphériques

Les aérosols sont des particules de matière (PM) en suspension dans un gaz. Ces dernières sont de différentes tailles. Dans la nature, on trouve des particules fines avec un diamètre de quelques nanomètres jusqu'aux particules grossières avec une taille supérieur à 10 μm (Figure I-6). On décrit la taille d'un mélange d'aérosols à l'aide d'une fonction de distribution de taille en nombre, voir équations (I.1)(I.2). Couramment, cette fonction est une distribution log normale, avec un rayon modal r_m et un écart type de σ_r , où n le nombre de particules du rayon r , et N_0 le nombre total de particules par unité de volume. On peut exprimer la distribution de taille des aérosols en termes de surface $a(r)$ ou en termes de volume $v(r)$ des particules. L'équation (I.4) montre celle exprimée en surface où A_0 est la surface totale des aérosols par unité de volume, μ_a le rayon de la surface médiane, et σ_a la déviation standard géométrique. La distribution en volume exprimée dans l'équation (I.6) montre une analogie avec celle exprimée en termes de surface, où V_0 est le volume total d'aérosols par unité de volume, μ_v est le rayon du volume médian, et σ_v la déviation standard géométrique [23].

$$n(r) = \frac{N_0}{r \ln(\sigma_r) \sqrt{2\pi}} \exp \left[-\frac{1}{2} \left(\frac{\ln\left(\frac{r}{r_m}\right)}{\ln(\sigma_r)} \right)^2 \right] \quad (\text{I.1})$$

$$n(r) = \frac{dN(r)}{dr} \quad (\text{I.2})$$

$$N_0 = \int_0^{\infty} n(r) dr \quad (\text{I.3})$$

$$a(r) = \frac{A_0}{\sigma_a r \sqrt{2\pi}} \exp \left[-\frac{1}{2} \left(\frac{\ln(r) - \mu_a}{\sigma_a} \right)^2 \right] \quad (\text{I.4})$$

$$A_0 = 4\pi \int_0^{\infty} r^2 n(r) dr \quad (\text{I.5})$$

$$v(r) = \frac{V_0}{\sigma_v r \sqrt{2\pi}} \exp \left[-\frac{1}{2} \left(\frac{\ln(r) - \mu_v}{\sigma_v} \right)^2 \right] \quad (\text{I.6})$$

$$V_0 = \frac{4}{3} \pi \int_0^{\infty} r^3 n(r) dr \quad (\text{I.7})$$

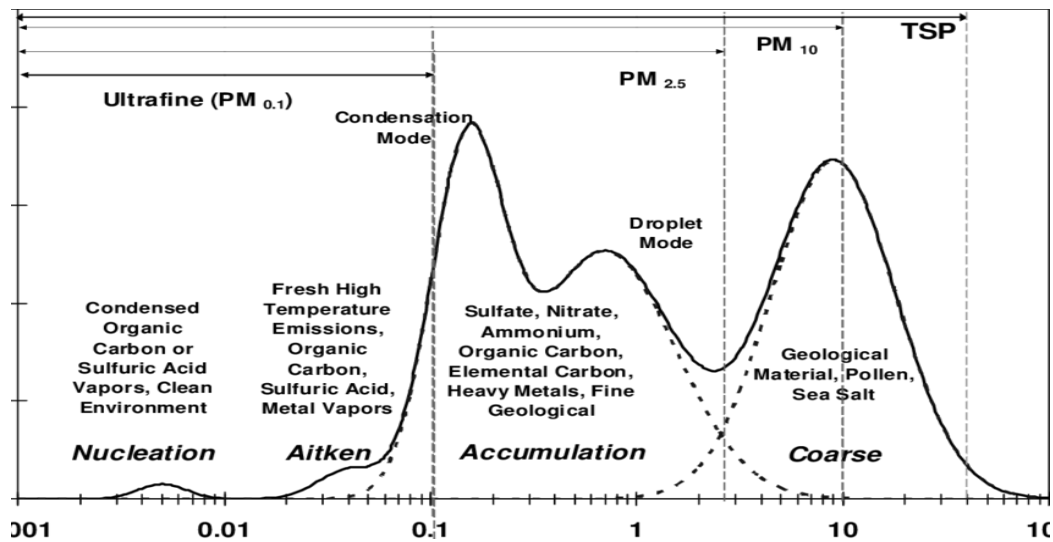


Figure I-6 Distribution de taille des différents aérosols atmosphériques [24].

Les concentrations des aérosols sont souvent exprimées en termes de concentrations massiques groupées par tailles à des intervalles de diamètres. On désigne par $PM_{2.5}$ la somme des concentrations des particules dont le diamètre est inférieur à $2.5 \mu\text{m}$, alors que les PM_{10} sont les particules dont le diamètre est supérieur à $10 \mu\text{m}$.

2.1. Classification des aérosols

On trouve plusieurs types d'aérosols dans l'atmosphère. Les aérosols les plus abondants en ordre décroissant sont les sels marins, la poussière minérale, le carbone organique, les aérosols sulfatés, puis le carbone suie. Ces particules en suspension peuvent être classifiées selon la taille (fine/grossière), l'origine (biogénique, anthropique), la composition (minérale, organique) ou la forme (sphéricité) des grains. Leurs impacts sur l'environnement et sur les êtres vivants varient selon ses caractéristiques.

Certains aérosols appelés aérosols primaires sont émis directement depuis la source, comme le sel marin, la cendre volcanique et la poussière désertique. Les particules de matière à base d'eau comme les gouttelettes de nuages et les cristaux de glace ne sont pas considérées comme des aérosols mais des hydrométéores [25]. À l'opposé, les aérosols secondaires se forment dans l'air par des interactions intermoléculaires des espèces gazeuses atmosphériques précurseurs, c'est le cas des aérosols de sulfates et de nitrates d'ammonium et quelques aérosols organique. Selon la durée de vie des particules et leurs réactivités, on peut en effet trouver des PM combinées, primaires et secondaires, comme illustré dans la Figure I-7.

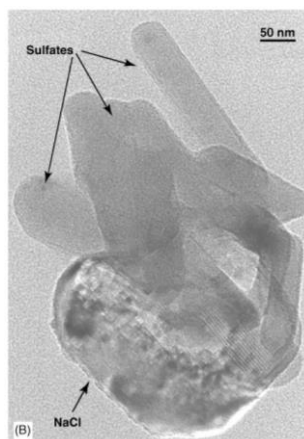
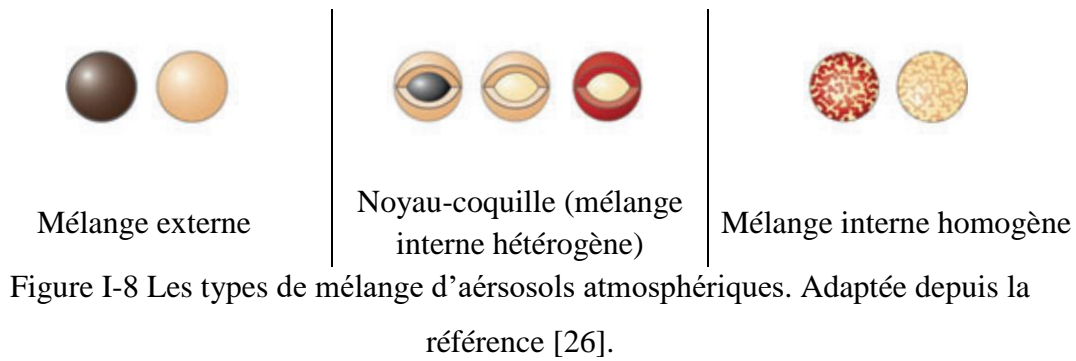


Figure I-7 Particule composée de sel marin et de sulfates [8].

2. Etat de mélange des aérosols

Dans l'atmosphère, on trouve souvent les PM constituées de mélange de plusieurs composés, on le qualifie selon sa constitution en mélange interne (homogène ou hétérogène), externe, ou une combinaison des deux (Figure I-8). La nature du mélange influence les caractéristiques optiques de l'aérosol.



2.2. Mécanismes de formation des aérosols atmosphériques

2.2.1. Aérosols issus de combustion

Les émissions produites par la combustion de carburants ont une grande variabilité, elles dépendent de la composition du combustible et des conditions de combustion. Si la combustion est incomplète la fumée contient des composés organiques volatils (COV) qui eux même forment des aérosols secondaires (SOA pour *Secondary Organic Aerosols*). Bien que l'activité humaine (comme le transport, chauffage, cuisson, industrie de pétrole, etc.) génère des aérosols issus de la combustion, les émissions de feux de la biomasse principalement végétale restent majoritaires. Ces derniers sont principalement des dérivés issus de la décomposition de la cellulose, des COV, de la résine, du carbone suie, et d'autre bio polymères [27]. La taille de ces particules est principalement dans le mode d'accumulation entre 0.1 et 1.3 μm (Figure I-6), environ 10% (principalement les cendres) sont des particules grossières de diamètre autour de 10 μm [28].

Une partie des émissions des feux de biomasse est due au brulage des résidus de l'agriculture, mais la majorité se produit spontanément en période de sécheresse, les feux peuvent être initiés aussi par les éclairs. On compte globalement environ 70 000 feux de forêt en moyenne chaque année [29]. Les régions tropicales sont les principales zones d'émission en termes de carbone. Dans certaines régions comme l'Australie, les savanes sont caractérisées par des feux de faible intensité mais de haute fréquence, à l'opposé des forêts permanentes où les feux sont plus intenses et moins fréquents, c'est le cas également des forêts boréales de Canada, Alaska et la Sibérie.

Des études ont montré que le changement climatique a engendré une augmentation de la longueur de la saison de feux, de la superficie brulée et de la fréquence des feux [30]. Parfois les feux sont si intenses, que les produits de la combustion atteignent la basse stratosphère [31].

Les émissions de combustion de la biomasse forment des SOA, ces derniers diffusent le rayonnement en contribuant directement et indirectement refroidissement de la planète, mais le carbone organique brun et le carbone suie contrarient cet effet et réchauffent la planète en absorbant le rayonnement [7]. Le réchauffement de l'air environnant se traduit par une ascension (*self-lofting*) des panaches d'environ 1 km par jour [32]. C'était le cas du *Black Summer* en Australie en 2019/2020 où les émissions ont atteint 35 km d'altitude.

2.2.2. Poussière minérale

Les aérosols de poussière minérale constituent le deuxième plus abondant type d'aérosol dans la planète. Ce sont des aérosols grossiers de taille typiquement supérieur à 1 μm , composés de mélanges complexes de différents minéraux principalement de l'argile, calcite, quartz, feldspaths, et des oxydes de fer [33]. Les poussières se forment sous l'effet du soulèvement du sable désertique. Leurs émissions se produisent en deux étapes : le vent exerce une force sur les agrégats de particules de sol supérieure aux forces qui les maintiennent ; l'énergie cinétique du vent mobilise les agrégats au sol dans un processus appelé saltation. Des chocs entre les agrégats et le sol libèrent les matériaux fins du sol et des agrégats, c'est le processus de sablage (*sandblasting*) [34]. Ce sont ces poussières plus fines qui sont ensuite soulevées par les vents et transportées à des régions lointaines parfois à des milliers de kilomètres. L'estimation des émissions de poussière minérale désertique est à ce jour incertaine [35].

2.2.3. Aérosols de sels marins

Les sels marins sont les aérosols les plus abondants sur Terre. Ces sels généralement de taille grossière sont principalement composés des ions de chlorure de sodium (NaCl), mais près du continent, ils se mélangent avec des composés anthropiques tel que les ions des sulfates, nitrates, magnésium, calcium, potassium, et d'ammonium [36]. Les sels marins sont émis à la couche limite marine, et ont une courte durée de vie, mais aux régions côtières (pour l'Europe), les sels transportés au continent auront une durée de vie étendue [37].

2.2.4. Aérosols secondaires

Les particules secondaires se forment en suivant des étapes d'agrégation successive, à partir de composés précurseurs qui s'accumulent en formant des nucléus. La nucléation homogène est le processus de formation de particules liquides ou gazeuses à partir des

molécules gazeuses précurseurs, tel que le dioxyde de soufre (SO_2) et les oxydes d'azote (NO_x). Par ailleurs, les COV sont oxydés par l'ozone, les radicaux d'hydroxyle, de chlore ou les nitrates pour donner des produits de faible volatilité. Ces derniers se rapprochent en créant des clusters (Figure I-9). Les particules issues de cette étape ont une taille de l'ordre du nanomètre, elles peuvent être composées de molécules similaires appelés des nucléus homo-moléculaires, ou différentes qui résulte d'une nucléation hétéro-moléculaire. Par ailleurs, la nucléation hétérogène est le processus d'agrégation de molécules sur un corps ou une surface étrangère, comme un ion ou une particule. A ce stade, les particules ont une taille inférieure à $0,01 \mu\text{m}$ (Figure I-6). Une fois que les nucléus sont formés, ils prennent du volume dans la phase d'accumulation pour former des aérosols de taille moyenne. La température, l'humidité relative ainsi que la viscosité sont des facteurs influençant sur la composition des particules formées [38].

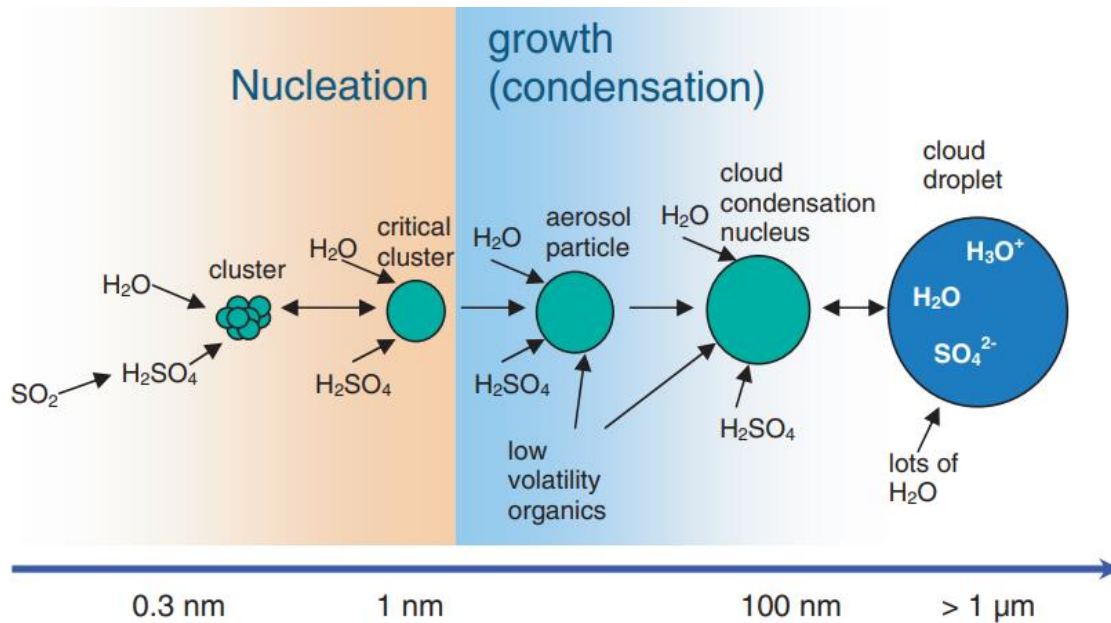


Figure I-9 Processus de formation des aérosols sulfatés.[39].

Le processus de nucléation est fréquemment observé dans les forêts, mais aussi en milieu urbain. On voit dans la Figure I-10 que la phase de nucléation des particules se déroule entre 9:00 et 12:00 du matin, puis elles grossissent en volume pendant le reste de la journée [40]. Dans les zones urbaines, les alcanes qui sont les espèces prédominantes des COV ont une durée de vie qui va d'environ 12 h à 34 h. Généralement, elles sont détruites par les radicaux OH [41]. D'autre part, la durée de vie des aérosols sulfatés secondaires est corrélée avec l'humidité si celle-ci est au-delà de 80 %. En ce qui concerne l'effet de la température, une

augmentation exponentielle de la durée de vie peut être observée lorsque la température diminue [42].

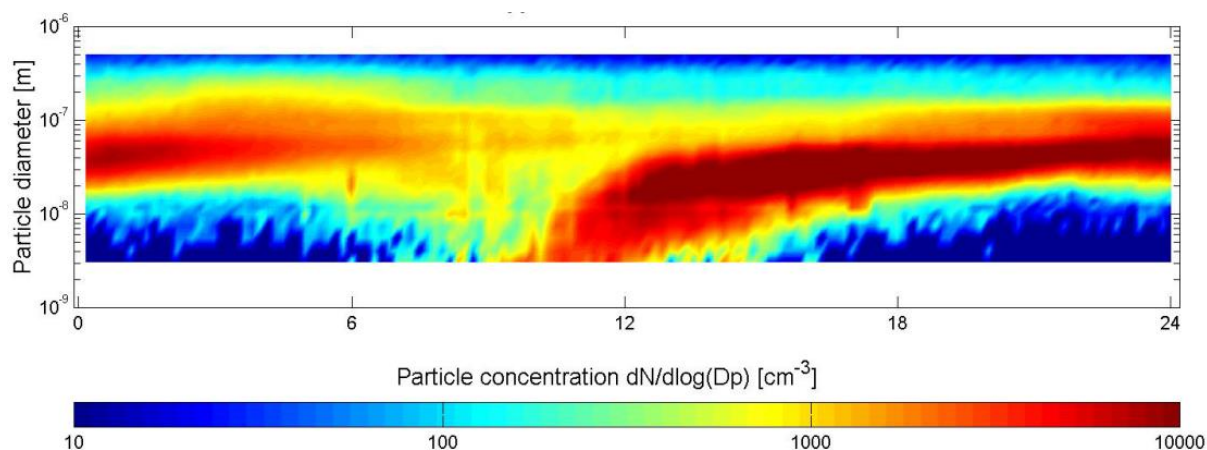


Figure I-10 Nucléation des particules dans une forêt boréale pendant une journée, concentration mesuré à 2 m de hauteur à la forêt de Hyytiälä en 19-05-1999. D_p est le rayon de la particule et N le nombre de particule par unité de surface [40].

Des aérosols inorganiques comme les gouttelettes d'acide sulfurique aqueux, se forment dans la troposphère par transformation du dioxyde de soufre (SO_2), sulfure de diméthyle (DMS) ou le disulfure de carbone (CS_2). Grâce à la longue durée de vie de l'oxysulfure de carbone (OCS), il atteint la stratosphère. L'oxydation de ce dernier constitue la source majeure des aérosols de fond dans la stratosphère, en effet, sous l'action des rayonnements UV il se dissocie puis forme l'acide sulfurique [17]. Des aérosols de nitrates sont également présents, ils sont générés à l'issue de l'interaction des sels marins avec les émissions anthropiques.

2.3. Durée de vie des principaux aérosols atmosphériques

Le vent transporte les aérosols atmosphériques sur de longues distances. Il existe différents mécanismes qui modifient la concentration des particules au cours de leurs déplacements. D'abord, par l'effet de la gravité, ces derniers se sédimentent en se déposant progressivement sur la surface. La turbulence atmosphérique limite ce processus via le brassage vertical. Un dépôt humide est aussi observé dans la nature, la pluie emporte vers la surface ces particules ce qui a un effet de « lessivage » de l'atmosphère.

Les particules de poussière sont transportées par le vent pendant des durées qui peuvent atteindre plusieurs journées, et atteignent des altitudes élevées (2.5 à 6 km) [43]. Au nord de

l'Afrique, ces émissions constituent la source majeure (46% à 60%) des émissions et dépôts de poussière désertique globale [44]. Une estimation basée sur la modélisation [45] dont le résultat est résumé au Tableau I-1 montrent que la quantité globale annuelle moyenne de poussière émise est de 1 650 Tg.an⁻¹. La majorité se dépose sans pluie. La quantité globale des aérosols sulfatés est de 39.3 Tg.S.an⁻¹ en moyenne. Environ 12 % des sulfates se dépose par sédimentation et le reste par lessivage. Pour les particules de carbone organique (OC) 104 Tg.C.an⁻¹, le dépôt sec est estimé à 19% de la quantité totale en moyenne. Pour le carbone suie (BC) 18 Tg.C.an⁻¹ est émis dont 22% se dépose par sédimentation. Enfin, la plus grande quantité d'émission est celle des sels marins qui vaut 6 950 Tg.S.an⁻¹ dont seulement 1% se dépose par lessivage.

La durée de vie moyenne de ces aérosols est comparable Tableau I-1, surtout pour les sulfates, BC, et les poussières, environ 6 jours. Un temps de 5 jours est estimé pour le carbone organique. Alors que pour le sel marin, le temps de vie est minimal, presque une demi-journée (0.6 jour) car la grande taille de la particule de sel et la forte humidité relative de l'environnement favorise le dépôt.

Tableau I-1 Bilan global moyen des années 1990, 1996 et 1997 de différents types d'aérosols [45].

Type d'aérosol	Source Tg.M.an ⁻¹	Dépôt sec Tg.M.an ⁻¹	Dépôt humide Tg.M.an ⁻¹	Temps de vie jours
Sulfates	39.30	4.73 (12%)	34.57 (88%)	6.07
OC	104.00	20.23 (19%)	83.40 (80%)	5.03
BC	18.00	3.97 (22%)	14.03 (78%)	6.07
Poussière	1,650.00	1,483.33 (89%)	183.33 (11%)	6.10
Sel marin	6,950.00	6,870.00 (99%)	74.00 (1%)	0.60

2.4. Impacts des aérosols

2.4.1. Effets sur la santé

Les aérosols ont un effet néfaste sur la santé de l'homme. On estime que plus de 4 millions de mortalité attribuée à l'exposition aux particules fines (PM_{2.5}) en 2015 [46]. Ces particules détériorent le système cardiovasculaire [47] et respiratoire [3] suite à l'exposition aigu ou chronique.

La toxicité dépend de la nature, la quantité et la taille des particules. Celles de taille inférieur à 2 μm se déposent aux alvéoles pulmonaires. Le dépôt d'aérosols dans les poumons est principalement régi par les mécanismes d'impaction inertielle, de sédimentation gravitationnelle, de diffusion brownienne et dans une moindre mesure, par la turbulence, la précipitation électrostatique et l'interception [48].

Les aérosols peuvent transporter des agents pathogènes microbiens fongiques et viraux, ce qui peut entraîner la propagation des maladies dans de différentes régions du monde [49]. Plusieurs études de poussière collectée sur les surfaces de fermes avicoles, porcines et laitières ont montré que la poussière contenait des champignons et des bactéries [50]. Des études sur la fièvre aphteuse ont montré la possibilité du transport du virus à plusieurs centaines de kilomètres en présence des conditions météorologiques favorables [50].

L'exposition chronique aux particules acides pourrait avoir un effet délétère à la croissance, le développement et au fonctionnement des poumons [4]. Les particules de matière peuvent causer même des maladies cardiovasculaire [51] et plusieurs recherches montrent un lien entre le cancer et l'exposition aux aérosols.

2.4.2. Effets sur l'environnement

Les aérosols de poussière minérale ont un effet bénéfique pour la biodiversité dans la planète. En effet, les poussières minérales transportées par le vent constituent une source de fertilisation en fer pour les océans [52]. Pareillement, la poussière du nord d'Afrique contribue fortement à la fertilisation de la forêt d'Amazon [53]. La contribution est très variable d'une année à l'autre, en moyenne on estime 28 Tg de poussière par an [54]. Par contre, ces aérosols contribuent à la réduction de la visibilité (Figure I-11) ce qui pourrait perturber la circulation aérienne. En plus, ils se déposent sur les surfaces exposées, en modifiant l'environnement urbain, voir réduire le rendement des panneaux photovoltaïques.



Figure I-11 Dégradation de la visibilité à Paris lors de l'épisode de pollution du 14 mars 2014 causée par les PMs. © AFP.

www.lexpress.fr.

Les aérosols qui atteignent l'altitude des nuages comme certaines émissions de feux de forêts, et de volcans changent les caractéristiques optiques de ces derniers et modifient la fréquence, l'intensité et l'étendu des précipitations [55]. Les aérosols tels que les sulfates et le carbone organique jouent un rôle important dans la formation des nuages en agissant comme des noyaux de nucléation des gouttelettes [56]. En revanche, les aérosols d'acide sulfurique (H_2SO_4) et d'acide nitrique (HNO_3) résultent des pluies acides qui impactent négativement la vie aquatique. Elles causent en plus de la mortalité de poissons, un échec de la reproduction, une réduction de la croissance, des déformations squelettiques et une augmentation de l'absorption des métaux lourds [57]. Le nombre de phytoplanctons est impacté négativement également [57]. L'acidification du sol cause la déplétion des nutriments tels que le potassium, le calcium et le magnésium nécessaires à la croissance des plantes, ce phénomène est appelé la lixiviation. Enfin, on observe un endommagement des arbres des forêts, au niveau des feuillages et des racines. D'autres effets négatifs de ces pluies touchent directement la vie et l'économie locale tels que la dégradation des matériaux de construction et la contamination de la nourriture [57]. D'autre part, le dépôt de l'azote inorganique cause l'eutrophication des surfaces marines et impactent la biodiversité en conséquence [58].

Par ailleurs, une modification de la stabilité du champ électrique entre la terre et l'atmosphère peut être engendrée. En effet, en l'absence d'aérosols, l'eau et d'autres gaz forment des clusters ionisés par le rayonnement cosmique. Ces clusters sont les responsables de la conductivité de l'atmosphère. La présence d'aérosols crée un gradient dans le champ électrique car les aérosols (tel que les poussières) ont généralement une conductivité électrique faible.

De manière générale, les aérosols sombres tels que le carbone suie absorbent le rayonnement solaire en contribuant dans le réchauffement climatique, alors que les aérosols de couleur claire, principalement les aérosols secondaires, réfléchissent le rayonnement vers l'espace en contrariant l'effet de serre. Ce forçage radiatif direct dû à l'interaction rayonnement-matière (voir page 46 de cette thèse) s'additionne avec l'impact semi-direct qui est le changement de la stabilité de l'atmosphère et la couverture nuageuse causé par le chauffage de la couche d'air locale contenant les aérosols. Enfin, les aérosols contribuent aussi indirectement au forçage radiatif en modifiant les caractéristiques des gouttelettes de nuages comme leurs tailles et leurs concentrations ce qui aura un effet sur la durée de vie du nuage et son albédo [59,60]. La combinaison des forçages radiatifs positifs des aérosols et des gaz à

effet de serre, en plus du forçage négatif des aérosols est estimé dans la Figure I-12. On voit que le carbone suie contribue positivement à cause de l'absorption du rayonnement, alors que le carbone organique a un effet refroidissant sur la planète avec une contribution partagée entre l'effet direct de diffusion du rayonnement, et d'effet indirect lié à l'interaction avec les nuages. Ces derniers, s'ajoutent à la contribution conséquente de -0.94 W.m^{-2} des émissions de dioxyde de soufre qui est un gaz précurseur des aérosols sulfatés, mais ne compensent pas quand même le réchauffement causé par les gaz à effet de serre (CO_2 , CH_4 ...). D'autre part, l'effet des émissions d'ammoniac qui est un gaz précurseur aux aérosols de nitrates contribuent uniquement de -0.03 W.m^{-2} ce qui est quasiment négligeable par rapport à celle causé par le dioxyde de soufre (Figure I-12).

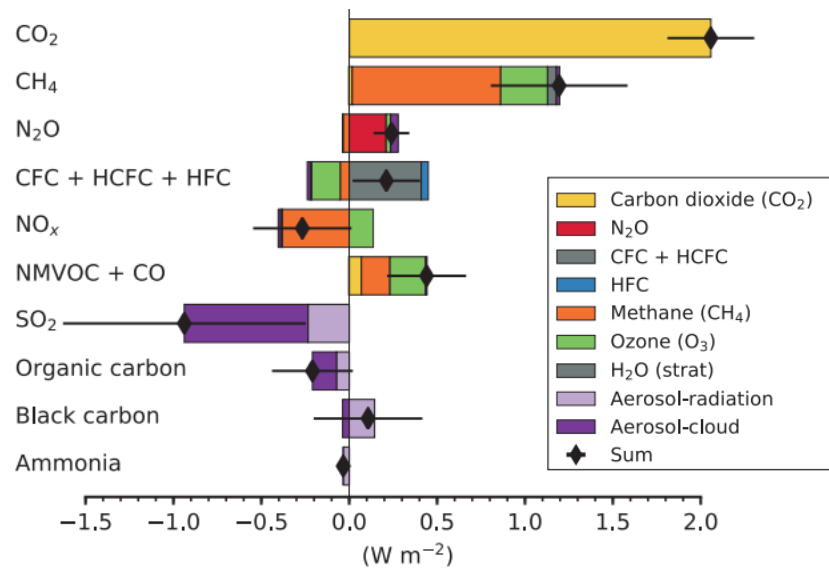


Figure I-12 Estimation de la contribution au forçage radiatif effectif de la planète pour quelques émissions entre 1970 et 2019 [61].

3. Interaction des particules avec le rayonnement

Les propriétés optiques des aérosols atmosphériques sont sujet d'étude depuis plusieurs décennies afin de pouvoir quantifier leurs abondances et modéliser leurs impacts. Les interactions entre les particules et le rayonnement électromagnétique peuvent avoir lieu dans de nombreuses gammes spectrales, et elles dépendent des caractéristiques intrinsèques de la particule. Cette section est consacrée aux interactions microphysiques utiles pour le prochain chapitre mais pas aux réactions photochimiques.

3.1. Le rayonnement électromagnétique

L'onde électromagnétique est un champ électrique couplé à un champ magnétique oscillant à la même fréquence, de manière perpendiculaire au sens du déplacement et avec une vitesse d'environ $3 \cdot 10^8 \text{ m s}^{-1}$ dans le vide. Si la lumière oscille sur seul plan, on dit que la lumière est **polarisée**, la polarisation peut être linéaire, circulaire, ou elliptique, selon le sens de la vibration. Si les oscillations se produisent sur plusieurs plans, on dit que la lumière est **non polarisée**. L'intensité totale d'un rayonnement I_{tot} est la somme du rayonnement polarisé I_{pol} et du rayonnement non polarisé I_{npol} .

$$I_{\text{tot}} = I_{\text{pol}} + I_{\text{npol}} \quad (\text{I.8})$$

Le degré de polarisation est défini comme le ratio $I_{\text{pol}}/I_{\text{tot}}$ [62]. On décrit le rayonnement dans le cadre des études atmosphériques par quatre quantités: l'intensité, le degré de polarisation linéaire, direction de polarisation, et le degré de polarisation circulaire (ellipticité) [62].

L'onde électromagnétique se propage en véhiculant de l'énergie E qu'on peut quantifier par l'équation suivante :

$$E = h\nu \quad (\text{I.9})$$

où ν est la fréquence et h est la constante de Planck.

L'émission d'un rayonnement électromagnétique provient de l'agitation thermique des atomes et des molécules. A la fin de l'année 1900, Max Planck (1858-1947) a formulé l'équation du spectre d'émission électromagnétique d'un corps chaud, qui est en accord avec l'expérimentation. Cette dernière, ci-dessous, complète l'hypothèse de son collègue Wilhelm Wien (1864-1928) qui stipule que la longueur d'onde maximale du spectre d'émission par un corps chaud est inversement proportionnelle avec sa température [63] (Figure I-13).

$$B_{\nu}(T) = \frac{8\pi\nu^2}{c^3} \frac{h\nu}{e^{h\nu/kT} - 1} \quad (\text{I.10})$$

Dans la loi de Planck, B_{ν} est la densité d'énergie émise depuis un corps à température absolue T . La valeur de k est la constante naturelle de Boltzmann, et c est la vitesse du rayonnement dans le vide.

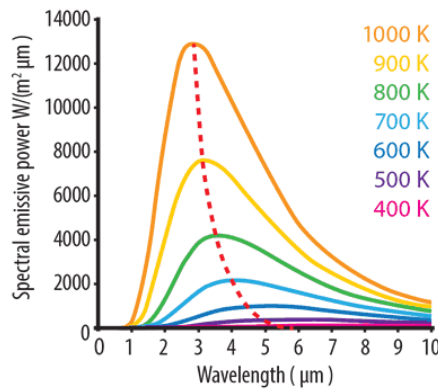


Figure I-13 Distribution d'énergie en fonction de la longueur d'onde pour quelques corps de température différentes [64].

Un spectre électromagnétique regroupe donc plusieurs fréquences radiatives. Elles peuvent aller de quelques Hz jusqu'à quelques centaines de GHz (Figure I-14). La fréquence détermine la quantité d'énergie véhiculée avec ce dernier.

Le soleil émet un rayonnement polychromatique dont le spectre correspond à une température d'environ 5 778 K. Mais chaque corps chaud, comme par exemple le sol, l'air, les nuages ou les aérosols émet un spectre de rayonnement électromagnétique correspondant à sa température. Les capteurs mesurent la luminance à plusieurs longueurs d'ondes, certains mesurent sa polarisation également. Ces derniers peuvent être utilisées pour déduire la composition de l'atmosphère en termes d'aérosols et de gaz.

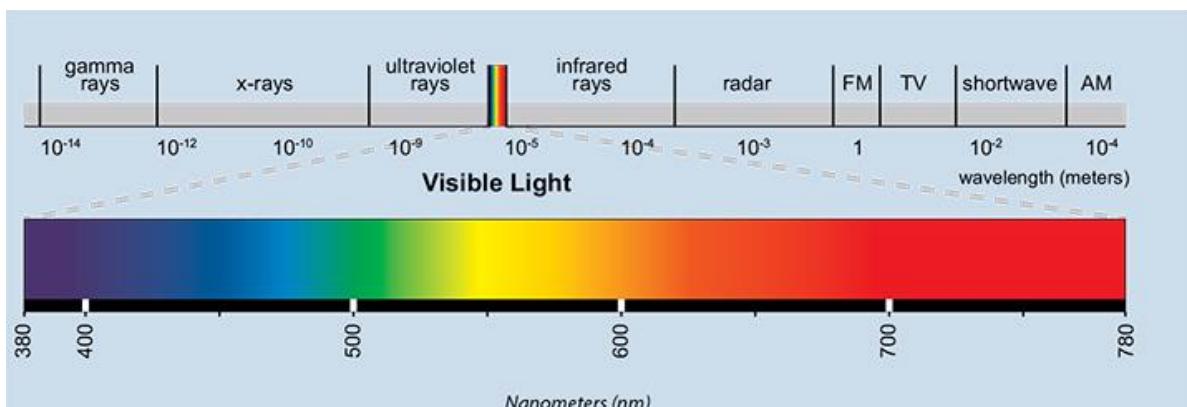


Figure I-14 Le spectre électromagnétique solaire

Le flux radiatif se propage dans l'espace de manière directionnelle, on définit la luminance aussi appelée radiance monochromatique I_λ pour une longueur d'onde λ comme le

flux d'énergie E traversant une unité de surface S par l'angle solide élémentaire Ω . Elle est souvent exprimée en $\text{W}\cdot\text{m}^{-2}\cdot\mu\text{m}^{-1}\cdot\text{sr}^{-1}$ [12].

$$I_\lambda = \frac{dE_\lambda}{dt ds d\lambda d\Omega} \quad (\text{I.11})$$

3.2. Absorption

La matière est composée d'atomes et de molécules. Ces particules sont généralement dans un état énergétique stable. Néanmoins, elles peuvent être excitées en absorbant de l'énergie pour atteindre un niveau d'énergie plus haut. L'absorption du rayonnement est sélective et quantifiée. Seuls les photons d'énergie correspondant à une transition énergétique autorisée sont absorbés. Il peut y avoir des transitions avec changements de niveau d'énergie électronique, vibrationnelle, ou rotationnelle de la particule (atome/molécule). L'absorption sélective nous permet d'élaborer des spectres d'absorption (Figure I-15) qui servent ensuite à identifier la composition d'un échantillon de matière aérosols par exemple en utilisant un spectromètre électromagnétique.

Afin de quantifier l'absorption du rayonnement par une particule, on utilise la **section efficace** d'absorption σ_{abs} (l'indice *abs* signifie absorption dans ce qui suit), c'est une grandeur exprimée en termes de superficie et elle est liée à la probabilité d'une particule d'absorber un rayonnement, elle est calculée comme le rapport du **coefficient d'absorption** k_{abs} sur la concentration C .

$$\sigma_{abs} = \frac{k_{abs}}{C} \quad (\text{I.12})$$

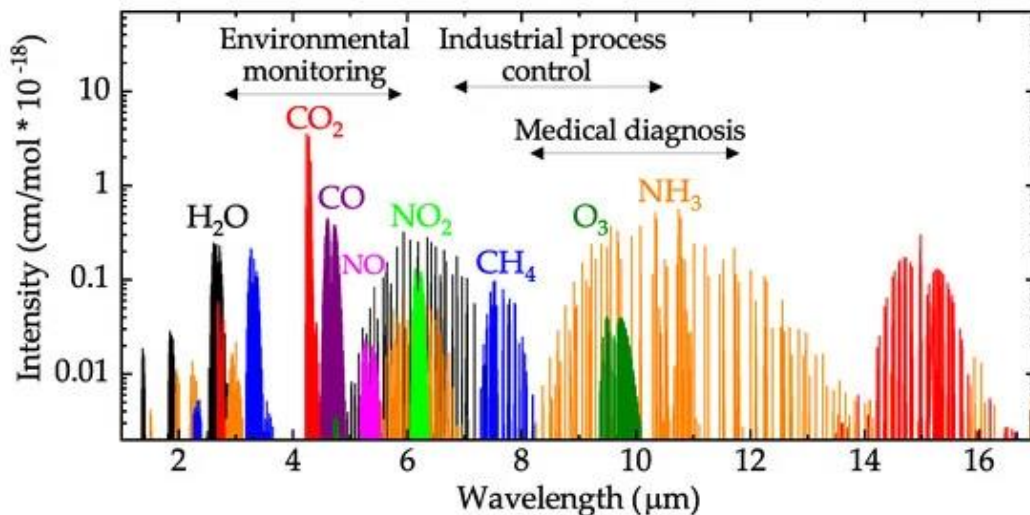


Figure I-15 Spectre d'absorption de quelques espèces atmosphériques [65].

3.3. Diffusion simple

La diffusion du rayonnement par une particule est la redirection du photon. Ceci dépend de la forme, du diamètre r de la particule et de la longueur d'onde du rayonnement λ . Le comportement du rayonnement lors de sa diffusion à travers les particules sphériques peut être décrit suivant deux modèles. Le **paramètre de taille** x (ci-dessous) nous permet de savoir la manière dont la particule diffuse le rayonnement, et ainsi le modèle le plus adéquat pour décrire cette diffusion. Les particules de petite taille $0.002 < x < 0.2$ diffusent le rayonnement suivant un modèle développé par le physicien Lord Rayleigh [1842–1919], alors que pour les particules de grande taille $0.2 < x < 2000$ le modèle de diffusion de Gustav Mie [1868–1957] est le plus réaliste. Au-delà de ces valeurs, les particules seraient trop grandes et leurs interactions avec le rayonnement suivent les lois de l'optique géométrique (Figure I-16).

$$x \cong \frac{2 \pi r}{\lambda} \quad (\text{I.13})$$

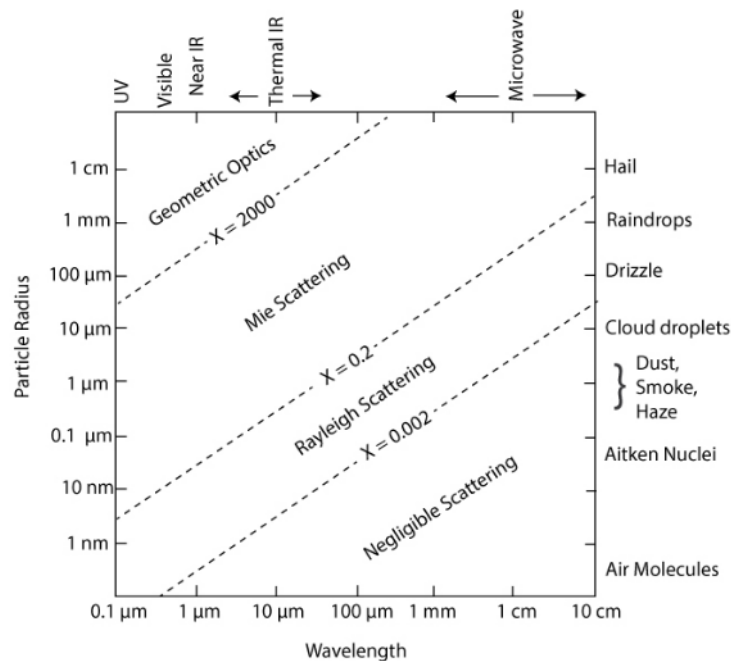


Figure I-16 Paramètre de taille x en fonction du rayon de la particule et de la longueur d'onde. Le type du rayonnement et de la particule sont montrés en haut et à droite respectivement [66].

L'angle de diffusion du rayonnement dans l'espace est modélisé par la **fonction de phase** notée $P(\theta)$ qui représente la probabilité qu'un photon venant à un angle θ soit diffusé dans une direction particulière. Elle dépend de la taille de la particule (Figure I-17). Si l'intensité du rayonnement diffusé est la même pour toutes les directions, on dit que la fonction de phase est isotrope, sinon elle est anisotrope.

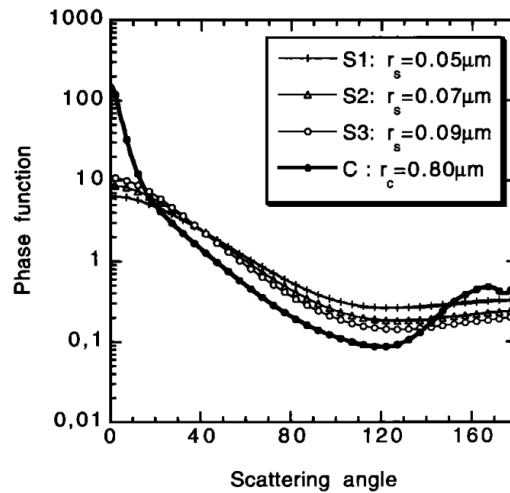


Figure I-17 Fonction de phase à 865 nm pour un aérosol grossier C, et trois aérosols fins S1, S2 et S3 [67].

3.3.1. Diffusion Rayleigh

Pour que la diffusion Rayleigh se produise, il faut que la particule soit dans un champ diélectrique homogène; et que le rayonnement pénètre la particule si rapidement que le champ propre de la particule est mis en place en un temps qui est court par rapport à la période du rayonnement incident [62]. Quand le rayonnement diffusé garde la même fréquence du rayonnement incident on parle de la diffusion élastique, sinon on dit que c'est une diffusion inélastique, comme les diffusions Raman et Rayleigh-Brillouin.

L'intensité du rayonnement diffusé I dans une direction θ pour une particule de petite taille par rapport à la longueur d'onde est égale à :

$$I(\lambda, \theta) = \frac{1}{\lambda^4} I_0(\lambda) \frac{2\pi^2(n^2-1)^2}{3N} \rho P(\theta) \quad \text{I.14}$$

La valeur I_0 est l'intensité du rayonnement incident avant la diffusion, n est l'indice de réfraction de l'air, ce dernier est le rapport entre la vitesse de la lumière dans le vide et sa vitesse dans l'air, ρ est la densité de l'air, et N la densité moléculaire [68].

La fonction de phase P pour un rayonnement non polarisé est montrée dans l'équation ci-dessous, avec μ égale à $\cos(\theta)$, pour un rayonnement polarisé, elle varie selon l'angle d'incidence du rayonnement par rapport au plan de polarisation (Figure I-18).

$$P(\mu) = \frac{3}{4}(1 + \mu^2) \quad (\text{I.15})$$

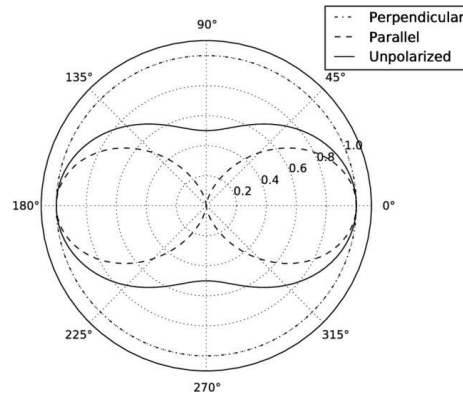


Figure I-18 Fonction de phase de la diffusion Rayleigh pour un rayonnement polarisé et non polarisé [69].

La section efficace moyenne σ_{sca}^{moy} pour une particule de gaz anisotrope pour une diffusion Rayleigh est égale à:

$$\sigma_{sca}^{moy} = \frac{k_{diff}}{N} = \frac{8\pi^3 (n_\theta^2 - 1)^2}{3 \lambda^4 N^2} \frac{6+3\delta}{6-7\delta} \quad (\text{I.16})$$

où k_{diff} est le coefficient de diffusion (l'indice *diff* sera utilisé pour désigner diffusion dans ce qui suit), N est le nombre de molécules par unité de volume, n_θ est l'indice de réfraction, δ est le **facteur de dépolarisation** qui est le ratio des intensités parallèles et perpendiculaires du plan de la diffusion. Ce facteur prend en général des valeurs $0 \leq \delta \leq 0.5$. Pour la diffusion isotrope $\delta = 0$ [62].

Rayleigh a expliqué la couleur bleue du ciel par la diffusion du rayonnement solaire par les molécules de l'air [70]. Comme la diffusion est inversement proportionnelle à la longueur d'onde (équation ci-dessus), le rayonnement bleu est diffusé plus comparé aux ceux de plus faibles fréquences comme le rouge.

3.3.2. Diffusion de Mie

La théorie de Mie présente un modèle pour les particules (aérosols) sphériques. La fonction de phase de Henyey-Greenstein introduite en 1941 traduisait bien le comportement du

rayonnement pour la diffusion de Mie. Plus le paramètre de taille x est grand, plus le rayonnement diffusé est focalisé vers le sens du flux radiatif (Figure I-19). Le **facteur d'asymétrie** $g \in [-1,1]$ présenté dans l'équation (I.17) nous renseigne sur cette particularité, et il dépend de la longueur d'onde (Figure I-20). Si $g > 0$ le rayonnement est diffusé favorablement vers l'avant, alors que à $g < 0$ le rayonnement est principalement rétrodiffusé. En 1992 W. Cornette et J. Shanks [71] proposèrent une fonction de phase plus réaliste qui assure la continuité des deux théories (Rayleigh et Mie). En effet, cette dernière, exprimée dans l'équation (I.18) converge vers la théorie de Rayleigh quand g tend vers 0.

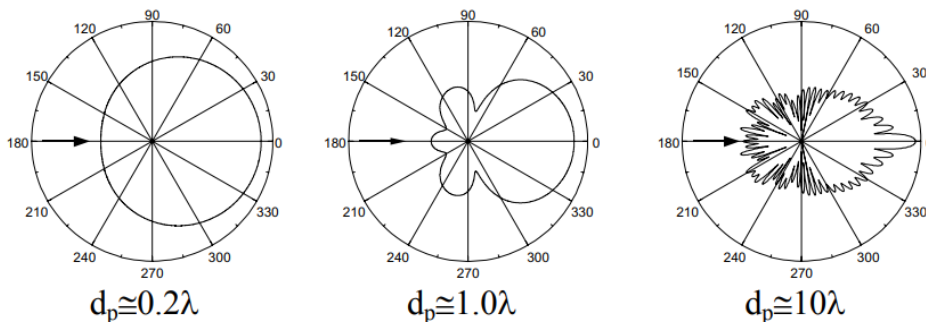


Figure I-19 Fonction de phase calculée avec la théorie de Mie pour trois paramètres de tailles différents [72].

$$g = \frac{1}{2} \int_{-1}^1 P(\mu) \mu d\mu \quad (\text{I.17})$$

$$P(\mu, g) = \frac{3(1 - g^2)}{2(2 + g^2)} \frac{(1 + \mu^2)}{(1 + g^2 - 2g\mu)^{3/2}} \quad (\text{I.18})$$

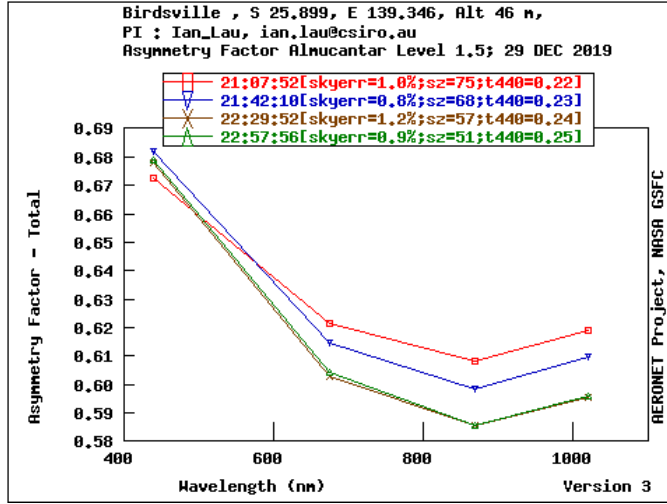


Figure I-20 Facteur d'asymétrie des aérosols observés par la station Birdsville en Australie pour quatre heures de la journée et pour quatre longueur d'onde [73].

3.3.3. Albédo de diffusion simple

L'albédo de diffusion simple ϖ est un nombre adimensionnel qui définit la fraction de l'énergie diffusée de l'énergie totale atténuée du rayonnement :

$$\varpi = \frac{k_{diff}}{k_{ext}} = \frac{\sigma_{diff}}{\sigma_{ext}} \quad (I.19)$$

tel que k_{ext} est le coefficient d'extinction (l'indice *ext* désigne extinction). Pour une particule donnée, la section efficace d'extinction σ_{ext} est la somme de la section efficace d'absorption σ_{abs} et celle de diffusion σ_{diff} .

$$\sigma_{ext} = \sigma_{abs} + \sigma_{diff} \quad (I.20)$$

La section efficace d'extinction d'une particule est liée à la section géométrique G (égal à πr^2 pour une sphère) par un facteur d'efficacité Q_{ext} .

$$\sigma_{ext} = G \cdot Q_{ext} \quad (I.21)$$

sachant que :

$$k_{ext} = k_{diff} + k_{abs} \quad (I.22)$$

$$\sigma_{ext} = \sigma_{diff} + \sigma_{abs} \quad (I.23)$$

$$Q_{ext} = Q_{diff} + Q_{abs} \quad (I.24)$$

La figure ci-dessous regroupe les valeurs d'albédo de diffusion simple de quelques types d'aérosols atmosphériques. Nous remarquons que les aérosols de poussière désertique diffusent plus de rayonnement que les aérosols de feux et les aérosols urbains. Ces derniers sont connus plutôt par leurs pouvoirs absorbants. Aussi, on voit que la dépendance spectrale de la diffusion est faible pour les poussières, alors que pour les aérosols de feux de biomasse et les aérosols urbains la diffusion diminue pour les grandes longueurs d'onde.

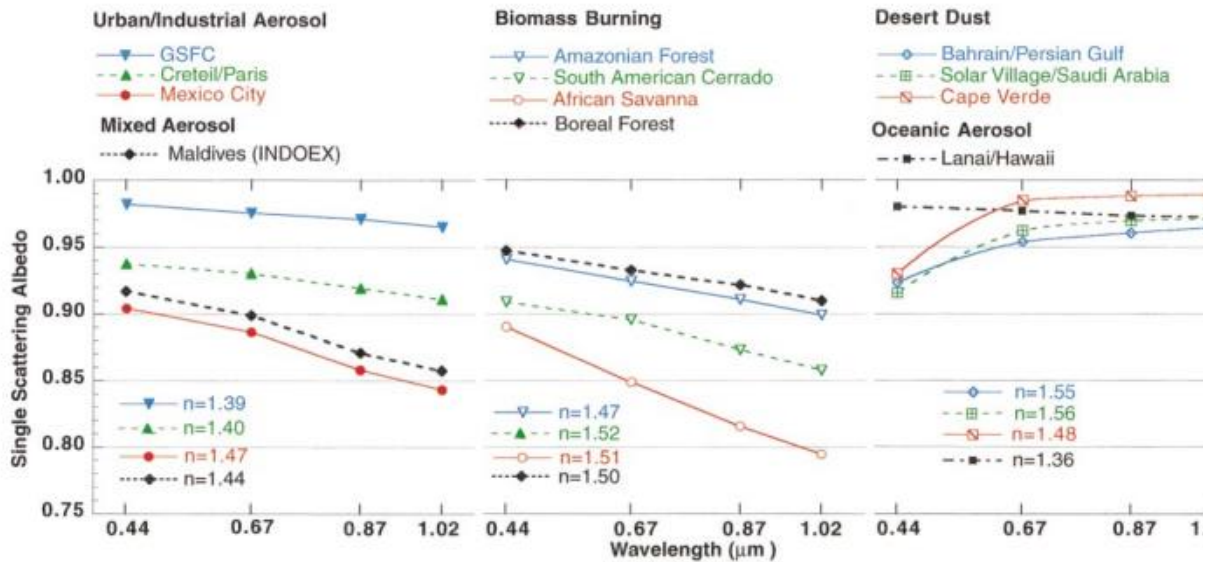


Figure I-21 Quelques valeurs d'albédo de diffusion simple de quelques types d'aérosols atmosphériques [74].

4. Atténuation du rayonnement dans l'atmosphère

Un milieu dilué homogène tel qu'une couche de l'atmosphère atténue le rayonnement qui la traverse en fonction du type et de la concentration des espèces qui la composent. Plus la concentration est grande, plus la lumière est atténuée, jusqu'à son extinction. Ce phénomène est décrit par la loi de Beer-Lambert ci-dessous pour un milieu gazeux d'épaisseur h .

$$I = I_0 e^{k_{ext}(\lambda) \rho h} \quad (I.25)$$

Lors de son passage, le photon de lumière peut être absorbé ou diffusé par les particules. L'**indice de réfraction complexe** m , sert à quantifier la capacité d'un corps à atténuer le rayonnement, il dépend de la longueur d'onde du rayonnement tel que :

$$m(\lambda) = n(\lambda) + k(\lambda) \quad (I.26)$$

dans l'équation ci-dessus, la composante imaginaire k décrit l'absorption du rayonnement par le milieu. La partie réelle n nous renseigne principalement sur le pouvoir diffusant du milieu.

Le coefficient d'absorption est lié avec la partie imaginaire de l'indice de réfraction par la relation :

$$k_{abs} = 4\pi k / \lambda \quad (I.27)$$

La dépendance spectrale de l'indice de réfraction complexe des différents aérosols de poussière est montrée dans la figure ci-dessous.

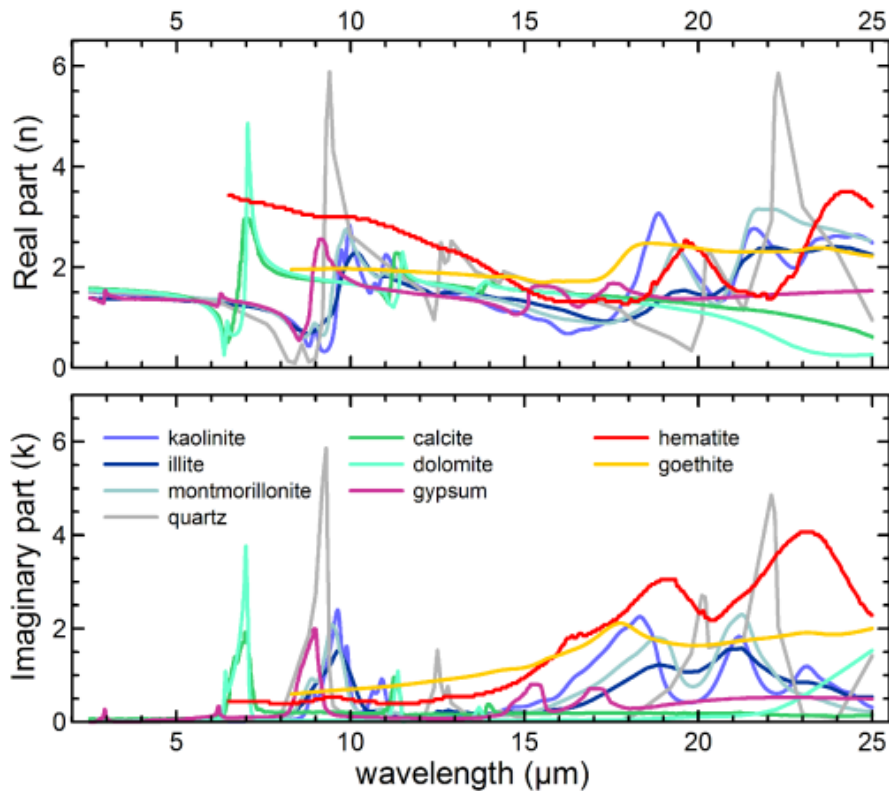


Figure I-22 Spectre d'indice de réfraction complexe de quelques composants minéraux de poussière dans la gamme spectrale de 2.5 à 25 μm, partie réelle tracée dans le panneau du haut et la partie imaginaire en bas [36].

Dans l'atmosphère, on trouve souvent les aérosols sous forme de panache constituant un milieu semi-transparent. Si le rayonnement est diffusé plusieurs fois, avant de quitter ce milieu, on parle d'une **diffusion multiple**. Pour calculer la diffusion multiple pour une atmosphère planétaire, celle-ci est généralement considérée composée de couches planes parallèles [62].

4.1. Épaisseur optique

On peut définir l'épaisseur optique τ (AOD pour *Aerosol Optical Depth*) comme une quantité caractérisant le pouvoir d'une couche d'aérosols d'atténuer (absorber et/ou diffuser) un rayonnement qui la traverse. Pour une longueur d'onde λ on écrit:

$$\tau = \int_h^\infty k_{ext}(\lambda, h) \rho dh \quad (\text{I.28})$$

On peut exprimer l'AOD aussi comme le logarithme naturel du rapport entre la luminance incidente I_λ^i et la luminance transmise I_λ^t de la couche d'aérosol :

$$\tau = \ln\left(\frac{I_\lambda^i}{I_\lambda^t}\right) \quad (\text{I.29})$$

L'épaisseur optique dans l'UV et le visible varie graduellement avec la longueur d'onde et la granulométrie des particules. En conséquence, les approches de télédétection des aérosols peuvent utiliser des capteurs satellitaires tel que celui de MODIS et VIIRS mesurant le rayonnement intégré sur des canaux (bandes) spectraux. En outre, en mesurant l'AOD dans deux longueurs d'onde λ_1 et λ_2 , nous pouvons déduire le **coefficient d'Ångström** α , exprimé ci-dessous. Ce dernier décrit la dépendance spectrale de l'AOD et il est lié à la taille des aérosols de manière inversement proportionnelle. On peut utiliser donc α pour déterminer l'AOD inconnu pour une longueur d'onde λ_2 si on connaît sa valeur pour une autre longueur d'onde λ_1 .

$$\alpha = \frac{\ln\left(\frac{AOD(\lambda_1)}{AOD(\lambda_2)}\right)}{\ln\left(\frac{\lambda_1}{\lambda_2}\right)} \quad (\text{I.30})$$

5. Conclusions

Dans ce chapitre, nous avons présenté quelques aspects des aérosols atmosphériques, leurs diversité, impacts et particularité en ce qui concerne leurs interactions avec la lumière. Ceci montre l'importance de l'évaluation de l'abondance des aérosols dans l'atmosphère non seulement via l'épaisseur optique, mais aussi en termes d'altitudes à laquelle ils se trouvent. La connaissance de la nature des interactions entre la particule d'aérosol et la lumière nous permet de comprendre comment estimer leurs abondances à distance dans le chapitre suivant.

II. Chapitre 2 : Quantification des aérosols dans l'atmosphère

Ce chapitre présente un aperçu des différentes méthodes employées pour estimer l'abondance des aérosols, avec une vue plus détaillée sur la télédétection. Le principe de la simulation régionale de la répartition spatio-temporelle des aérosols par la modélisation de chimie-transport est aussi décrit. Ces deux méthodes sont employées dans les travaux de la thèse, présentés dans les Chapitres III et IV. Les limitations de chacune de ces approches sont également discutées.

1.	Mesures in situ	61
2.	Télédétection.....	62
2.1.	Technique active	62
3.2.	Technique passive.....	64
3.2.1.	Capteur au sol	64
3.2.2.	Capteurs embarqués	65
4.	Observation des aérosols depuis l'espace.....	71
4.1.	Méthode d'inversion de mesures passives de luminance	72
4.2.	Équation de transfert radiatif	74
4.2.1.	Modélisation direct du transfert radiatif	75
4.2.2.	Inversion	78
4.2.3.	Évaluation qualitative de la qualité de l'inversion.....	79
4.2.4.	Analyse des erreurs de l'inversion.....	80
4.2.5.	Régularisation Tikhonov-Phillips	81
5.	Modélisation régionale de la qualité de l'air.....	81
5.1.	Le modèle de chimie-transport CHIMERE	82
5.2.	Émission.....	83
5.3.	Advection.....	83
5.4.	Dépôt sec et humide.....	83
5.5.	Chimie.....	84
5.6.	Propriétés optiques.....	84
6.	Conclusion	85

1. Mesures in situ

Il existe plusieurs techniques pour estimer le nombre de particules d'aérosols, la distribution de taille, la composition chimique, et les propriétés optiques. Les mesures peuvent être réalisées par des mesures au niveau de la surface ou en altitude à l'aide d'un système aéroporté. D'autre part, l'analyse de l'air peut se faire en continu ou en utilisant des filtres à aérosols pour piéger les particules à analyser au laboratoire ultérieurement.

La microbalance oscillante à élément conique fournit une mesure de la concentration massique des PMs en temps quasi réel. Elle utilise un principe gravimétrique : un filtre est placé à l'extrémité d'un tube conique fixé d'une seule extrémité et en oscillation libre. Le dépôt de particules sur le filtre engendre une augmentation de sa masse et ralentit les oscillations en conséquence [75].

Afin de mesurer une distribution de taille des PM, on utilise les analyseurs de mobilité différentielle. Ces instruments reposent sur une technique adaptée pour les particules chargées et se base sur le fait que les particules chargées se déplacent dans un champ électrique de manière qui dépend de leurs diamètres et de leurs charges élémentaires qu'elles portent. Faire varier la tension électrique permet d'avoir des mesures de concentration d'aérosols par taille.

Le nombre total d'aérosols est mesuré avec un compteur de particules à condensation. L'instrument peut généralement détecter la taille des aérosols qui va de quelques nanomètres jusqu'à quelques micromètres. Le principe est de faire passer le flux d'air contenant les aérosols ultrafins à mesurer dans un milieu gazeux super saturé qui va se condenser rapidement sur les particules en les faisant grossir en conséquence, jusqu'à une taille détectable par l'appareillage optique. Les particules sont comptées par détection du rayonnement diffusé au passage par un faisceau de laser. La technique est parfois utilisée en combinaison avec un module thermique afin d'enlever les composés volatiles pour effectuer des analyses d'aérosols sélectives. On peut également faire une mesure de la distribution de taille en la combinant avec un analyseur différentiel [76].

Bien que la mesure in situ des aérosols soit précise, elle présente quelques inconvénients. Les campagnes de mesures aéroportées sont coûteuses. En outre, ces mesures sont très localisées et les stations de surveillance sont concentrées dans les pays développés ce qui rend difficile l'étude à l'échelle globale.

2. Télédétection

Il est possible de mesurer à distance les propriétés optiques et microphysiques des aérosols. Le principe repose sur le fait que les aérosols diffusent et absorbent la lumière (voir Section 3 du Chapitre I). La concentration des espèces atmosphériques intégrée sur une colonne atmosphérique de rayon élémentaire peut être mesurée par télédétection, l'unité de mesure dans ce cas est particules par élément de surface (mole.cm^{-2}). On distingue deux types de télédétection pour les aérosols dans l'atmosphère. Une méthode passive, qui exploite le rayonnement naturel des corps chauds célestes rayonnant (le soleil par exemple), ou des corps illuminés (la lune par exemple). La deuxième méthode dite active utilise une source de rayonnement artificielle comme un laser.

Les montages de télédétection d'aérosols existants sont déployés soit au sol ou embarqués. Les stations d'observation depuis le sol bénéficient de quelques avantages des stations de mesures in situ (maintenance et calibrations périodique, l'amélioration de l'équipement en continue, etc.). Par contre, l'observation est localisée, alors que les télémessures aéroportées bénéficient de leurs mobilités en échange de la précision.

2.1. Technique active

La détection active d'aérosols à distance est possible grâce à la modélisation de la rétrodiffusion du rayonnement émis. On envoie de puissantes pulsations de faisceau de laser dans l'atmosphère, les particules présentes dans le trajet optique du rayonnement renvoient une partie de l'énergie dans la direction de la source. Le capteur du télescope, monté près du laser, reçoit les photons rétrodiffusés et les convertit en signaux numériques. Le délai temporel entre l'émission du laser et la réception du signal renseigne sur l'altitude des aérosols, alors que l'intensité du signal sert à estimer leurs abondances. Cette technologie appelée LIDAR (*Light Detection and Ranging*) peut être montée au sol comme sur les avions et les satellites. L'avantage de ces systèmes est qu'ils sont plus précis comparés aux systèmes passifs, en plus, ils sont capables de détecter les multiples couches d'aérosols même durant la nuit.

3. Lidar embarqué

Des mesures directes des profils verticaux des aérosols sont obtenues grâce aux capteurs actifs tel que celui de l'instrument satellitaire CALIOP (*Cloud-Aerosol Lidar with Orthogonal Polarization*). Cet instrument est un lidar à bord de CALIPSO (*Cloud Aerosol*

Lidar and Infrared Pathfinder Satellite Observations) (Figure II-1) qui mesure les profils verticaux de rétrodiffusion des aérosols et des nuages à résolution spatiale fine de 30 m entre la surface et 8 km d'altitude à la longueur d'onde 532 nm. L'instrument opère à 1064 nm également avec une résolution verticale de 60 m. La résolution horizontale est néanmoins limitée à la trace du satellite, et ne dépasse pas 5 km [77] ce qui ne permet pas la distribution 3D des observations. La couverture globale correspond donc à des transects latitudinaux espacés d'environ 2000 km en longitude pour un même jour. Grâce à la mesure du degré de dépolarisation, des produits sont dérivés des mesures tel que des spécificités typologiques sur les scènes observées par rapport aux aérosols et aux nuages. On observe également les bases et les sommets des couches de ces derniers, ainsi que les profils d'extinction des aérosols (voir Figure II-2) [77]. Ces mesures lidar sont utilisées dans ma thèse pour la validation des observations de la distribution verticale des aérosols avec la méthode développée qu'on va aborder au Chapitre III.



Figure II-1 Image de synthèse du satellite CALIPSO [78].

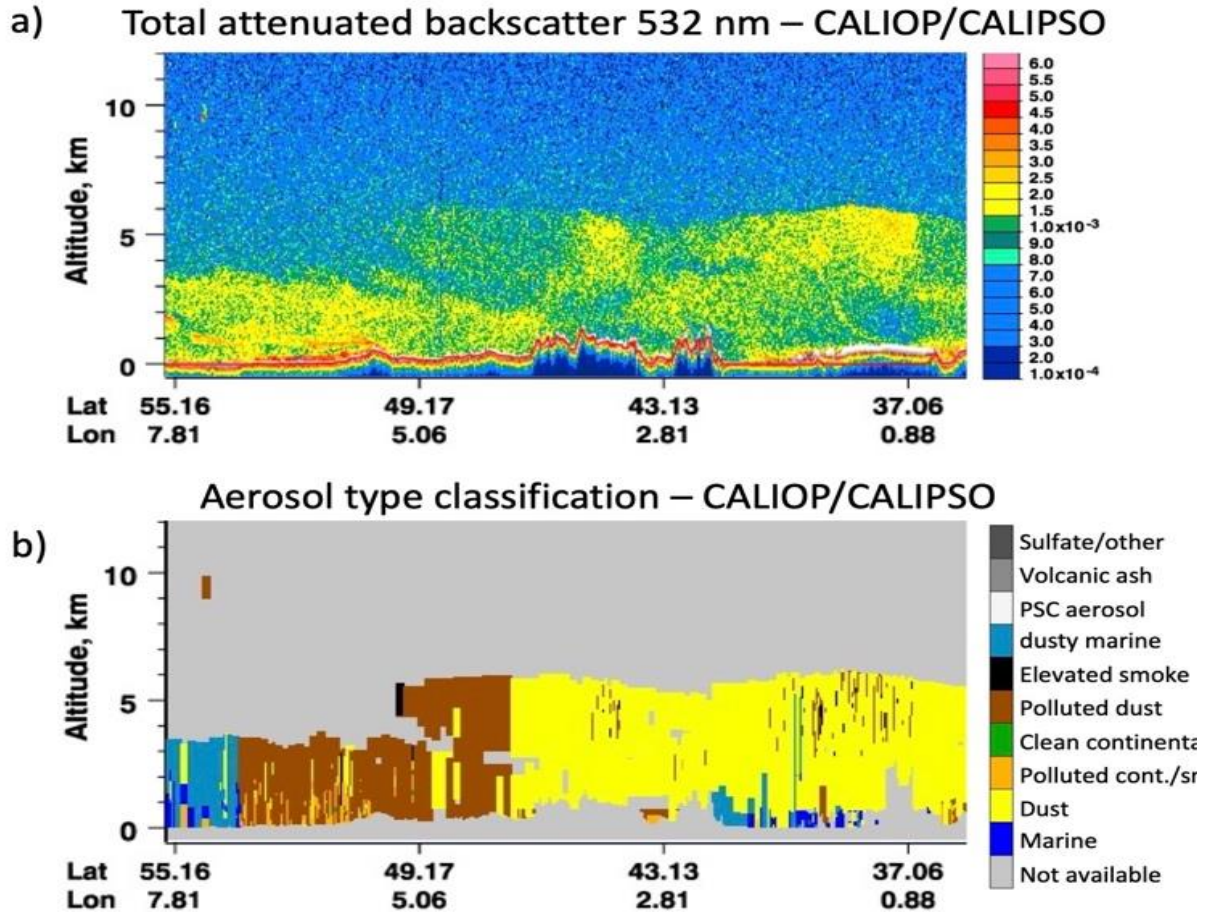


Figure II-2 Transects des profils de (a) rétrodiffusion atténuée des aérosols observée par CALIOP embarqué sur CALIPSO et (b) types d'aérosols déduits des mesures CALIOP, le 26 juin 2019. [échange privé J. Cuesta, 2022].

3.2. Technique passive

Les systèmes de télédétection passive des aérosols utilisent des photomètres pour mesurer le rayonnement solaire à plusieurs longueurs d'onde qui sera utilisé pour restituer les caractéristiques des aérosols présents dans l'atmosphère entre la source du rayonnement et le capteur photosensible de mesure.

3.2.1. Capteur au sol

Les observations photométriques passives au sol sont largement utilisées pour la recherche scientifique, la caractérisation des aérosols, la validation des produits satellitaires, et pour la synergie avec d'autres bases de données. Le réseau AERONET (Aerosol Robotic Network) [79] est le plus populaire dans cette catégorie, il est constitué de plus de 1500 sites (Figure II-3a) de surveillance permanents et temporaires équipés de photomètres solaires

passifs (Figure II-3b). Le réseau regroupe un ensemble de partenaires internationaux mis en place par la NASA (*National Aeronautics and Space Administration*) et PHOTONS (PHOtométrie pour le Traitement Opérationnel de Normalisation Satellitaire) composé de l'université de Lille 1, CNES (Centre National des Etudes Spatiale), et CNRS-INSU (Centre National de la Recherche Scientifique, Institut national des sciences de l'Univers du CNRS). La couverture d'AERONET est étendue grâce à d'autres réseaux et aux collaborateurs d'agences nationales, d'instituts, d'universités, de scientifiques individuels et des partenaires.

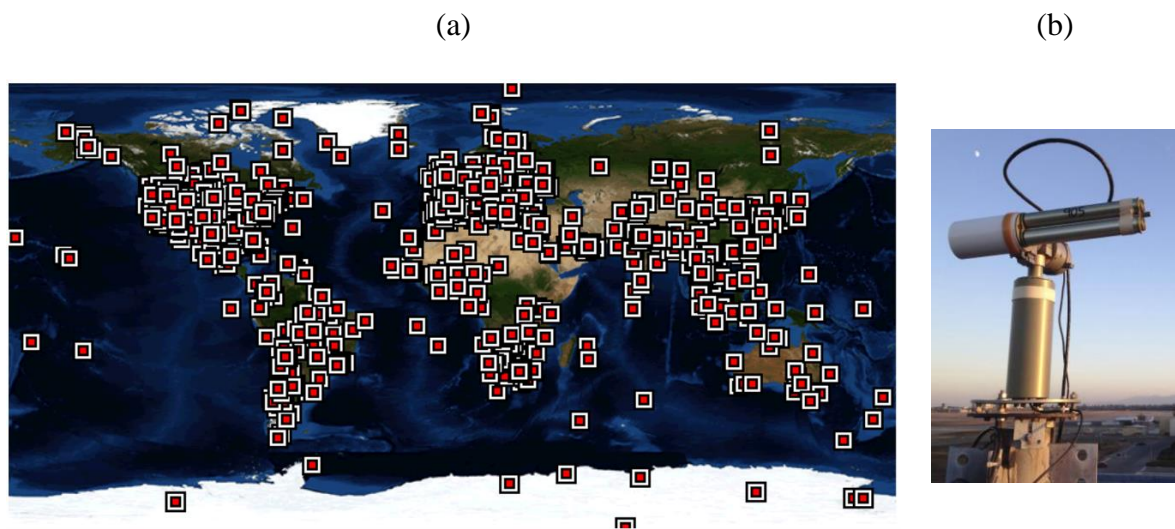


Figure II-3 (a) Carte de la distribution globale des stations du réseau AERONET [80], (b) Image d'un photomètre solaire du réseau AERONET [81].

Chaque station est constituée de bras robotisés pointant régulièrement vers le soleil pendant la journée; de l'information est ensuite extraite de l'irradiance solaire et de la radiance atmosphérique mesurées avec le photomètre [82]. Le même dispositif est utilisé pour caractériser les aérosols pendant la nuit grâce à la luminance lunaire. Le réseau impose la standardisation et la calibration des instruments. Les données restituées sont les propriétés optiques (AOD, albédo de diffusion simple, indice de réfraction imaginaire) et microphysiques (distribution en taille, diamètres effectifs...) des aérosols. Voir section 3 du Chapitre I pour la définition de ces grandeurs.

3.2.2. Capteurs embarqués

Les systèmes de télédétection de panaches d'aérosols, depuis l'espace sont constitués des spectromètres embarqués à bord des satellites orbitant la terre. Les spectromètres enregistrent la luminance terrestre à multiples canaux spectraux ou sur tout un spectre. Ces

mesures seront utilisées pour restituer les caractéristiques des aérosols atmosphériques. L'observation des aérosols depuis l'espace n'est pas nouvelle. En effet, le premier instrument spatial pour observer la terre date des années 1970s. En quelques jours de son déploiement, Landsat 1 fournit la première image de grands panaches de feux en Alaska. Les images de Landsat 1 de 80 m de résolution ont été utilisées pour étudier la surface également (cartographie par exemple). De nouvelles générations d'instruments ont remplacé les anciens, la calibration en continu est nécessaire pour éviter la détérioration de la qualité des observations. Ainsi, certains instruments peuvent assurer des mesures fiables au-delà de leur durée de service conçue.

Les études spatiales des aérosols atmosphériques utilisent généralement soit des instruments défilants en orbites quasi-polaire de basse altitude (~750 km), ou ceux de haute altitude en orbite géostationnaire (~36 000 km) (Figure II-4). Les satellites en basse altitude passent au-dessus de chaque région de la terre au moins deux fois tous les 24h. Pour les instruments à visée au nadir, le couloir couvert lors de chaque passage –appelé fauchée– peut avoir une largeur transversale supérieure à 2500 kilomètres. Une quinzaine de passages donc constituent une couverture quasi complète du globe. Certains instruments observent les aérosols au limbe, c'est-à-dire avec un angle de visée tangent au couche atmosphérique (Figure II-5) en mesurant l'irradiance solaire.

Les instruments en haute altitude situant dans des orbites géostationnaires ont une période de révolution autour de la terre égale à une journée, en se synchronisant avec sa période de rotation. Avec une visée fixée vers une zone à l'équateur, on ne voit pas les régions polaires, mais l'avantage est la résolution temporelle qui peut aller jusqu'à 15 min.

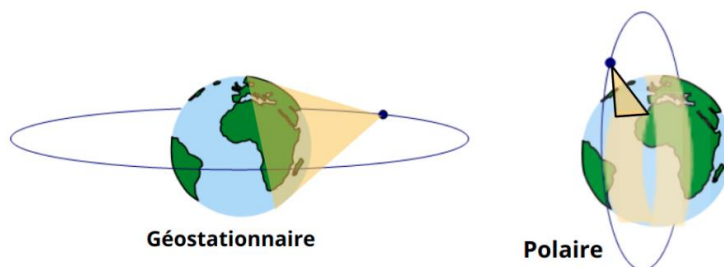


Figure II-4 Illustration d'un instrument en orbite géostationnaire (à gauche) et en orbite polaire (à droite). Adaptée depuis la référence [83].

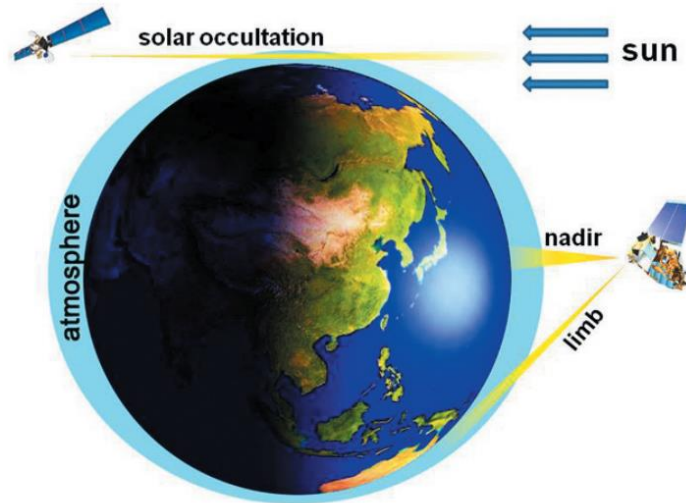


Figure II-5 Différents systèmes de d'observation d'aérosols depuis l'espace [84].

Le choix de longueur d'onde utilisé dans l'instrument de télédétection est important. L'irradiance solaire doit atteindre la couche d'aérosols et conserver une certaine intensité détectable par l'appareillage optique. La présence de molécules dans l'air comme l'eau atténue fortement l'irradiance dans certaines gammes spectrales tels que l'infrarouge thermique, il est donc préférable d'éviter ces domaines spectraux et observer la luminance dans des longueurs d'onde nommées « fenêtres spectrales atmosphériques » où les composants moléculaires de l'atmosphère sont plutôt transparents. Ces bandes spectrales sont visible dans la (Figure II-6) par exemple entre 1500 et 1750 nm.

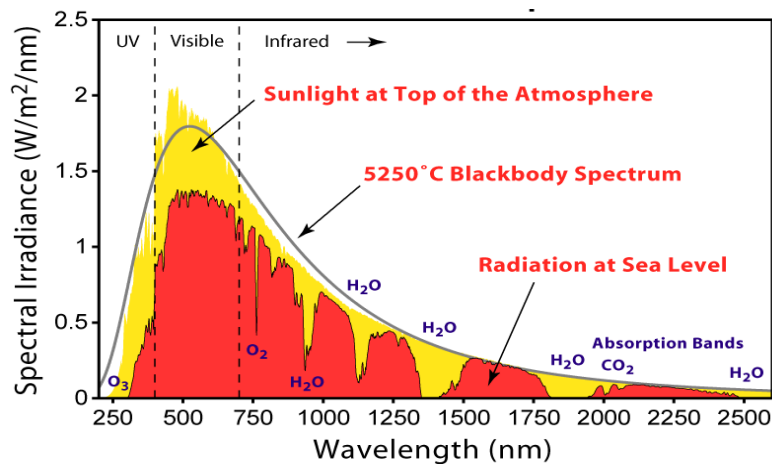


Figure II-6 Spectre de l'irradiance solaire au sommet de l'atmosphère (en jaune) et au niveau du sol (en rouge) montrant les fenêtres spectrales atmosphériques [85].

a- MODIS

Le radiomètre spatial MODIS (*Moderate Resolution Imaging Spectroradiometer*) à bord des satellites Aqua (Figure II-7) et Terra, lancé respectivement en 2002 et 1999 [86] fournit des images multi spectrales de la terre avec un temps de passage à environ 13h30 (Aqua) et 10:30 (Terra) heure locale. L'orbite est héliosynchrone quasi polaire à 705 km d'altitude. MODIS observe la radiance à 36 bandes spectrales (de 405 nm à 14.385 μm) avec une résolution horizontale allant de 250 à 1 000 m et une fauchée au nadir de 2 330 km de large, couvrant quotidiennement la majorité de la surface de la terre [87].

Ci-dessous quelques caractéristiques de l'instrument [88]:

- Dimensions : 1.0 x 1.6 x 1.0 m
- Poids : 228.7 kg
- Puissance moyenne : 162.5 W
- Débit de données moyen : 6.1 Mbps
- Durée de vie : 6 ans

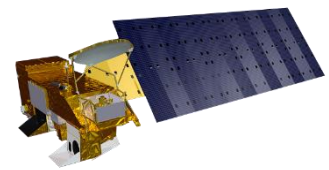


Figure II-7 Image de synthèse de la plateforme Aqua (crédit :NASA).

De nombreux produits de niveau 2 ont été développés à partir des observations de radiance de MODIS, tel que la typologie de couverture de la surface, l'anomalie thermiques, ou encore l'AOD à 550 nm. Ce dernier est montré dans la carte de la Figure II-8 qui permet d'identifier les régions à forte pollution aux aérosols. L'AOD de MODIS est un produit fiable et utilisé par beaucoup d'études comme une valeur de référence. Un instrument plus récent, VIIRS (*Visible Infrared Imaging Radiometer Suite*) fourni des produits basés sur l'héritage de MODIS, il est lancé en 2011 à bord de Suomi NPP (*National Polar-orbiting Partnership*) et déployé à une orbite similaire à Aqua.

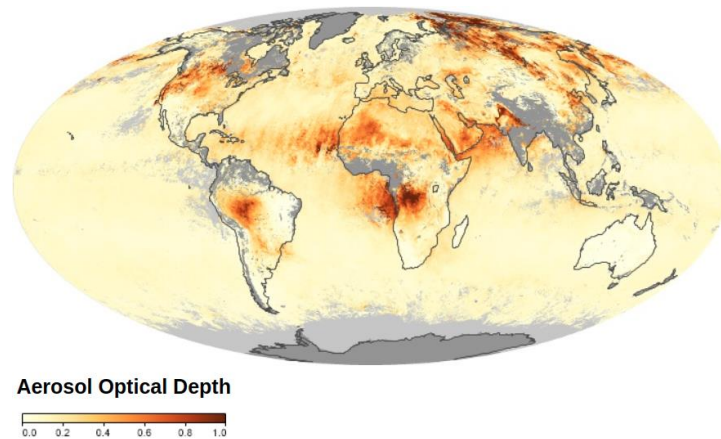


Figure II-8 AOD moyenne pour le mois d'aout 2021 vue par MODIS Terra [89].

b- TROPOMI

Le spectromètre TROPOMI [90] (*TROPospheric Ozone Monitoring Instrument*) est plus récent, et se diffère de MODIS par ses capteurs de haute résolution spectrale. Il est embarqué sur le satellite S5P (*Sentinel-5 Precursor*) (Figure II-9), mis sur une orbite polaire héliosynchrone en 2017 avec un passage ascendant à 13h30 heure locale. Le spectromètre à réseau mesure des spectres de radiance au sommet de l'atmosphère dans 8 bandes spectrales (Tableau II-1), couvrant des longueurs d'ondes de l'ultraviolet (UV) jusqu'à l'infrarouge de courtes longueurs d'onde (SWIR). Régulièrement, il effectue aussi une mesure de l'irradiance solaire afin d'obtenir le spectre de réflectance, qui est le rapport entre la radiance terrestre et l'irradiance solaire. La réflectance est moins sensible aux dégradations radiométrique du capteur [91] c'est pour cela que c'est une grandeur couramment utilisée pour la restitution des propriétés atmosphériques. Ce sont ces mesures de réflectance qui sont utilisés dans ce travail de thèse pour la caractérisation des aérosols (Chapitre III).

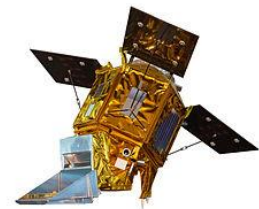


Figure II-9 Image d'un modèle du satellite S5P [Wikipedia].

Tableau II-1 Les 8 bandes spectrale de TROPOMI

Bande	Spectromètre	Résolution spectrale [nm]	Gammes spectrale [nm]
1	UV	0.49	270 - 300
2			300 - 320
3	UVIS	0.54	320 - 405
4			405 - 500
5	NIR	0.38	675 - 725
6			725 - 775
7	SWIR	0.25	2305-2345
8			2345-2385

L'instrument fonctionne dans une configuration en balai-poussoir (sans balayage mécanique) avec une fauchée d'environ 2 600 km à la surface de la terre. La taille de pixel typique (près du nadir) est de 5.5 x 3,5 km² pour toutes les bandes spectrales, à l'exception de la bande UV1 (7 x 28 km²) et des bandes SWIR (7 x 7 km²) [92].

Quelques caractéristiques de l'instrument sont mentionnées ci-dessous :

- 4 spectromètre électroniquement divisé en deux bandes
- Précision radiométrique (absolue) : 1.6% (SWIR) à 1.9% (UV) de la réflectance.
- Masse totale : 204.3 kg sans inclure le module de contrôle (16.7 kg).
- Dimensions : 1.40 x 0.65 x 0.75 m
- Puissance consommée en moyenne : 155 W
- Données générées : 139 Gbits par orbite.
- Durée de vie : 7 ans

Plusieurs produits ont été développés en exploitant les mesures de TROPOMI, certains se basent sur l'héritage des instruments prédécesseurs tel que SCIAMACHY (*SCanning Imaging Absorption spectroMeter for Atmospheric CHartographY*), OMI (*Ozone Monitoring Instrument*) et GOME-2 (*Global Ozone Monitoring Experiment*), pour estimer les profils verticaux d'ozone (O₃) ou les quantités intégrées sur la colonne de dioxyde d'azote (NO₂), dioxyde de soufre (SO₂), méthane (CH₄), formaldéhyde (HCHO), et le monoxyde de carbone (CO). On trouve aussi des produits pour estimer l'abondance de nuages ainsi que l'abondance qualitative (indice d'aérosols) et la hauteur moyenne de la couche prédominante d'aérosols [93].

Grâce à la haute résolution spatiale, nous pouvons observer la forte concentration de NO_2 issue des agglomérations urbaines, ainsi que les émissions récurrentes issues des bateaux dans l'océan atlantique et la mer méditerranéenne [94]. Voir la Figure II-10.

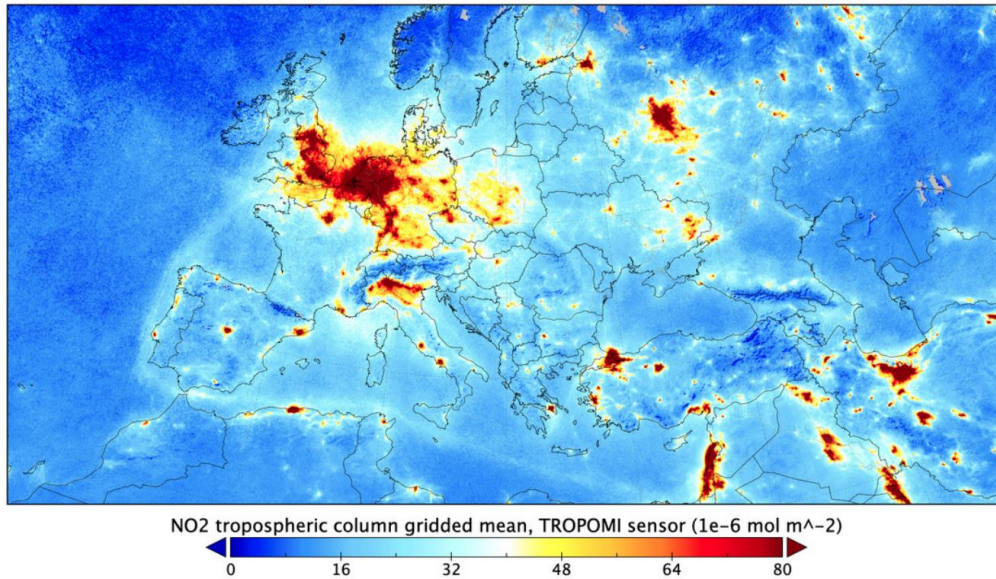


Figure II-10 Carte de la colonne totale moyenne de NO_2 pour le mois d'avril 2018 [95].

4. Observation des aérosols depuis l'espace

Classiquement on observe l'abondance des aérosols depuis l'espace à travers la mesure de l'AOD tel montré dans la Figure II-8. La carte constitue une représentation spatiale intégrée à la verticale, ce qui donne une vue bidimensionnelle (2D) de l'endroit où les aérosols se trouvent dans l'atmosphère. Bien que cette estimation est suffisante pour évaluer l'ampleur de l'impact radiatif des aérosols sur la planète au long termes, ou pour déterminer l'évolution de la tendance de pollution, le manque d'information sur l'altitude des aérosols limite l'étude du transport vertical de ces derniers. La connaissance de la hauteur des aérosols est très importante puisque cette dernière conditionne la gravité de leurs impacts sur l'environnement et sur la santé (voir Chapitre I, section 2.4). Pour estimer l'étendu de la pollution particulaire dans la colonne verticale atmosphérique, nous nous servons des mesures de profils de rétrodiffusion de haute résolution verticale des aérosols mesurés par exemple avec le lidar CALIOP (Figure II-2), mais comme l'angle d'observation est très étroit, on obtient uniquement des transects latitudinaux 2D, donc pas d'observation tridimensionnelles (3D).

Il a fallu développer des techniques sophistiquées d'inversion de spectres électromagnétiques afin d'obtenir des représentations 3D, à l'instar de la méthode AEROIASI

[96] pour observer la forme des panaches de poussière désertique. Cette méthode utilise le spectre de l'irradiance thermique terrestre pour estimer un profil d'aérosols de poussière minérale. Comme la méthode utilise le capteur infrarouge passif polaire IASI (*Infrared Atmospheric Sounding Interferometer*), la couverture globale est assurée, et on arrive à constituer la représentation 3D du panache de poussière. Au-delà des aérosols de poussière désertique qui ont de large diamètre, la méthode AEROIASI ne permet pas d'observer les aérosols fins, car la signature spectrale de ces derniers dans l'infrarouge thermique est négligeable. Dans le chapitre III, je vais donc exploiter une autre bande spectrale qui est le visible et le proche infrarouge. La réflectance terrestre mesurée par l'instrument satellitaire (TROPOMI) sera utilisée pour observer pour la première fois en 3D, les panaches des aérosols issus de la combustion de la biomasse des forêts qui sont plus fins comparés au aérosols de sable désertique.

4.1. Méthode d'inversion de mesures passives de luminance

Comme la plupart des techniques de télédétection de constituants atmosphériques, la restitution de l'AOD est un problème mal posé ; les mesures ne suffisent pas pour déterminer la variable inconnue, car plusieurs situations peuvent produire le même signal. Le but est de trouver la valeur de la variable d'état à partir de la mesure instrumentale. Ce type de problème est souvent résolu par une approche Bayésienne, avec une contrainte sur la solution, ce qui rend cette dernière sensée et stable. Cette approche est souvent employée dans des domaines de l'imagerie (sismique, médicale, etc.).

Ces méthodes d'inversion qui se basent sur l'optimisation de fonction-coût sont lentes, car elles sont gourmandes en termes de calcul. Afin d'assurer des solutions en temps réelle, certaines méthodes d'inversion se base sur le principe de mappage entre des solutions pré-calculées et leurs variables d'état correspondantes. Une des implémentations de ce type d'inversion se fait par des tableaux de recherche multidimensionnelles (*look-up tables*). Cette technique classique a permis d'avoir une couverture globale en carte d'AOD pour le satellite MODIS en temps réel (l'algorithme *Dark Target*). Plus récemment, des méthodes utilisant des modèles à réseaux de neurones artificielles commencent à émerger dans le domaine de l'inversion des observations satellitaires.

Dans ce qui suit, on va décrire la méthode d'inversion par optimisation de fonction-coût. D'abord, un spectre théorique est simulé par ordinateur en utilisant les informations

disponibles sur l'état de l'atmosphère, en supposant des hypothèses de départ sur les variables inconnues. Cette étape est appelée la modélisation directe. Ensuite, le spectre simulé sera ajusté pour se rapprocher de l'observation réelle en modifiant successivement quelques valeurs a priori de départ.

La modélisation simule avec un certain degré de détail l'état du système satellite-atmosphère-surface. Ce dernier peut être plus ou moins complexe pour chacune des composantes simulées.

a- Le satellite

On modélise un satellite hypothétique qui aura le même comportement spectral et la même orientation dans l'espace qu'un satellite de référence. La résolution de l'ETR prendra en compte l'angle zénithal, l'angle azimutal relative entre le soleil et le capteur satellitaire, ainsi que la fonction de réponse spectrale caractéristique de ce dernier afin de simuler les imperfections de sa mesure de réflectance.

b- L'atmosphère

L'atmosphère est modélisée comme des couches uniformément illuminées du haut par le soleil et bornées en bas par un sol réfléchissant. Elle est couramment supposée composée de plans parallèles si l'angle zénithal θ du soleil (voir Figure II-11) n'est pas très grand (inférieur à 80°) [97]. Dans le cas où le soleil est près de l'horizon, il est nécessaire de modéliser la sphéricité de l'atmosphère [97] pour une simulation plus fidèle à la réalité. De plus, l'atmosphère simulée doit considérer les différents composants qui contribuent à l'atténuation du rayonnement (aérosols, gaz, nuages), ainsi que la prise en compte d'autres facteurs météorologiques tel que la température et la pression.

c- La Surface

La surface est modélisée en termes d'émissivité dans l'infrarouge thermique, et en termes de réflectivité dans les courtes longueurs d'ondes tel que l'UV, visible ou infrarouge proche. La réflectivité est connu grâce aux observation satellitaire de différents typologie de la surface (couverture végétale, désert, neige, océans, etc.). Chacune des surfaces est caractérisée par un albédo qui est la fraction réfléchi de la luminance totale incidente au sol. La surface de ce dernier peut être considéré comme Lambertienne, réfléchissant la lumière incidente uniformément dans toutes les directions. Toutefois, pour obtenir une représentation réaliste des

caractéristiques optiques de la surface, il est préférable de combiner la réflexion diffuse et spéculaire (directionnelle).

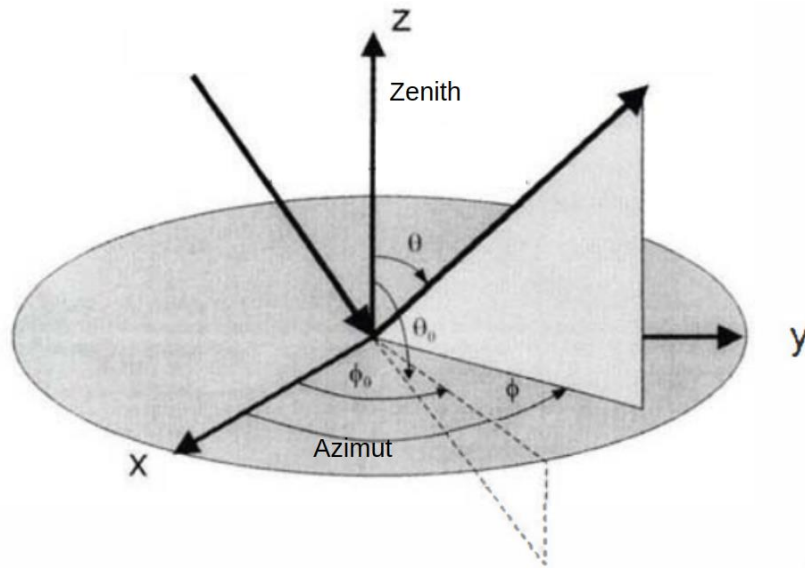


Figure II-11 Illustration de l'angle d'incidence et l'angle de diffusion dans les coordonnées polaires. Adaptée depuis la référence [98].

Le principe scientifique utilisé pour observer les aérosols à distance repose sur la modélisation de l'équation de transfert radiatif (Chapitre I, section 4.2) par une méthode inverse qui sera développée dans les sections suivantes.

4.2. Équation de transfert radiatif

Le calcul de transfert radiatif dans l'atmosphère est l'estimation de l'intensité du rayonnement qui la traverse à la suite des interactions avec la matière par absorption, diffusion et émission. L'équation de transfert radiatif (ETR) décrit le comportement du rayonnement dans l'atmosphère. Pour un rayonnement monochromatique traversant une atmosphère planétaire supposée composée de plans parallèles et en admettant l'équilibre thermodynamique locale, l'ETR est décrite dans [98–101] comme suit :

$$\mu \frac{dI(\tau, \mu, \varphi)}{d\tau} = I(\tau, \mu, \varphi) - S(\tau, \mu, \varphi) \quad (\text{II.1})$$

où I est l'intensité du rayonnement et S la fonction source, μ est égale à $|\cos(\theta)|$, en coordonnées sphériques θ est l'angle zénithal du rayonnement par rapport à la normale du plan de la couche diffusante, et φ l'angle azimut (Figure II-11).

La fonction source S est égale à :

$$S(\tau, \mu, \varphi) = \frac{\varpi(\tau)}{4\pi} \int_{-1}^{+1} \int_0^{2\pi} p(\tau, \mu, \varphi, \mu', \varphi') I(\tau, \mu', \varphi') d\varphi' d\mu' \quad (\text{II.2})$$

$$+ (1 - \varpi)B(T) + Q(\tau, \mu, \varphi)$$

le premier terme quantifie l'émission de la couche d'air dû à la diffusion multiple [102]. Le second terme représente l'émission thermique où B est la loi de Planck, et ϖ l'albédo de diffusion simple (revoir les équations (I.10) et (I.19) respectivement). Le dernier terme Q représente la fonction source non homogène dû à la diffusion simple [102], elle s'écrit :

$$Q(\tau, \mu, \varphi) = \frac{\varpi(\tau)}{4\pi} p(\tau, \mu, -\mu_0, \varphi - \varphi_0) I_0 \exp(-\tau/\mu_0) \quad (\text{II.3})$$

où μ_0 et φ_0 sont définis comme μ et φ respectivement mais pour le rayonnement entrant avant l'atténuation.

4.2.1. Modélisation direct du transfert radiatif

Lors de la résolution de l'ETR, les quantités τ , ϖ , et $p(\tau, \mu, \varphi, \mu', \varphi')$ sont déterminées à l'aide d'un *Solver* numérique. Le *solver* le plus cité pour la modélisation du transfert radiatif dans la science de la géophysique est **DISORT** (*Discrete Ordinates Radiative Transfer*). Cette résolution qui utilise la méthode des ordonnées discrètes est formalisée dans les années 1980s [100,103], mais elle reste largement utilisée dans le monde entier [104]. Elle repose sur la discrétisation de l'ETR en remplaçant l'intégrale des angles par une formule de quadrature de Gauss. Plusieurs autres modèles de transfert radiatif se base sur ce *solver* tel que VLIDORT (utilisé dans ce travail) et RTTOV (décrits ci-après). Le calcul étant couteux et lent, des efforts ont été fait pour accélérer ce calcul [105]. D'autres *solvers* tel que SMART-G (*Speed-Up Monte Carlo Advanced Radiative Transfer using GPU*) [106,107] utilise l'approche stochastique rapide de Monte Carlo.

Il existe une multitude de modèle de transfert radiatif disponible pour l'étude de la composition atmosphérique. Le modèle doit simuler le comportement de la lumière avec une précision acceptable, être suffisamment rapide et stable numériquement [104]. Le choix du code à utiliser dépendra de la précision de calcul souhaitée, la puissance de calcul mise à disposition, le type d'accélérateur matériel envisagé, et la plage spectrale à laquelle le calcul sera réalisé. Ci-dessous, je cite quelques codes de transfert radiatif populaires :

VLIDORT (*Vector Linearized Discrete Ordinate Radiative Transfer*) [108]: Il s'agit du code de transfert radiatif que j'ai utilisé pour la restitution de l'AOD des aérosols de feux de forêts (voir Chapitre III). VLIDORT utilise la méthode des ordonnées discrètes pour résoudre l'ERT pour chaque fréquence (ligne-par-ligne) pour une atmosphère diffusante à multiples couches parallèles ou pseudo-sphériques. Il peut être utilisé pour les restitutions avec la méthode de l'occultation solaire (visé au limbe) comme pour les restitutions des instruments à visé près du nadir. Il supporte le calcul de polarisation du rayonnement également [108].

RTTOV (*Radiative Transfer for TIROS Operational Vertical Sounder*) : initialement développé dans les années 90s par ECMWF (*European Center for Medium-range Weather Forecasts*) pour assimiler les radiances du sondeur vertical radiométrique TIROS (*Television infrared observation satellites*). Il s'agit d'un code rapide de transfert radiatif (de régression) développé spécifiquement pour les applications de télédétection par satellite. Adapté pour simuler 90 différents capteurs satellitaires, du micro-onde jusqu'au visible [109], et il est souvent utilisé pour les prévisions météorologiques et l'assimilation des observations en temps réelle ou encore et la prévision du changement climatique.

MODTRAN (*MODerate resolution atmosphere TRANsmission*) [110]: il s'agit d'un code développé par *Air Force Research Laboratory*, il modélise le transfert radiatif atmosphérique de l'ultraviolet jusqu'à l'infrarouge avec un calcul de ligne-par-ligne. Il supporte la diffusion multiple pour des applications sur l'étude des aérosols, les gaz et les nuages. Il est utilisé pour des applications de recherche ou commerciales pour la prédiction et l'analyse des spectres de rayonnement [111]. Le code est disponible sous forme de progiciel autonome et de service web. La version récente est dotée d'interface de programmation orientée objet et modulaire pour faciliter l'intégration aux autres applications. Par contre, il peut être coûteux en calcul, en particulier pour une résolution spectrale élevée ou de grands domaines spatiaux. Aussi, la nature propriétaire du code implique moins de transparence aux utilisateurs.

LBLRTM (*Line-By-Line Radiative Transfer Model*) [112]. Développé à partir de **FASCODE** (*Fast Atmospheric Signature Code*) [113]. Il calcule la transmittance et la radiance. Sa bonne précision de calcul permet une marge d'erreur de 0.5% ce qui est très faible comparé à celle attribuée au calcul numérique qui est cinq fois supérieur. LBLRTM se positionne comme une norme de référence pour d'autres codes de transfert radiatif rapides comme RTTOV. Le code peut modéliser une large gamme de conditions atmosphériques, y compris

les effets d'équilibre thermodynamique non local [114]. Néanmoins, ceci nécessitera beaucoup de ressources computationnelles.

Le modèle direct désigné F (équation II.4) représente le transfert radiatif dans l'atmosphère et simule un spectre de réflectance terrestre y qui est un vecteur adimensionnel de taille m . Ce spectre est lié avec le vecteur d'état x de taille n qui sera restitué, et des paramètres additionnels b par le modèle F . Les vecteurs b et x prennent la dimensions des variables modélisés. L'ensemble de l'équation (II.4) décrit le système simulé avec une erreur ϵ .

$$y = F(x, b) + \epsilon \quad \text{II.4}$$

Ce modèle direct suppose des hypothèses (Figure II-12) pour simplifier le calcul, comme l'équilibre thermodynamique locale, la symétrie des particules, etc. De plus, il doit considérer la configuration de mesure de l'instrument comme l'angle de visée et la fonction de réponse spectrale. Il utilise également des paramètres additionnels b qui influencent la simulation, mais qui ne sont pas ajustés lors de l'inversion. Ceux-ci sont les données spectroscopiques, les profils de pression et de température, les propriétés de la surface, la présence d'autre espèces atmosphérique, l'effet Ring, etc. (Figure II-12).

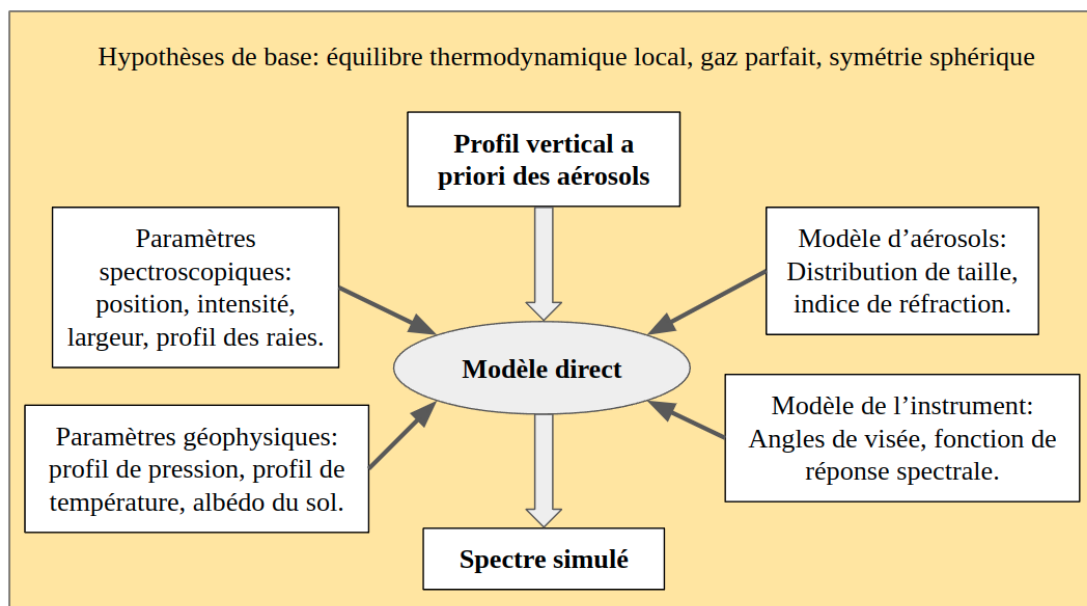


Figure II-12 Schéma de modélisation du transfert radiatif par le modèle direct.

4.2.2. Inversion

Dans le contexte de la télédétection atmosphérique, l'inversion consiste à restituer l'état de l'atmosphère et de la surface qui a donné lieu au spectre de radiance observé par le satellite. Dans ma thèse, j'utilise une méthode appelée « estimation optimale » décrite par C. Rodgers [115]. Cette dernière est basée sur l'approche probabiliste bayésienne qui permet de restituer la meilleure estimation d'un paramètre x à partir d'un a priori, la mesure, et leurs incertitudes.

D'après le théorème de Bayes la fonction de densité de probabilité P d'un événement x en connaissant y est :

$$P(x|y) = \frac{P(y|x)P(x)}{P(y)} \quad (\text{II.5})$$

Sachant qu'une distribution de probabilité gaussienne d'un vecteur x de dimension n , centré autour de la valeur a priori μ et une déviation standard autour de cette dernière représentée par la matrice covariance S_μ est égale à :

$$P(x) = \frac{1}{\sqrt{(2\pi)^n \det(S_\mu)}} \exp\left(-\frac{1}{2}(x - \mu)^T S_\mu^{-1}(x - \mu)\right) \quad (\text{II.6})$$

La résolution du problème inverse est synonyme de trouver l'état x le plus probable (maximum de probabilité) en connaissant la mesure y . Le problème se réduit donc en la minimisation de la fonction-coût L correspondant à la meilleure estimation a posteriori. Pour des distributions gaussiennes de toutes les quantités, L est égale à :

$$L = -2 \ln P(x|y) + c \quad (\text{II.7})$$

$$= [y - F(x)]^T S_\epsilon^{-1}[y - F(x)] + (x - x_a)^T S_a^{-1}(x - x_a) + c \quad (\text{II.8})$$

le premier terme de L représente le résidu entre la mesure et la simulation qu'il faut minimiser, et le second terme peut être considéré comme une contrainte sur la solution x . S_ϵ est la matrice de covariance des erreurs associées à la mesure, et S_a la matrice de covariance a priori du vecteur d'état, qui correspond à la variabilité de l'état autour d'une solution a priori x_a , et c une constante qui n'a pas d'impact sur l'optimisation de L .

Comme F est non linéaire, la résolution de L se fait avec une méthode d'approximations successives. Nous pouvons utiliser la méthode de Gauss–Newton avec ajustement du profil vertical estimé initialement x_i , où il est recommandé de prendre le profil a priori comme profil initial lors des itérations. Pour obtenir cette solution, nous développons

le modèle direct autour de l'estimation initiale sous forme de série de Taylor à l'ordre premier en ignorant les termes supérieurs comme décrite dans [116]

$$y = F(x_i) + \frac{\partial F}{\partial x}(x_{i+1} - x_i) \quad (\text{II.9})$$

$$= y_i + K_i(x_{i+1} - x_i) \quad (\text{II.10})$$

K_i est la fonction de poids ou la matrice Jacobienne de l'itération i , représentant la sensibilité du modèle direct envers le changement au vecteur d'état x . Ce dernier prend la forme suivante pour chaque itération selon la méthode d'estimation optimale [115] en prenant en compte un état a priori:

$$x_{i+1} = x_i + (S_a^{-1} + K_i^T S_\epsilon^{-1} K_i)^{-1} [K_i^T S_\epsilon^{-1} (y - F(x_i)) - S_a^{-1} (x_i - x_a)] \quad (\text{II.11})$$

L'algorithme s'arrête quand on aboutit à une valeur de x_i au-delà de laquelle L devient quasiment constant, le spectre simulé sera similaire au spectre mesuré et on dit que l'algorithme de minimisation a convergé. Le critère de convergence est déterminé en trouvant un compromis entre la rapidité de la restitution et de la précision souhaitée, soit en fixant un nombre d'itération à faire ou en fixant un dL acceptable au-delà duquel la minimisation s'arrête. Le vecteur d'état obtenu correspondra à une estimation optimale par rapport aux contraintes établies dans la fonction minimisée.

4.2.3. Évaluation qualitative de la qualité de l'inversion

- La matrice des noyaux moyens

La fonction de lissage caractérise la sensibilité de l'état restitué par rapport à la variation du vecteur d'état réel. On la désigne par la matrice carrée A qui nous renseigne sur la qualité de l'inversion. Les lignes de la matrice sont les noyaux du lissage (*averaging kernels*).

$$A = GK = \frac{\partial \hat{x}}{\partial x} \quad (\text{II.12})$$

où \hat{x} le vecteur d'état restitué, et G est la matrice de contribution ou la matrice de gain, elle est définie comme ceci :

$$G = \frac{\hat{x} - x_a}{y - Kx_a} \quad (\text{II.13})$$

et égale à :

$$G = (K^T S_\epsilon^{-1} K + S_a^{-1})^{-1} K^T S_\epsilon^{-1} \quad (\text{II.14})$$

- Le degré de liberté

Le degré de liberté de la restitution DOF (*degree of freedom*) est la somme des éléments diagonaux a_{ii} de la matrice A et il prend une valeur maximale égale à la taille du vecteur d'état. Il indique le nombre d'éléments indépendants qui sont restitués lors de l'inversion.

$$DOF = \sum_{i=1}^n a_{ii} \quad (\text{II.15})$$

4.2.4. Analyse des erreurs de l'inversion

La matrice de covariance de l'erreur totale S_{tot} associé à la restitution est la somme des matrices de covariance S_s , S_m , et S_f constituant respectivement les contributions de trois sources d'erreurs : des erreurs dues au lissage en lien avec les poids de la matrice A , des erreurs de mesures et des erreurs dues à l'incertitude des paramètres du modèle direct :

$$S_{tot} = S_s + S_m + S_f \quad (\text{II.16})$$

- Erreur de lissage (*smoothing*) S_s

S_s est la matrice de covariance des erreurs de lissage causées par la non indépendance des éléments du vecteur d'état restitué, sa formule est la suivante :

$$S_s = (A - I) S_e (A - 1)^T \quad (\text{II.17})$$

où S_e est la matrice de covariance d'un ensemble de vecteur d'état réel par rapport à la moyenne.

- Erreur des mesures S_m

Le bruit de la mesure influence la restitution, les erreurs induites par ce dernier sont représentées d'après Rodgers [115] par cette équation :

$$S_m = G S_e G^T \quad (\text{II.18})$$

- Erreur liée aux paramètres du modèle direct S_f

Pour l'inversion, on utilise une estimation des paramètres du modèle direct non restitués \hat{b} . On prend en compte les erreurs associées à ces paramètres, qui peuvent être systématiques ou aléatoires, en calculant la matrice suivante :

$$S_f = A_b S_b A_b^T \quad (\text{II.19})$$

tel que A_b et S_b sont respectivement la matrice des noyaux moyens et la matrice de covariance des paramètres b .

4.2.5. Régularisation Tikhonov-Phillips

La méthode de régularisation Tikhonov-Phillips est pratique dans le cas où il est difficile de trouver la matrice de covariance a priori S_a , car cette dernière est construite à partir de plusieurs mesures indépendantes des variables du vecteur d'état. Dans ce cas, une solution est de modifier L en remplaçant S_a^{-1} par une matrice R . Cette dernière contraint uniquement la forme du vecteur d'état de la solution grâce à l'utilisation de l'opérateur dérivée première L_1 pondéré par un coefficient α . Cette méthode est décrite dans [117].

$$R = \alpha L_1^T L_1 \quad (\text{II.20})$$

$$\mathbf{L}_1 = \begin{bmatrix} -1 & 1 & 0 & \dots & 0 \\ 0 & -1 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & -1 & 1 \end{bmatrix} \quad (\text{II.21})$$

Il est également possible d'utiliser l'opérateur de dérivée seconde. Le vecteur d'état restitué avec les contraintes de dérivées première et/ou seconde sera contraint uniquement vis à vis de la forme (degré de lissage du profil) et non pas par rapport à l'intensité (concentration). Cette dernière sera ajustée suivant l'intensité du spectre de mesure.

Le coefficient α est un paramètre qui détermine la force de la contrainte. Le choix de ce dernier est critique car le degré de liberté de la restitution tend vers zéro à des valeurs larges de α ou les erreurs de restitution sont trop élevées pour des α trop faibles.

5. Modélisation régionale de la qualité de l'air

Les modèles numériques traduisent des phénomènes observés liés à la composition et aux processus atmosphériques sous forme de relations mathématiques. La solution numérique de ces équations souvent complexes est une approximation de la grandeur souhaitée pour un usage spécifique. Dans le contexte de l'étude de la qualité de l'air, plusieurs phénomènes modifient l'état de l'atmosphère et la concentration des espèces, un modèle prend en compte le

transport et la réactivité des constituants de l'atmosphère pour chaque instant. Il existe plusieurs types de modèles utilisés pour estimer la qualité de l'air en terme d'aérosols. Certains se basent sur des approches CFD (*computational fluid dynamics*) pour résoudre l'équation de transport des flux, on trouve également la modélisation gaussienne de la dispersion des polluants, en plus des modèles de chimie-transport (CTM pour *chemistry-transport model*).

Un CTM est utilisé pour simuler les panaches à grand échelle (régionale/globale) et pour calculer les champs 3D des concentrations des espèces. Certains CTMs adoptent un référentiel Eulérien où la résolution de la grille de simulation est fixe, d'autre adoptent un référentiel lagrangien où chaque particule simulée est traquée tout au long de la durée de simulation. Dans ma thèse, j'utilise des simulations d'AOD issues du modèle Eulérien CHIMERE (Chapitre IV). Une description du modèle est présentée ci-après.

5.1. Le modèle de chimie-transport CHIMERE

CHIMERE est un modèle de chimie-transport qui simule l'évolution spatio-temporelle d'une centaine d'espèces d'aérosols et de gaz dans l'atmosphère. Il est capable de simuler des phénomènes de l'échelle locale jusqu'à l'échelle hémisphérique [118] avec une résolution horizontale qui va de 1 km jusqu'à quelques degrés [119]. Plusieurs travaux de recherches sur les aérosols et ses précurseurs utilisent le modèle [34,120–123], et il est utilisé en production comme outil de prévision et d'analyse de la qualité de l'air en termes de pollution particulaire et gazeuse [124], mais dans ce travail on va se concentré sur l'aspect modélisation d'analyse (du passé) et non pas sur la prévision (du futur).

Comme CHIMERE est un modèle Eulérien, le domaine de simulation est discrétisé en grille tridimensionnelle fixe, les différents champs de concentration sont calculés dans chaque maille composant la grille. En respectant le principe de conservation de la masse. La variation de la concentration C_i d'une espèce dans chaque maille étant la somme des variation de sa concentration issues des contributions de l'émission, l'advection, le dépôt sec et humide, et de la réactivité chimique (équation ci-dessous) :

$$\frac{\delta C_i}{\delta t} = \left(\frac{\delta C_i}{\delta t}\right)_{emission} + \left(\frac{\delta C_i}{\delta t}\right)_{advection} + \left(\frac{\delta C_i}{\delta t}\right)_{dep\dot{o}t} + \left(\frac{\delta C_i}{\delta t}\right)_{chimie} \quad (II.22)$$

On modélise également d'autres phénomènes qui contribuent à la concentration finale calculée dans la maille comme la nucléation, la coagulation, l'absorption [118]. Pour les

régions désertiques, le soulèvement des particules de poussières minérales désertiques sont également pris en compte [125].

Le modèle CHIMERE est régional, la simulation utilise donc des champs de climatologie pré calculés par un autre modèle de circulation global tel que LMDz-INCA (Laboratoire de Météorologie Dynamique couplé avec le schéma INteraction Chimie et Aérosols) [126] afin d'initier la simulation aux bords du domaine de simulation.

5.2. Émission

Les émissions anthropiques sont des valeurs d'inventaires d'émissions tel que les données EDGARv5.0 qui est une compilation des émissions de gaz à effet de serre, émissions de pollution particulaire primaire (PM_{2.5}, PM₁₀) pour tous les pays et pour la majorité des secteurs activités [127].

L'émission de la poussière désertique est modélisée premièrement par l'effet du sablage, ce phénomène libère les particules de poussière du sable et des agrégats en mouvement à la surface. Ce mouvement est initié par le vent quand la vitesse de friction U_* dépasse une certaine limite U_*^T . Ceci constitue ce qu'on appelle la saltation. Cette dernière est modélisée suivant la relation développée dans [128], Le flux massique $F(D_p)$ est égale à :

$$F(D_p) = \frac{K}{g} \rho (U_* - U_*^T) (U_* + U_*^T)^2 \quad (\text{II.23})$$

où K est une constante de proportionnalité, ρ est la masse volumique de l'air, et D_p le diamètre de la particule. La vitesse de friction limite U_*^T est calculée suivant le schéma décrit dans [129]. Le modèle de flux de sablage utilisé [125] est proportionnel au flux de saltation et au cube du diamètre massique moyen de la particule.

5.3. Advection

Le schéma d'advection utilisé est celui de Van Leer [130] en utilisant les meilleurs champs météorologiques disponibles réanalysés tel que les données d'ECMWF, ou NCEP (National Centers for Environmental Prediction).

5.4. Dépôt sec et humide

D'après l'article [131], le dépôt est modélisé dans CHIMERE 2017 en terme de résistances [132] tel la vitesse totale de dépôt v_d d'une espèce i est égale à l'inverse de la

somme de la résistance aérodynamique R_a , la résistance de la couche limite R_b , et la résistance de générale de la surface R_c .

$$v_{d,i} = \frac{1}{R_a + R_b + R_c} \quad (\text{II.24})$$

Quant au dépôt humide, le taux de lessivage au niveau des nuages est proportionnel à la quantité d'eau perdue par précipitation [133], alors que le taux de lessivage en-dessous des nuages tient compte de l'efficacité de collision ainsi que la dissolution irréversible des particules dans la gouttelette d'eau de pluie.

5.5. Chimie

Les processus chimiques sont représentés dans CHIMERE à l'aide d'un schéma chimique qui désigne l'ensemble des réactions chimiques qui peuvent se produire dans l'atmosphère avec leurs constantes cinétiques. Ces constantes dépendent souvent de la température et parfois de la pression atmosphérique et du rayonnement solaire. Dans un schéma chimique on trouve des espèces chimiques et des espèces représentatives de famille d'espèces.

Le schéma d'interaction chimique à la phase gazeuse adopté par CHIMERE est MELCHIOR2 (Modèle Lagrangien de Chimie de l'Ozone à l'échelle Régionale) [134], pour les aérosols organiques les schémas adoptés sont : le modèle d'aérosol organique hydrophile/hydrophobe H^2O [135] et SOAP [136], ce dernier modélise l'interaction thermodynamique des aérosols organiques secondaires. Quant aux aérosols inorganiques secondaires on utilise ISOROPIA 2.1 [137]. Le développement du module d'aérosols organique et inorganique utilisé dans CHIMERE est présenté dans une publication de Couvidat [131].

5.6. Propriétés optiques

Les propriétés optiques des aérosols utilisées par CHIMERE 2017 sont décrites dans l'article [118], il est également possible d'utiliser d'autres modèles tel que le mécanisme de photolyse Fast-J [138]. Les aérosols sont répartis en plusieurs classes (appelées *bins*) de tailles différentes, et l'AOD est calculée sous une hypothèse de mélange externe.

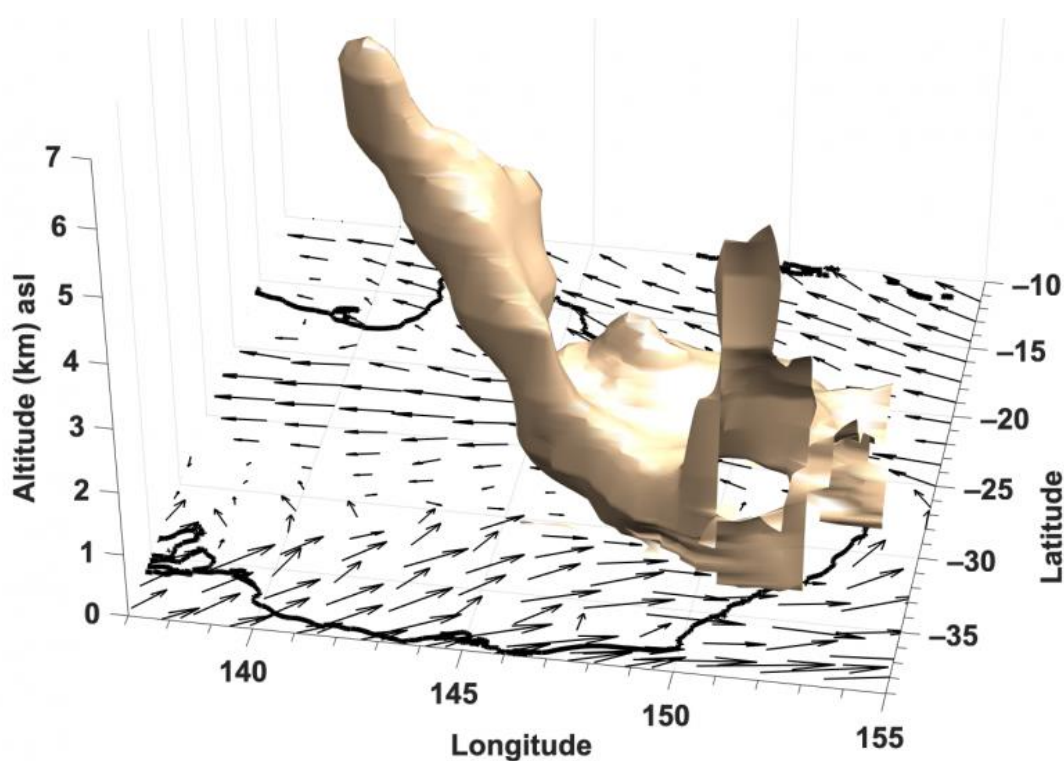
6. Conclusion

Dans ce chapitre, nous avons vu les méthodes d'estimation des aérosols atmosphériques. En particulier les méthodes basées sur la télédétection, et celles basées sur la modélisation numérique. Les avantages et les inconvénients de chacune des méthodes sont présentés. En effets, le point faible de la télédétection se résume dans l'incapacité de restituer passivement une vue 3D des aérosols fins, alors que l'estimation issue de la modélisation manque de précision dans les régions peu denses en stations in situ. Dans les deux chapitres suivants nous allons corriger l'insuffisance de chacune de ces méthodes.

III. Chapitre 3 : Télédétection tridimensionnelle des aérosols fins

Three-Dimensional Distribution of Biomass Burning Aerosols from Australian Wildfires Observed by TROPOMI Satellite Observations

Résumé.....	89
1. Three-dimensional distribution of biomass burning aerosols:.....	91
2. Materials and methods	94
2.1. The AEROS5P method.....	94
2.2. Inputs.....	96
2.3. Forward calculations of reflectance spectra.....	103
2.4. Inversion procedure	105
2.5. Other aerosol observations used for comparisons with AEROS5P	107
3. Comparison of AEROS5P retrievals with other aerosol products.....	109
3.1. Aerosol optical depth	109
3.2. Aerosol extinction profiles.....	114
3.3. Daily evolution of the 3D distribution of biomass burning aerosol plumes	115
4. Conclusion	120



Résumé

Une meilleure connaissance de la distribution tridimensionnelle des aérosols dans l'atmosphère nous permet de mieux comprendre leurs cycles de vie et leurs impacts environnementaux. Les sources d'aérosols sont principalement à la surface, et la présence de ces particules dégradent la qualité de l'air que nous respirons. Mais ces aérosols sont souvent transportés dans différentes couches atmosphériques, ils atténuent le rayonnement incident sur terre, modifient les propriétés des nuages, ainsi que les propriétés chimiques et dynamiques des couches d'air. L'altitude d'injection dans l'atmosphère des aérosols est en particulier une information importante pour l'estimation de leur distribution spatiale, étant donné les incertitudes des processus atmosphériques modélisés.

Comme présenté dans le Chapitre II, seule la distribution 3D des aérosols grossiers tels que les poussières désertiques ont été observée à l'échelle journalière passivement depuis l'espace. Cette observation a été obtenue en analysant des mesures du spectre de luminance terrestre dans l'infrarouge thermique, où la signature spectrale des aérosols dépend de leur température et donc de leur altitude. Cependant, seuls les aérosols grossiers ont une signature radiative suffisamment significative pour restituer leur altitude dans l'infrarouge thermique (longueurs d'onde supérieures à 8 μm).

Afin de contribuer à l'étude de la distribution 3D des aérosols, j'ai développé une nouvelle approche de télédétection passive par satellite pour l'observer. Il s'agit de la méthode AEROS5P qui observe la distribution tridimensionnelle des aérosols fins, qui n'avait jamais été observée par satellite avant mon travail. Cette méthode adaptée aux observations dans le spectre visible et le proche infrarouge restitue les profils verticaux d'extinction des aérosols à partir des observations en ciel clair (sans nuages) du capteur satellitaire hyper spectral TROPOMI de la mission *Sentinel 5 Precursor*. La méthode est basée sur la théorie de l'estimation optimale introduite par C. Rodgers [115] et utilise une régularisation de type Tikhonov-Phillips, qui détermine le profil vertical d'aérosols en ajustant de manière itérative un spectre obtenu par calcul de réflectance pour des longueurs d'ondes spécifiques (dont la bande-A d'oxygène). On utilise le modèle de transfert radiatif VLIDORT pour effectuer ces calculs. L'ajustement est fait simultanément pour la concentration des aérosols aux différentes couches de l'atmosphère et l'albédo de la surface. Le but est de réduire l'écart entre le spectre calculé et le spectre observé par le satellite.

Ce chapitre présente les premiers résultats de la méthode AEROS5P permettant l'observation de l'évolution journalière de la distribution tridimensionnelle des aérosols fins issus des incendies de forêts australiennes. Le premier cas d'étude détaillé a été effectué pour la période du 20 au 24 décembre 2019, étant donné que c'est une période avec une intensité des feux importante avec une couverture nuageuse particulièrement faible. Les AODs dérivées par intégration verticale de l'extinction des aérosols restituées par la nouvelle méthode sont comparées aux observations coïncidentes de VIIRS. Elles montrent un bon accord dans la distribution horizontale des aérosols, avec un coefficient de corrélation de 0,87 et une erreur absolue moyenne de 0,2, des comparaisons avec les observations de MODIS et AERONET montre également un accord acceptable. Les profils verticaux issus d'AEROS5P sont également comparés avec les mesures du lidar spatial CALIOP. Les altitudes moyennes des aérosols issues de ces deux mesures montrent un bon accord, avec un biais moyen de 185 m seulement et un bon coefficient de corrélation (0,83). De plus, les observations AEROS5P révèlent la hauteur d'injection des aérosols de feux de biomasse en 3D. Les hauteurs d'injection les plus élevées au cours de la période d'analyse coïncidaient avec la plus grande puissance radiative du feu dérivée de MODIS. Un accord est également trouvé avec la stabilité verticale de l'atmosphère.

L'approche AEROS5P fournit des restitutions d'aérosols fins pour les mesures sans contamination de nuages, bien qu'actuellement limitée à des situations avec une présence dominante de particules de fumée. Les développements futurs viseront également à observer d'autres espèces d'aérosols, comme les aérosols de nitrates.

Les sections suivantes reprenant les travaux de la publication scientifique [139] au journal *Remote Sensing* dans sa langue originale (anglais) et rajoutant des éléments additionnels.

1. Three-dimensional distribution of biomass burning aerosols:

Biomass burning (BB) aerosols have a major impact on the environment. Large-scale emissions of these particles have large impacts on the radiation balance and climate. Fresh and aged biomass burning plumes produce diabatic radiative heating, e.g., [140], and then can lead to self-lofting of the plume, e.g., [141]. They can produce a warming or cooling radiative forcing, depending on their optical properties and then their atmospheric lifetime [8]. If vertically and horizontally co-located with clouds, they may modify their properties and precipitation by acting as cloud condensation nuclei, e.g., [142], depending on the age and the composition of the particles, e.g., [143]. When transported near the surface, BB aerosols degrade air quality and cause negative long-term health effects, e.g., [144]. The characterization of BB aerosols and their environmental impacts is hampered by the large spatiotemporal variability of wildfires and the difficulty to predict their occurrence and intensity.

The altitudes at which BB aerosol plumes are injected into the atmosphere during emission by wildfires is also important for the estimation of their impacts and is very uncertain. Chemistry-transport models often predict them using semi-empirical methods based on satellite measurements of fire radiative power, burned surface areas, fire temperature, etc.; e.g., [145,146]. The limitations of these methods are shown by comparisons with observations of the fire plumes [147,148]. Their uncertainties are linked to the lack of sufficient knowledge on the characteristics of the fires, pyro-convection processes, their effects on atmospheric stability and feedback. Accurate knowledge on the smoke injection heights is important for studying the regions that these particles may reach and thus their impacts. This is linked to the fact that horizontal transport strongly varies depending on the altitude of the atmospheric layers.

A description of both the horizontal and vertical distributions of atmospheric constituents within BB plumes is required to study their evolution in time and space and their impacts on the environment. Passive remote sensing allows the observation of the abundance of several atmospheric gaseous species, such as ozone or water vapor, in terms of vertical profiles. This is done by using the dependence of the spectral absorption signatures of these gases with respect to their vertical distribution. On the other hand, classic approaches of aerosol passive remote sensing, using sensors such as MODIS (Moderate Resolution Imaging Spectroradiometer), have provided for several decades' total atmospheric columns of aerosol

optical depth, which only describe their horizontal distributions. This is partly linked to the fact that the spectral signature of aerosols only depends on their altitude for specific parts of the radiation spectrum, such as the oxygen absorption bands in the near infrared, e.g., [149], and the thermal infrared but only for coarse particles such as desert dust, e.g., [96]. Indeed, extinction efficiencies of fine particles much smaller than the wavelengths of thermal radiation are mostly very low and thus their signature in this spectral domain is mainly negligible. This is the case for most fine particles such as soot aerosols emitted by fires, for which thermal infrared extinction coefficients are at least a factor of ~ 20 smaller than in the visible [150].

Satellite high spectral resolution measurements in the near infrared, with strong spectral signatures of both fine and coarse aerosols, are only performed by a few spaceborne sensors, such as Global Ozone Monitoring Experiment-2 (GOME-2) [151] and TROPospheric Ozone Monitoring Instrument (TROPOMI) [90], respectively onboard MetOp-A/B/C since 2007 and Sentinel 5 Precursor (S5P) since 2017. Their analysis is challenging since the interactions between radiation and aerosols simultaneously depend on their abundance, altitude, size distribution [152] and composition [153]. Another difficulty for observing aerosols from space comes from the fact that the composition and thermodynamic state of the atmosphere are constantly evolving, and so are the optical properties of aerosols [154,155], which hampers tracing of the path of aerosol plumes in the atmosphere.

Several approaches allow the observation of the vertical profiles of particulate matter in the atmosphere. Airborne in situ sensors provide direct and accurate measurements of aerosol vertical profiles [156], but their availability is limited in space and time as they are mainly performed during field campaigns. Active remote sensing based on lidar instruments like the CALIOP lidar onboard the Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation (CALIPSO) satellite can retrieve latitudinal transects of aerosol vertical profiles [157] but only for a very narrow field of view (90 m wide footprints). On a given day, these measurements are spaced longitudinally by about ~ 2000 km at mid-latitudes (for either daytime or nighttime overpasses). Recently, passive remote sensing of the reflectance spectrum of the oxygen absorption A-band has allowed the retrieval of mean aerosol heights over dark targets, from GOME-2 [149,157] and TROPOMI [158,159] measurements with horizontal resolutions of respectively $40 \text{ km} \times 80 \text{ km}$ and $3.5 \text{ km} \times 5.5 \text{ km}$. Using nine viewing angles, aerosol mean heights are also derived from the Multi Angle Imaging Spectroradiometer (MISR) [160] on bright and dark surfaces at 4.4 km horizontal resolution [161,162]. The vertical profile of desert

dust coarse particles is retrieved from space using measurements of the thermal infrared spectrum from the Infrared Atmospheric Sounding Interferometer (IASI) instrument [163–165]. The latter observations are horizontally spaced by 25 km (at the nadir), so that they allow the observation of the 3D distribution of coarse particles having a strong radiative signature in the thermal infrared (such as desert dust). This may also be found for fine particles in some specific cases, due to the presence of significant absorption bands in the thermal infrared, as seen for sulfuric acid and ammonium sulfate [166], or for highly abundant aerosols emitted by large-scale outbreaks, e.g., [167]. However, this is not the case for most fine particles such as those originating from fires and most urban pollution, e.g. [150].

In this chapter, we present a new approach called AEROS5P for observing the 3D distribution of BB fine aerosols based on TROPOMI measurements. This method derives vertical profiles of aerosol extinction at 550 nm (linearly interpolated between 496 and 673 nm) for cloud-free pixels of TROPOMI. AEROS5P thus provides the daily observation of the full 3D distribution of fine BB particles, for cloud-free conditions. The determination of the vertical profile of aerosol extinction from passive spectral measurements is based on a Tikhonov–Philips-like auto-adaptive regularization. This technique was originally developed for desert dust aerosols and IASI measurements [96], lately applied also to volcanic sulphate aerosol [168], but it is here adapted for BB particles and TROPOMI data. For simultaneously constraining the aerosol vertical distribution and AOD, the method uses both near infrared and visible spectra from TROPOMI (details in Section 2.2). The vertical profiles of BB aerosols derived from AEROS5P can also be used to calculate AODs and mean aerosol layer heights.

We use AEROS5P for describing the 3D distribution of BB aerosols emitted during the extreme large-scale wildfires that occurred in Australia in late 2019 and the beginning of 2020. Our approach provides an unprecedented 3D description of the smoke injection height and transport pathways. We mainly focus on the period of 20–24 December 2019, which is particularly characterized by cloud-free conditions over most of the Australian continent during several days. The record-breaking Australian fires [169] caused a catastrophic impact on the economy, the population health, and biodiversity [170]. Massive amounts of smoke were injected into the upper troposphere and the stratosphere. The subsequent in-plume radiative heating contributed to the formation of a smoke-charged confined ascending vortex that reached 35 km of altitude [141]. The plumes were transported very long distances and dominated the stratospheric aerosol layer of the whole Southern Hemisphere for several

months, producing the most important impact on the climate system linked to stratospheric aerosol perturbations since the eruption of Pinatubo in 1991 [8].

This chapter provides a description of the AEROS5P algorithm backbone, and all additional datasets used in this work (Section 2.2). Section 3 shows comparisons of AEROS5P retrievals with respect to observations of the AOD from AERONET network stations, its horizontal distribution from MODIS and VIIRS (Visible Infrared Imaging Radiometer Suite) and vertical profiles derived from the CALIOP spaceborne lidar. The following section describes the 3D distribution of smoke plumes and their injection heights for the period of analysis, and the link with respect to the wildfire activity and meteorological conditions. Section 4 draws conclusions and discusses perspectives of this work.

2. Materials and methods

2.1. The AEROS5P method

An overview scheme of the AEROS5P algorithm is presented in Figure III-1. The approach simulates radiative transfer (RT) in the surface–atmosphere system and calculates reflectance spectra in the visible and near infrared (see Section 4.1 of Chapter II of this manuscript for more detail about the simulation of atmospheric radiative transfer). The simulated reflectance spectra are then compared with those measured by TROPOMI. The retrieval algorithm estimates the vertical profile of aerosol content in the atmosphere with a regularized iterative process, as done for the profiles of desert dust by [96,163] and ozone by, e.g., [171,172]. As typically done in the Tikhonov–Philips regularization [117] and the optimal estimation method [115], the cost function that is minimized is the sum of the residuals of the measurement vector and the difference between the state vector and a priori values. These last ones are weighted by respectively radiometric noise and the regularization constraint (as described in [117] and Section 4.2.5 of Chapter II of the manuscript).

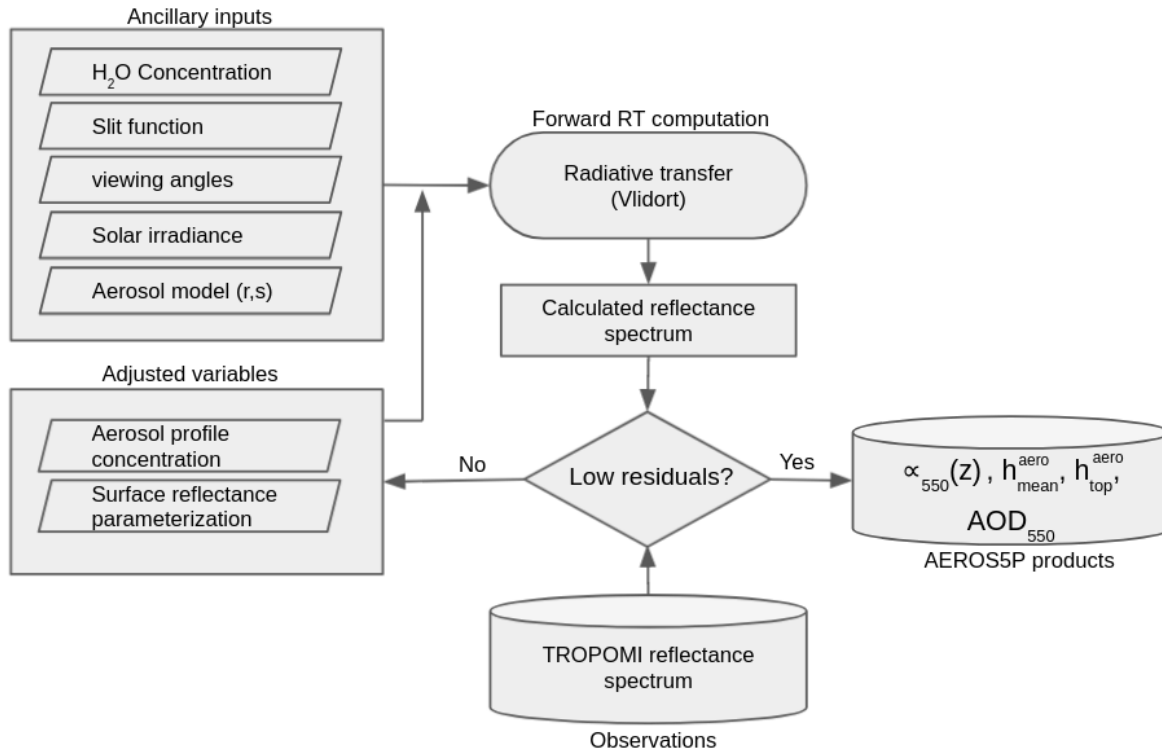


Figure III-1 Scheme of the AEROS5P algorithm for retrieving the 3D distribution of fine BB aerosols. r : aerosol size distribution modal radius; s : aerosol size distribution standard deviation; h_{mean} : aerosol layer mean height; α_{550} : aerosol extinction profile at 550 nm; h_{top} : aerosol layer top height; AOD_{550} : aerosol optical depth at 550 nm.

In absence of clouds, the spectral signature at the oxygen A-band near 760 nm provides information on the altitude of aerosols (this has been shown by e.g., [173] and see an explicative scheme in Figure III-2). This is explained by the fact that the effective depth of the layer of oxygen absorbing radiation depends on the altitude of aerosols. When scattering particles are located at higher (lower) atmospheric layers, solar radiation penetrates less close (closer) to the surface and thus less (more) oxygen molecules absorb radiation. This explains why the oxygen absorption band is stronger when aerosols are closer to the surface and vice versa (Figure III-2a,b). On the other hand, the other TROPOMI measurements in the visible spectrum used by AEROS5P do not depend on the altitude of aerosols and thus are jointly employed for deriving their abundance in terms of AOD (this is the case at 775 nm in Figure III-2b).

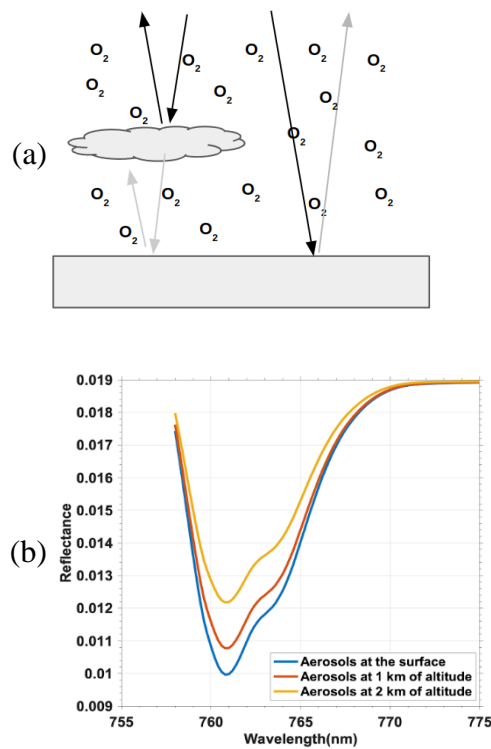


Figure III-2 Scheme describing the aerosol altitude retrieval principle. (a) radiation path length traversing the atmosphere in the presence and absence of aerosols. (b) Oxygen A-band absorption band intensity in different aerosol altitude situations.

The final products of AEROS5P are vertical profiles of aerosol extinction at 550 nm (α_{550}), aerosol top layer height (h_{top}), mean aerosol height (h_{mean}) and AODs for cloud-free TROPOMI pixels both over land and ocean, at a daily scale. The following paragraphs describe more in detail the inputs, RT calculations and inversion procedure of our method.

2.2. Inputs

AEROS5P uses TROPOMI measurements in the visible and near infrared. This instrument is a spectrometer onboard the Copernicus S5P mission. Since its deployment in 2017, the sensor has been tested and calibrated extensively [174], and its quality is periodically monitored [175]. It has a swath width of around 2600 km, offering daily global coverage with a ground pixel resolution of $3.5 \times 5.5 \text{ km}^2$ (after 6 August 2019) and an overpass time of 13:30 local time. The instrument observes the spectrum of radiation emitted by the sun and backscattered by the Earth and the atmosphere towards the sensor, from the ultraviolet to the short-wave infrared (8 bands). The sensor also observes the sun to measure the solar irradiance spectrum.

AEROS5P uses reflectance measurements, derived as the ratio between radiance backscattered by the Earth and solar irradiance at each wavelength. This quantity is less dependent on calibration and aging of the instrument than absolute measurements of backscattered radiance alone. For each pixel, our approach uses reflectance measured at 194 wavelengths (Figure III-3), corresponding to all channels within 12 spectral microwindows (see the central wavelength of these microwindows in Table III-1). Radiation at these wavelengths is mainly affected by aerosols, surface properties and oxygen absorption, while avoiding the spurious influence of other gasses. They sample the visible spectrum from violet to red (specific wavelengths are taken from the list suggested by [176] and the longest wavelength microwindow covers the whole oxygen absorption A-band (from 758 to 775 nm)). These measurements are taken from the instrument bands 4, 5 and 6 and were downloaded from [177]. Two pixels on the west side of the swath of bands 5 and 6 are ignored, since the band 4 has two more ground pixel count compared to the other bands. While not optimal, this temporary workaround seems to work fine for our case study and it did not influence significantly the quality of the retrieval, but a proper treatment of the two bands by co-registration is in consideration. The present chapter only shows results for 1 out of 13 pixels of TROPOMI for reducing computation time, although AEROS5P can process all cloud-free pixels. The effective horizontal resolution is approximately $0.12 \times 0.2^\circ$ (near $12 \times 20 \text{ km}^2$). Only TROPOMI cloud-free pixels are processed. Cloudy pixels were screened out using the VIIRS level 2 cloud mask data CLDMSK_L2 [178] downloaded from [179] and co-localized with TROPOMI pixels. Pixels flagged as “cloudy” and “probably cloudy” were removed for robust cloud filtering.

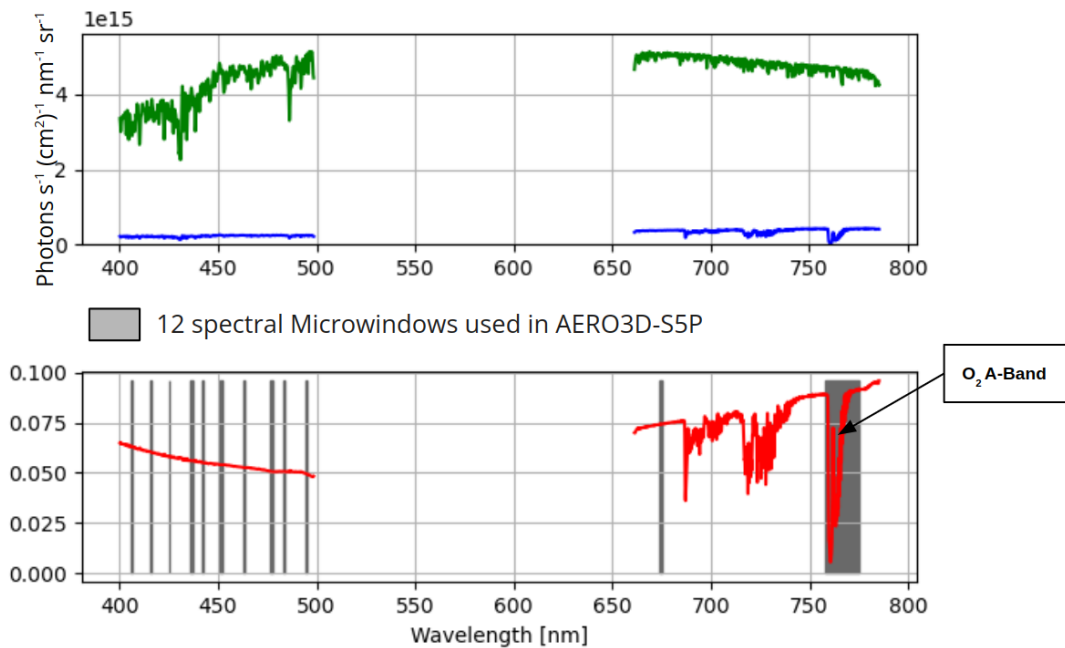


Figure III-3 Irradiance (bleu) and radiance (green) spectrum measurement by TROPOMI and their ratio corresponding to reflectance (red) spectrum with microwindows highlighted in grey that are used in AEROS5P algorithm.

In addition, we applied a so-called “soft correction” for improving the calibration of TROPOMI measurements. This correction is an adaptation from the work of [180,181] developed for the OCRA cloud retrieval for GOME-2 and TROPOMI measurements. It reduces the effects of parasite light sources, and it depends on the across-track position, the month of the year, land/ocean surface and instrument aging. These corrections improve the consistency of AOD and altitude estimation of successive scans.

Ancillary inputs of the algorithm describe the state of the atmosphere and geophysical properties. Meteorological conditions are taken from ERA-Interim reanalysis from the European Center for Medium Weather Forecast [182] in terms of temperature, pressure, and water vapor profiles. A vertically constant oxygen mixing ratio of 0.209 is assumed and oxygen absorption spectroscopic parameters from HITRAN 2012 [183] are used. Other geophysical inputs are a high spectral resolution spectrum of solar irradiance measured by [184] and monthly directionally dependent Lambertian-equivalent reflectivity (DLER) of the surface derived from TROPOMI measurements within the Sentinel-5p+ Innovation project [185] at 19 discrete wavelengths and $0.125 \times 0.125^\circ$ spatial resolution. Figure III-4 shows the diffuse minimum (diffuse) LER of that database at several wavelengths in the Australian region. We

see the reflectivity contrast between land and ocean, which is more apparent at the longer wavelength (754 nm) compared to it at the shorter wavelength (402 nm). In addition to LER, the TROPOMI reflectivity database is modeled with a three order polynomial as described in the reference [186] and shown in the equation below, where c_i are coefficients and θ_v is the satellite viewing angle. The relative azimuth angles between TROPOMI and the sun are also considered to get the directional reflectance.

$$DLER = LER + c_0 + c_1\theta_v + c_2\theta_v^2 + c_3\theta_v^3 \quad (\text{III.1})$$

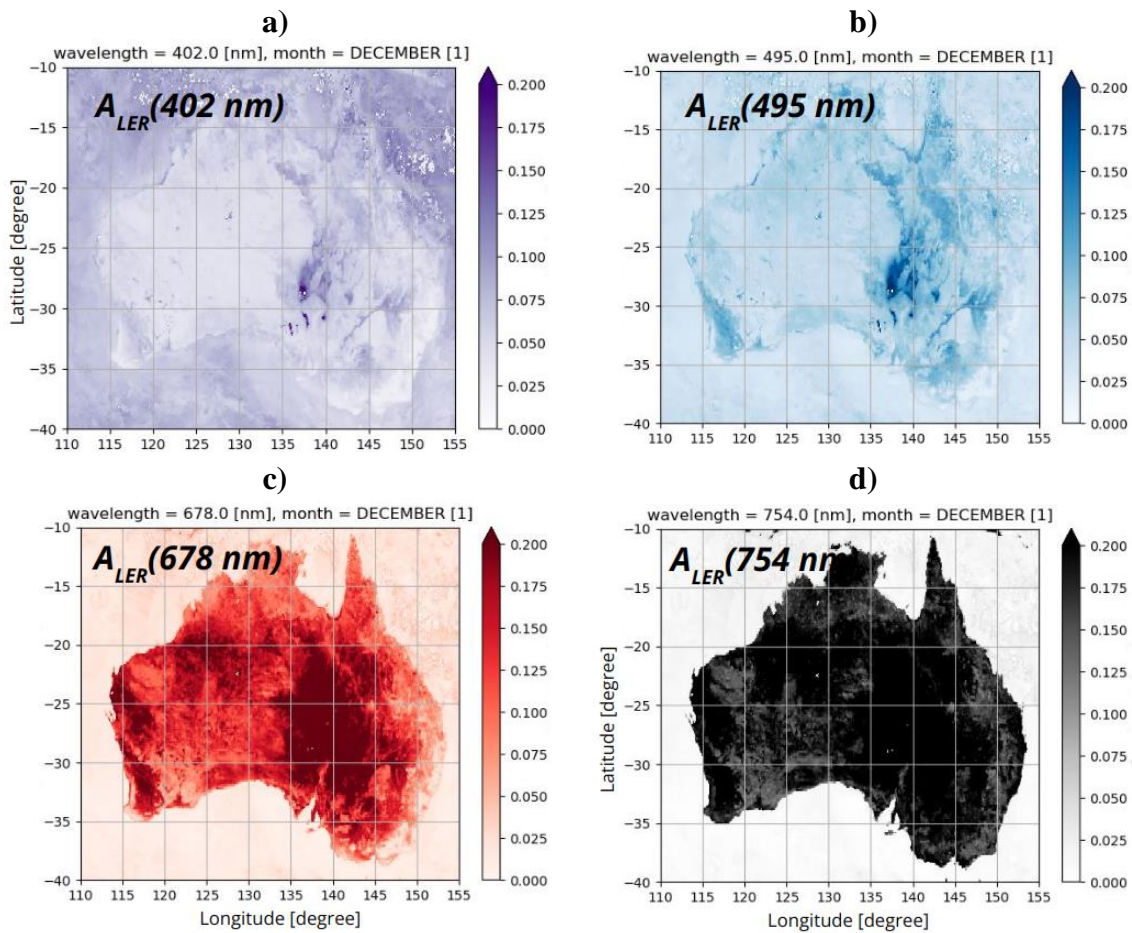


Figure III-4 TROPOMI minimum (diffuse) surface LER reflectivity map of the month of December over Australia shown for 4 discrete wavelengths; a) 402 nm; b) 495 nm; c) 678; and d) 754 nm which are close to that used in AEROS5P.

While several aerosol types may be found in Australia, most of the particulate matter responsible for the high AOD is fine BB aerosols or smoke during the period we analyzed (see Figure III-5). Thus, we assume the microphysical properties of the latter in the retrieval. BB aerosols are assumed to have a monomodal lognormal size distribution with modal radius of $0.1 \mu\text{m}$ and a width of 1.43 taken from the AERONET station measurement at Birdsville (see Figure III-6) and refractive indexes from [187], All retrievals consider a unique a priori aerosol vertical profile of background concentrations, homogenous up to 9 km of altitude and decreasing above. A priori total column aerosol concentration was set to consider a background AOD of 0.03 at 550 nm.

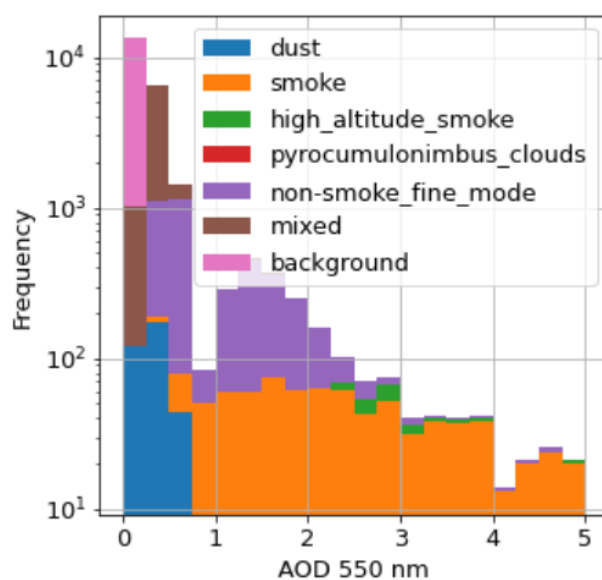


Figure III-5 VIIRS AOD_{550 nm} per aerosol type from 20 to 24 December 2019.

Birdsville , S 25.899, E 139.346, Alt 46 m,
 PI : Ian_Lau, ian.lau@csiro.au
 Size Distribution Alnucantar Level 1.5; 22 DEC 2019

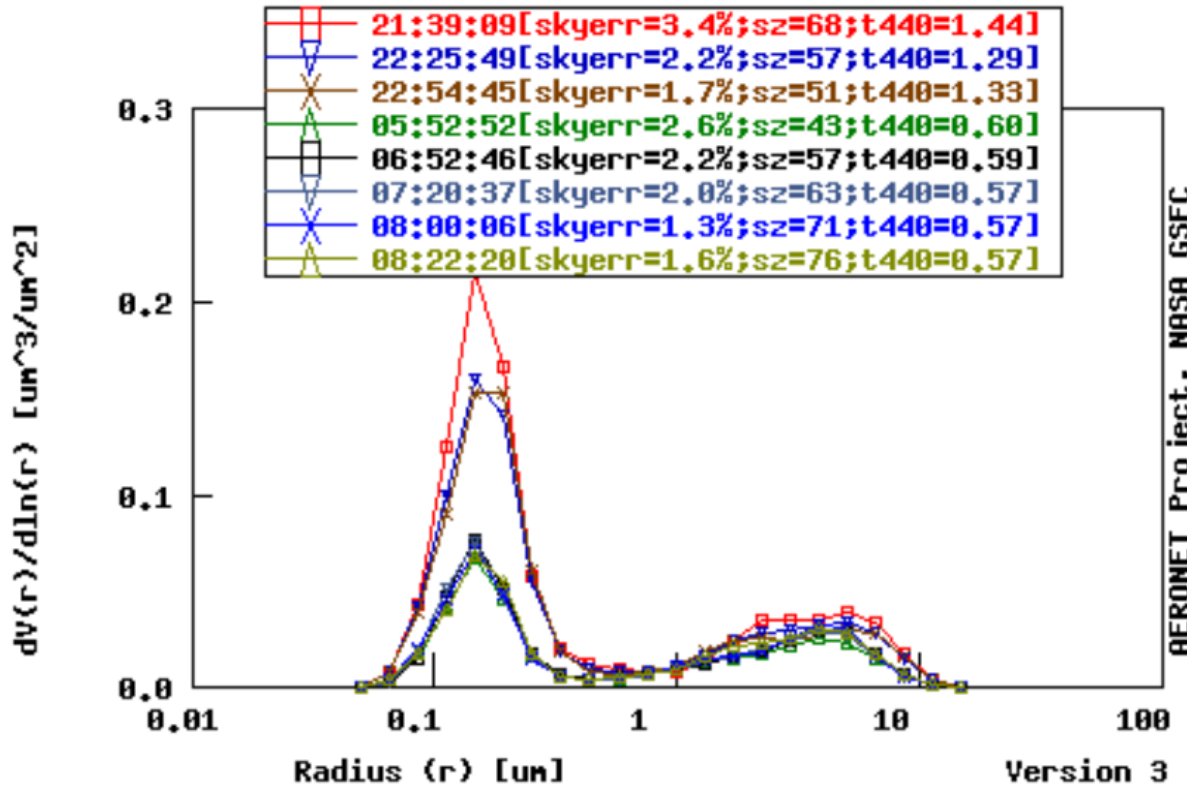


Figure III-6 Distribution de taille d'aérosols observée par la station Birdsville en Australie du réseau AERONET pour la journée 29 décembre 2019 [80].

Table III-1. AEROS5P inputs and their corresponding sources.

	Variable	Characteristics
TROPOMI measurements	Reflectance spectra	12 microwindows centered at 406, 416, 425, 436, 442, 451.5, 463, 477, 483, 494.5, 674.5 and 766.5 nm
Meteorological state	Temperature profile	ERA-Interim reanalyzes
	Pressure profile	ERA-Interim reanalyzes
	Water vapor profile	ERA-Interim reanalyzes
Atmospheric species	A unique a priori aerosol concentration profile	Homogeneous concentration up to 9 km corresponding to a background AOD of 0.03 at 550 nm
	BB aerosol refractive indices	From [187]
	Aerosol size distribution	A modal radius of 0.1 μm and width of 1.43 taken from AERONET retrievals at Birdsville on 22 December 2019
	Oxygen vertical profile	A constant mixing ratio of 0.209
Gas absorption cross sections	Oxygen absorption cross sections	Calculation using HITRAN 2012 [183] spectroscopic parameters, line mixing, H ₂ O broadening and collision-induced absorption following [188]
	H ₂ O absorption cross sections	From HITRAN 2012 [183]
Geophysical state	Solar spectrum	High spectral measurements from [184]
	Surface albedo	DLER at 19 wavelengths, 0.125x0.125°, version 0.3 from the TROPOMI s5p+ Innovation project [185]
Instrumental calibration	Soft correction	Adapted from the soft correction for the OCRA cloud retrieval [180,181]
	Instrumental spectral response	Slit function provided for each swath position and wavelengths from [189]

2.3. Forward calculations of reflectance spectra

Radiative transfer calculations are performed with the VLIDORT (Vectorized Linear Discrete Ordinate Radiative Transfer) model version 2.7 [108,190]. It performs plane-parallel scalar computations accounting for light multiple scattering and absorption in the atmosphere at each wavelength. VLIDORT configuration is chosen for providing acceptable accuracy while reducing computation time. It is set for using a linearized Mie code (assuming spherical particle shape), six aerosol moments, two discrete radiative streams and no vector calculations. Computations are performed line by line for 262 wavelengths spaced by 0.2 nm covering the 12 spectral microwindows used in the approach.

Oxygen absorption cross sections at the A-band are derived for each RT calculation considering line mixing, water vapor broadening and collision-induced absorption following [188]. They are calculated at a high spectral resolution of 0.01 nm and then convolved by the instrument response function for deriving effective cross sections at the RT coarser spectral resolution. This technique significantly speeds up computations while keeping sufficient accuracy, as done for ozone retrievals by [171,172]. Sensitivity tests based on the retrieval of one day of measurements over Australia (20 December 2019) suggest that retrievals of the mean altitude of the aerosol layers using effective oxygen cross sections are highly correlated (correlation coefficient R of 0.96) with those based on high spectral resolution radiative transfer calculations. A mean bias of -270 m (lower altitudes with effective cross sections) and a root-mean-squared (RMS) difference of 560 m was also seen between the two datasets. This indicates that this source of error is not negligible, but we consider it acceptable for analyzing the 3D distribution of BB aerosol plumes in the present study.

We assume that most of the smoke aerosols were located within the troposphere during the analysis period from 20 to 24 December 2019. During this Australian outbreak, transport of smoke aerosols up to the stratosphere was only seen after 29 December 2019 using all available measurements from the CALIOP spaceborne lidar [141,191]. Therefore, in the present chapter we only consider the atmosphere from the surface up to 12 km of altitude, divided into layers of equal depth of 1 km. Each layer contains water vapor, oxygen, and an initial background concentration of aerosols. The choice of the altitude limit comes from a computational cost perspective, yet it is possible to expand the atmosphere model to the stratosphere to study other extreme cases.

Surface albedo ω_0 is spectrally resolved (see an example in Figure III-7b) and directionally dependent. For each wavelength λ of the RT calculations, it is derived as a linear combination of high spectral resolution empirical orthogonal functions (EOF) according to the following equation:

$$\omega_0(\lambda) = \sum_{i=1}^n X_i \cdot EOF_i(\lambda) \quad (\text{III.2})$$

the EOFs are derived from 400 to 800 nm at 1 nm resolution from a principal component analysis of a large database of high spectral resolution measurements of land reflectance for several surface types (vegetation, soils, man-made and ocean from [192]) and the United States Geological Survey (USGS) digital spectral library (version splib06a, [193]). We consider six and three spectrally resolved EOFs respectively over land and ocean. For an example pixel over land, the terms $X_i \cdot EOF_i$ are shown in Figure III-7c. The a priori values of the coefficients X_i are calculated using the TROPOMI-derived DLER database mentioned in Table III-1 for a given location and pointing geometry. The latter database is provided for specific wavelengths (see an example in Figure III-7a). As done by [192], the coefficients are obtained by solving the linear system above evaluated at the wavelengths of the DLER data. This is performed with the following operation:

$$X^{a \text{ priori}} = (A^T A)^{-1} A^T \Omega_{DLER} \quad (\text{III.3})$$

where $X^{a \text{ priori}}$ is a vector with the a priori values of the coefficients X_i , A is a matrix where each column is a EOF interpolated at the wavelengths of the DLER database, the superscript T denotes the transpose representation of a matrix, and the vector Ω_{DLER} is the DLER values from the TROPOMI-derived database. The values of the coefficients X_i are also jointly adjusted with the vertical profile of aerosols for better fitting TROPOMI reflectance spectra during the AEROS5P inversion process. Tests show that the TROPOMI surface reflectivity database seems to give a better reflectance fit compared with the MODIS or OMI (Ozone Monitoring Instrument) surface albedo database.

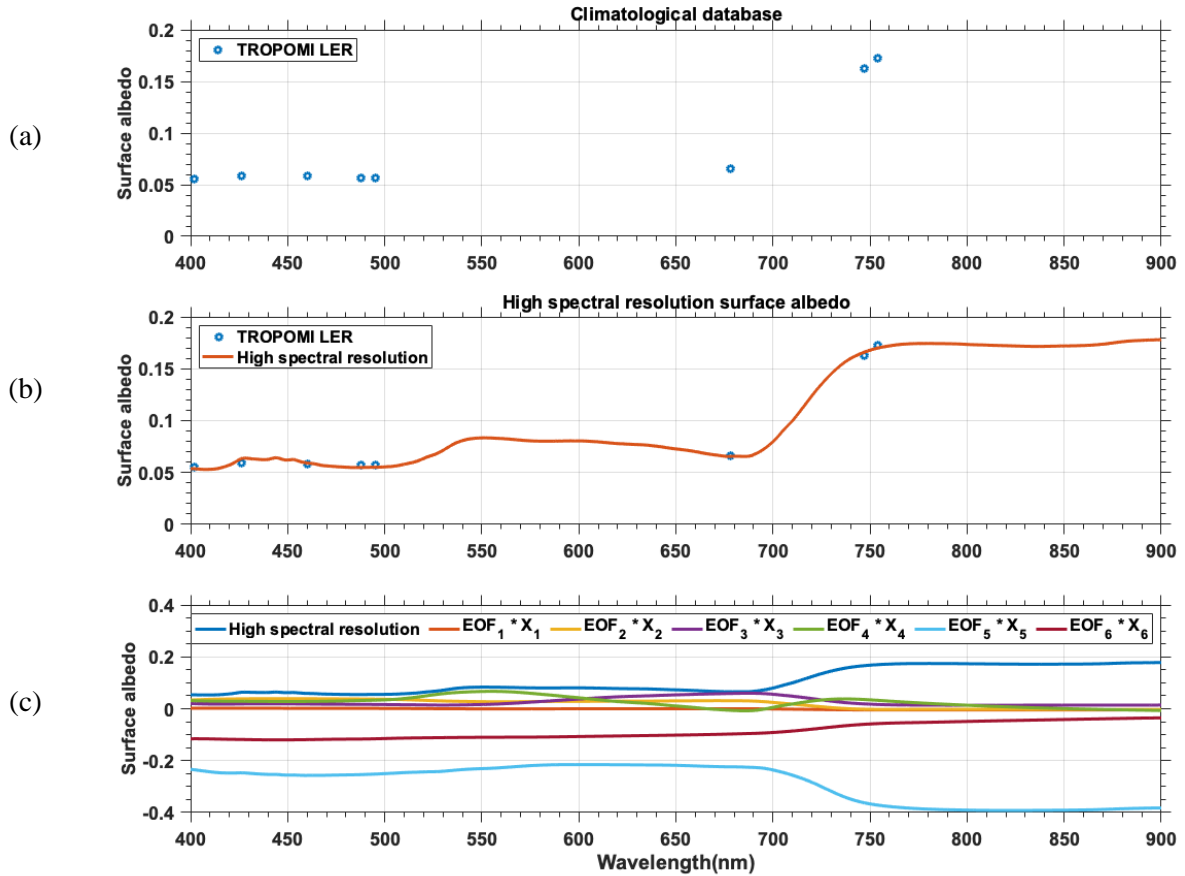


Figure III-7 (a) A random example of a pixel from TROPOMI showing climatological surface LER data that are provided discontinuously at 8 wavelengths. (b): Interpolated surface LER data with high spectral resolution used in AEROS5P (c) High resolution 6 principal components of surface albedo database used for the interpolation, that are developed by [192].

Once the reflectance spectra are simulated by VLIDORT, they are convolved by the instrument spectral response function (or slit function) of TROPOMI for enhancing the resemblance with measurements. These instrumental functions are provided for each swath position and wavelengths from [189].

2.4. Inversion procedure

The variables that are jointly adjusted for reducing the differences between measured and simulated reflectance spectra are given by the state vector $x_{AEROS5P}$:

$$x_{AEROS5P} = [N_{log\ aero}^T X^T]^T \quad (III.4)$$

Thus, the vertical profile of BB aerosol concentrations is obtained while simultaneously adjusting the weighting coefficients c_i in the vector C of the spectrally resolved EOFs of surface albedo. For avoiding negative values, the vector $N_{\log \text{aero}}^T$ contains the logarithm of aerosol number concentrations at each vertical level (spaced by 1 km). Moreover, only the four most significant coefficients c_i defining the surface albedo are adjusted for limiting the size of the state vector.

As TROPOMI measurements do not provide enough information to derive independent values of aerosol abundance at each vertical level, the approach constrains the solution for describing a single aerosol layer with a variable vertical distribution. This is done with Tikhonov–Philips constraints [194] that vary with altitude and auto-adapt to the convergence conditions as done in [117,195]. The constraint strengths are empirically derived from numerical simulations as done by [164] for desert dust but allowing for BB aerosols more variability at the upper atmospheric levels. During the iterative procedure, the constraints of the aerosol concentrations are automatically scaled by a Levenberg–Marquardt-type damping factor. following [164], see equation (II.20), this method releases (enhances) the constraints when converging to smaller (diverging to higher) differences between measured and simulated spectra. The number of iterations was empirically set to 10 as convergence is usually found and the state vector varies at most by 5% (see an examples of measured and simulated spectra in Figure III-8). Using these constraints, the number of independent pieces of information (or degrees of freedom DOF) in the retrieval of BB aerosols profiles varies during the iterative procedure. For the results shown in Section 3, a typical value of ~ 1.5 is reached during the first iterations, which allows the determination of the shape of the aerosol profile.

Once the iterative procedure is finished, the AEROS5P retrieval results for each pixel are taken from the iteration with the least difference between measurements and simulations. Quality assurance post-processing screens out pixels where the spectral differences remain too large ($>10\%$) or too large DOFs (>3). Some retrievals are eventually not usable especially those who are located near clouds, which express high spectral fit residue.

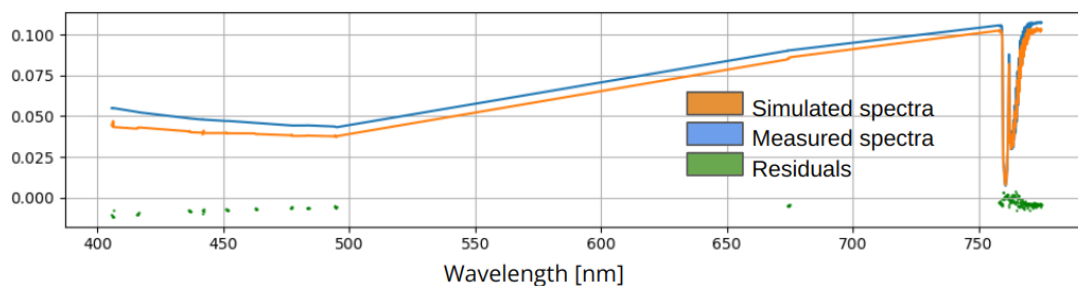


Figure III-8 Reflectance spectral fit of a typical cloud free pixel observation.

The final product of AEROS5P are aerosol extinction profiles at 550 nm, which are calculated with the adjusted concentration profiles and the assumed size distribution and refractive indexes, for cloud-free TROPOMI pixels over land and ocean. We consider that the aerosol profile in terms of extinction is better constrained by AEROS5P than that in terms of concentrations, since this last one directly depends on the accuracy of the assumed particle size distribution and refractive index. AODs are directly calculated by vertical integration of the aerosol extinction profiles, the mean aerosol height as the altitude at which half of the AOD is above and the other below, and the top aerosol height as the highest altitude with an aerosol extinction larger than a given threshold (we use here 0.1 km^{-1}). AODs and aerosol extinction at 550 nm are derived from linear interpolation between retrievals at 496 and 673 nm.

2.5. Other aerosol observations used for comparisons with AEROS5P

For assessing the quality of AEROS5P aerosol retrievals, we compared them with several widely used and validated aerosol products. First, a comparison was conducted with respect to AOD derived from ground-based and satellite measurements. We considered sun photometer AOD measurements from seven AErosol RObotic NETwork (AERONET) stations [79] in Australia with available data in the period of analysis (Birdsville, Fowlers Gap, Lake Argyle, Lake Lefroy, Learmonth, Lucinda and Tumbarumba). We used level 1.5 data downloaded from [73]. They are measurements taken around noon at 440 nm, which we extrapolated to 550 nm using the corresponding Angstrom exponent measurements between 440 and 870 nm.

The horizontal distribution of AOD derived from AEROS5P was compared to that from satellite products derived from the MODIS Dark Target and Deep Blue algorithms [196]. For a better horizontal coverage, we used the Multi-angle Implementation of Atmospheric Correction (MAIAC), which includes data from two MODIS sensors, as well as VIIRS using

Dark Target. We considered the MYD04_L2 product from MODIS onboard the Aqua platform, MCD19A2 combining MODIS/Aqua+Terra (MAIAC), and the AERDB_L2 product from VIIRS onboard the Suomi National Polar-orbiting Partnership (NPP) satellite. We under-sampled these products at the AEROS5P grid resolution. For comparison of the horizontal distribution of AOD, we used MODIS/MAIAC data as they offer an enhanced horizontal coverage compared to the MODIS approaches that use one sensor. Co-localized comparisons of AOD were done with respect to MODIS Aqua, as the overpass time is similar to that of TROPOMI.

Aerosol vertical profiles and mean heights retrieved from AEROS5P were validated against those derived from co-located measurements from CALIOP spaceborne lidar. This instrument measures latitudinal transects of total attenuated backscatter profiles at 532 and 1064 nm, with a 30 m vertical resolution below 8 km and 60 m above. These profiles qualitatively depict the vertical distribution of aerosols. For the cases analyzed here, the operational CALIOP product (v4.2) misclassified the aerosols emitted from fires as “polluted dust” and “marine”. Since the operational level 2 product of aerosol extinction is largely biased, and their spatial coverage is poor, we derived here aerosol extinction profiles using a classic Fernald lidar inversion procedure [197,198] (also called Klett inversion). The particle lidar ratio was adjusted for obtaining a rough match with respect to co-located measurements of AOD from MODIS. We used these aerosol extinction profiles to derive AODs and mean aerosol heights from CALIOP. It is worth noting that the operational TROPOMI product of aerosol layer height is not included in the comparison due to limited spatial coverage. Indeed, almost no data are available in coincidence with the transects of CALIOP and AEROS5P observations analyzed in this chapter.

In addition to aerosol products, we analyzed the results derived from AEROS5P in terms of the 3D distribution of BB aerosols with other datasets describing the fire activity and the atmospheric conditions. The distribution of carbon monoxide (CO) was examined with the TROPOMI operational product [199,200]. Fire activity was surveyed using fire detections and estimates of Fire Radiative Power (FRP) from merged MODIS data onboard Aqua and Terra (MYD14 and MOD14 respectively) products [201] gridded to $0.2 \times 0.2^\circ$ resolution.

Meteorological conditions are described with ERA5 reanalysis [182] from ECMWF at $0.25 \times 0.25^\circ$ resolution and 37 pressure levels (downloaded from the ESPRI IPSL mesocentre <https://mesocentre.ipsl.fr> (accessed on 21 September 2021)). In addition to wind fields, we

describe atmospheric static stability in terms of Brunt–Vaisala frequencies. This quantity describes the buoyancy of the air parcels, according to the surrounding meteorological conditions. It is calculated using ERA5 potential temperature, geopotential, and wind fields with formulas adapted from [16]. This variable does not consider processes as pyro-convection, whose explicit consideration needs a dedicated dynamical model that is out of the scope of the present study.

3. Comparison of AEROS5P retrievals with other aerosol products

3.1. Aerosol optical depth

The horizontal distribution of the AOD observed by AEROS5P was compared to that from MODIS on 20 and 21 December 2019 (Figure III-9). The two approaches showed very similar structures of AODs for the two days, both in variability and absolute values. On 20 December, a large and dense BB aerosol plume of an AOD larger than ~ 1 was depicted over eastern Australia extending from 150°E to 140°E ; their shape and AOD values were very similar for both satellites. South and west of the plume some intermediate AODs of up to 0.4 were also seen in both cases. On the following day, two BB aerosol plumes were similarly depicted in two AOD maps. A plume over central Australia of AODs of ~ 1 and a second one over eastern Australia of AODs above 1 was clearly seen with a similar shape from both satellite data, although moderately larger AODs were seen for AEROS5P. We note a larger horizontal coverage of MODIS/MAIAC data, provided the use of two sensors and a less conservative detection of clouds as compared to AEROS5P. AOD maps from MODIS/Aqua showed overall similar features (not shown for avoiding redundancy) and agreement with AEROS5P, but with a more limited horizontal coverage than MODIS/MAIAC.

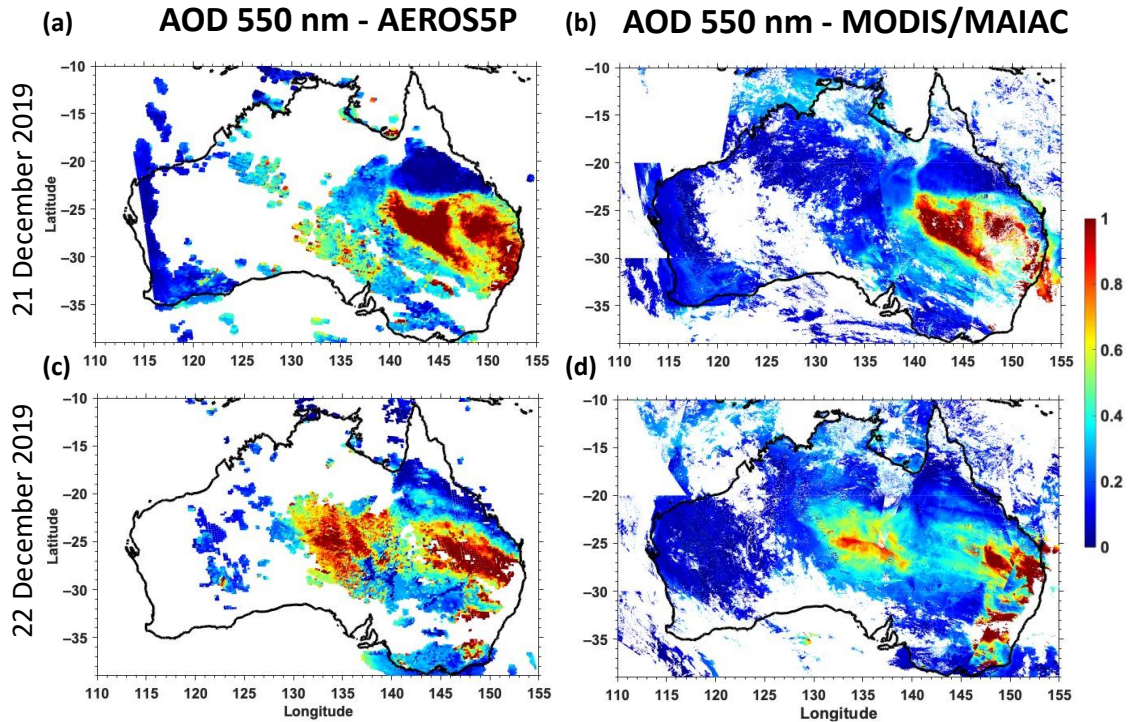


Figure III-9 Horizontal distribution of AOD at 550 nm from (a,c) AEROS5P and (b,d) MODIS using the MAIAC algorithm on (a,b) 21 and (c,d) 22 December 2019 over Australia.

Figure III-10 presents a comparison of the AOD at 550 nm measured at seven AERONET stations with the three co-located satellite products (AEROS5P, MODIS/Aqua and VIIRS). AOD from AEROS5P was in good agreement with VIIRS and MODIS/Aqua on most of the days of the analyzed intense fire episode on 20–24 December 2021. The agreements with respect to AERONET ground-based AOD observations were fairly good and similar for the three satellite products (including AEROS5P). They clearly showed the largest AODs near the fire regions (at Tumbarumba over the southeast of Australia), medium values at central Australia (at Birdsville) downwind from the fires and the lowest ones far from sources (near the eastern coast at Learmonth and Lucinda).

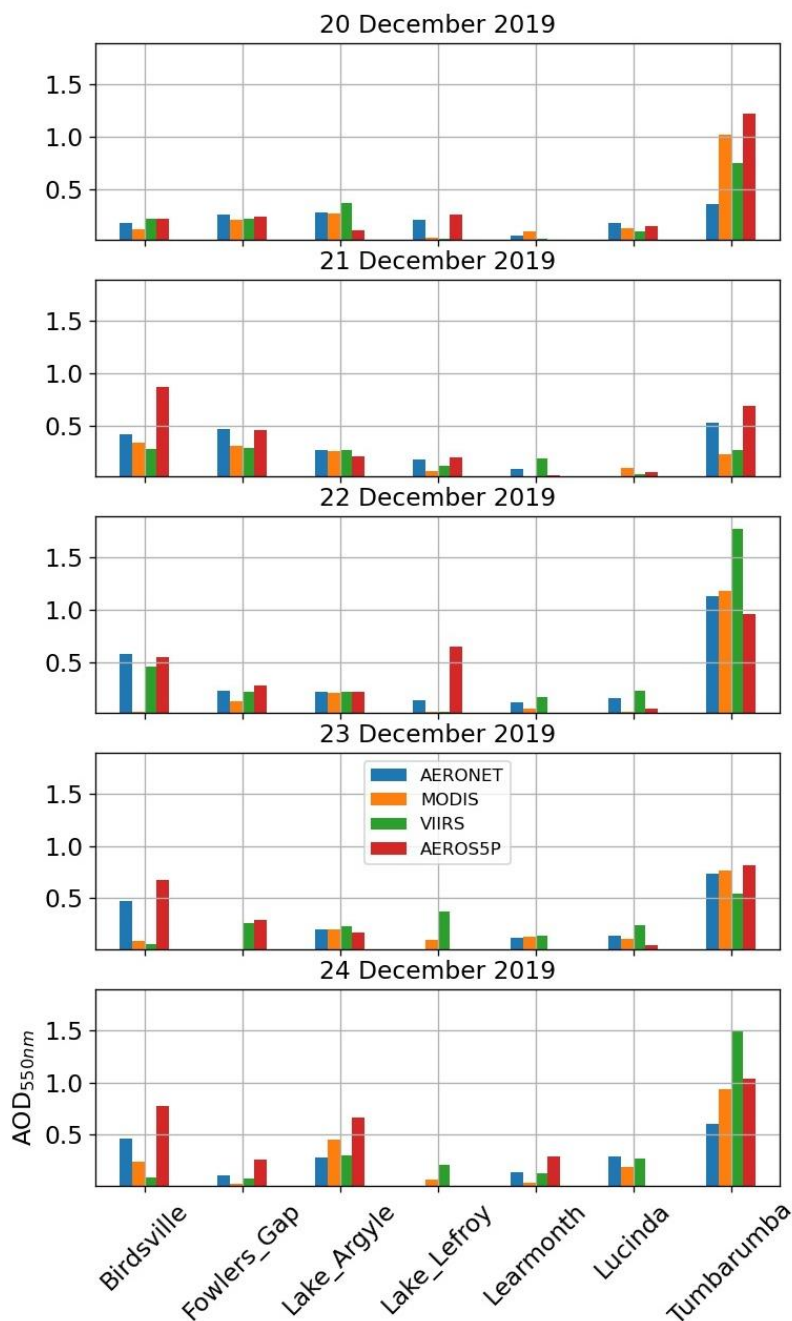


Figure III-10 Comparison of AOD at 550 nm measured at 7 AERONET stations in Australia by sun photometer with respect to three satellite products: AEROS5P, MODIS/Aqua and VIIRS on 20 to 24 December 2019 (one panel per day from the upper to lower one).

We noticed three cases where the AEROS5P AOD significantly differed from the AERONET measurements. First, on 20 December at the station Tumbarumba, larger AODs were derived by AEROS5P and MODIS (respectively ~ 1.2 and ~ 1 , around 13:30 local time) than those from AERONET (~ 0.4 , taken around noon). This is clearly linked to the time difference between the measurements, since AERONET measurements at this station showed

that AOD rose in the early afternoon of this day up to ~ 1.2 (according to measurements reported in [73]). The second case occurred on 21 December, when AEROS5P depicted an AOD around 0.9 and the sun photometer from the Birdsville station measured ~ 0.6 . This slight discrepancy likely comes from the horizontal heterogeneity of the smoke plume as nearby AEROS5P retrievals showed AODs around 0.6–0.7 (Figure III-11 on the 21 December). The last case of a mismatch between the AOD of AERONET and that of AEROS5P was on 22 December at the station Lake Lefroy. The AEROS5P AOD was then around 0.6 while ~ 0.1 was measured by the ground-based station. This difference was certainly caused by cloud contamination unfiltered by the VIIRS cloud mask used within AEROS5P. This is consistent with the fact that most nearby pixels were screened out and only isolated pixels near Lake Lefroy showed high AODs (Figure III-11). A better cloud filtering mask may improve the retrieval quality in the pixels neighboring cloudy situations.

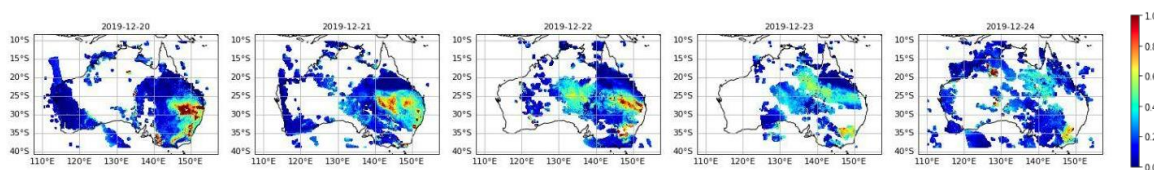


Figure III-11 AEROS5P AOD 550 nm from 20 to 24 December 2019.

A more systematic comparison of all co-located retrievals of AOD from AEROS5P, MODIS/Aqua and VIIRS is shown in Figure III-12. It covers the period of 20–24 December 2019. A good overall agreement was found between the three datasets, whereas the match was better between AEROS5P and VIIRS. This is shown by the fact that 65% of AEROS5P/VIIRS co-located pixels showed differences smaller than 0.2, a correlation coefficient of 0.87 and RMS (Root Mean Squared) differences of 0.3. These values are roughly similar to those found from the comparisons of AOD between MODIS and AEROIASI from the desert dust over the Sahara, with a correlation coefficient of 0.90 and an RMS difference of 0.2 [164]. The large value of our RMS difference is likely linked to the large AOD values of the comparison (up to 5 here, where it is at most 1 for [164]). With respect to MODIS Aqua, these indicators are moderately smaller, as their correlation coefficient and RMS differences were respectively 0.63 and 0.36, while 58% of the collocated pixels showed AOD differences smaller than 0.2. A mean absolute error around 0.2 is seen between AEROS5P and both other satellite instruments. The range of values of comparison for VIIRS is clearly greater (AOD up to 5) than for the comparison with MODIS, explaining partly the larger correlation in the first case. A comparison by the aerosol type detected by VIIRS shows that larger AOD values were mostly identified as “smoke” as expected (Table III-2). Intermediate values with AOD around 0.6 were

detected as “non-smoke fine aerosols”. For those two aerosol types (smoke and non-smoke fine aerosols), the match between AEROS5P and VIIRS was much better (correlation coefficient larger than 0.82). Most of the points of AOD below 0.5 correspond to pixels identified by VIIRS as “mixed” and “background” (Table III-2). For those cases, the AOD derived from AEROS5P that assumes the presence of only BB aerosols showed as expected larger differences (mainly overestimating) with respect to the retrievals from VIIRS and MODIS.

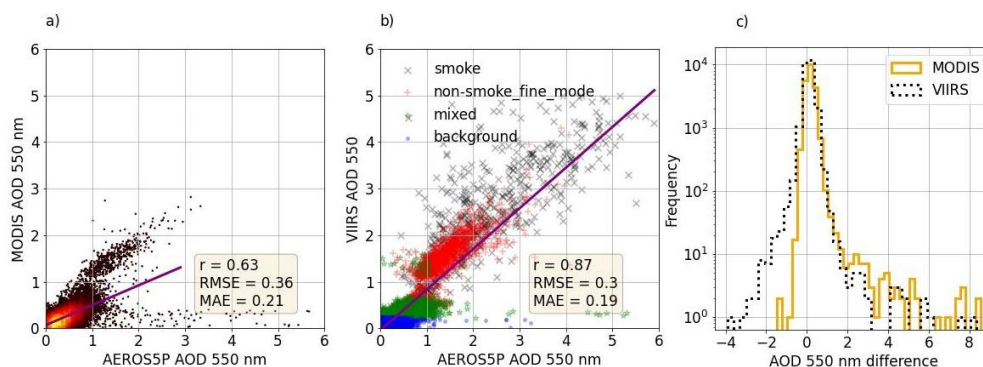


Figure III-12 Scatterplot comparison of AOD at 550 nm derived from AEROS5P with respect to co-located retrievals from (a) MODIS/AQUA and (b) VIIRS for the period 20–24 December 2019. Colors in panel (a) represent the number of points for each sector of 0.1×0.1 of AOD. For (a,b), purple lines are linear regression lines of the scatter of points, dotted lines are the 1:1 curve and blue shade is the $\pm 20\%$ interval. (c) Frequency histogram of differences of AOD at 550 nm of MODIS/AQUA and VIIRS with respect to AEROS5P. Symbols and colors in panel (b) indicate the aerosol type derived from the VIIRS approach.

Table III-2 Comparison of AOD from AEROS5P and VIIRS for each aerosol type detected by VIIRS. r: Pearson correlation coefficient; RMSE: Root Mean Squared Error; MAE: Mean Absolute Error.

Aerosol type	r	RMSE	MAE	Number of co-located pixels
smoke	0.82	0.69	0.49	667
non-smoke_fine_mode	0.86	0.30	0.24	3168
mixed	0.23	0.35	0.21	5868
background	0.24	0.20	0.13	8595

3.2. Aerosol extinction profiles

A comparison of two transects of aerosol extinction profiles derived from CALIOP and AEROS5P on 21 and 22 December 2019 is shown in Figure III-13. The vertical distribution of aerosols by the two datasets is in fairly good agreement. On 21 December, both retrievals showed a denser aerosol layer extending from 1.0 to about 2.5 km in the southern part, approaching the surface at 28–30°S and progressively increasing in altitude up to 1–3 km around 24°S. The transect of the following day showed in both cases a rather homogenous aerosol layer from the surface up to ~4 km. The aerosol layer top gradually varied from 4 km of altitude at the southern part up to 2 km of altitude near 16°S. Quantitatively, values of aerosol extinction profiles were quite similar between the CALIOP and AEROS5P transects. This particularly stood out on 22 December 2019 (Figure III-13b,d), when the vertically integrated AOD for AEROS5P and CALIOP was very similar to that derived from MODIS (Figure III-13f). The transect of 21 December 2019 showed moderately larger values of extinction for AEROS5P than those derived from the inversion of CALIOP measurements, while the AOD from AEROS5P was in better agreement with MODIS than that from CALIOP (it underestimated the AOD south of 28°S in Figure III-13e). Some other differences may have partly come from the presence of clouds and the difference in overpass times of CALIOP and TROPOMI of about 1.5 h. The mean altitude of the aerosol layers is depicted as dashed lines in the figure. There is a good agreement between the two datasets. Considering all coincidences in the two transects, they show a correlation of 0.83, a mean bias of 180 m and an RMS difference of about 500 m. The uncertainty of the mean aerosol layer height derived from AEROS5P is moderately weaker and in similar order of magnitude to that of the TROPOMI operational product (between 500 and 1000 m according to [202]).

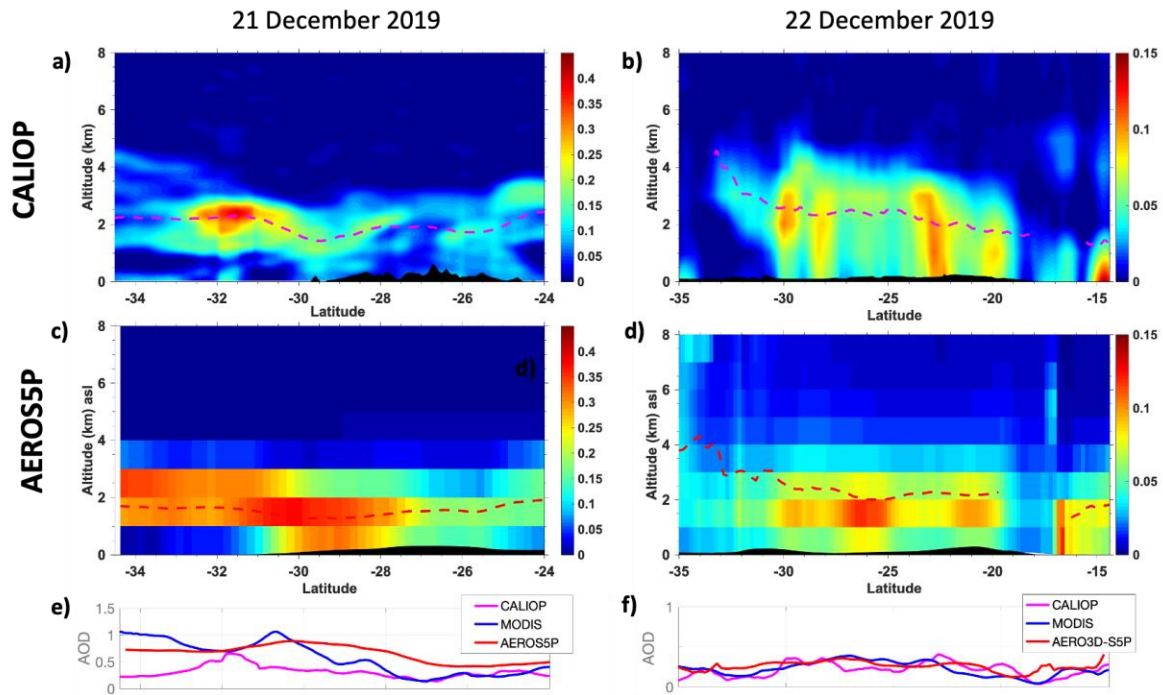


Figure III-13 Comparison of transects of aerosol extinction profiles (in km^{-1}) derived from (a,b) the CALIOP spaceborne lidar and (c,d) the AEROS5P passive approach on (a,c) 21 December and (b,d) 22 December 2019. Dashed lines indicate the mean altitude of the aerosol layers defined as the ones with half of the AOD above and below. The corresponding AOD at 550 nm from MODIS MAIAC and AEROS5P, and CALIOP at 532 nm is shown in panels (e,f).

3.3. Daily evolution of the 3D distribution of biomass burning aerosol plumes

During the period of analysis, exceptionally hot and dry conditions over Australia were responsible for numerous fires in a wide variety of locations. However, the most intense fires leading to the emission of massive amounts of smoke were seen principally located over eastern and southeastern Australia (principally in the New South Wales region [203]). This is seen in the example of fire detections from MODIS on 20 December, where the highest density of fire hotspots was located near the southeastern coast (Figure III-14c). As shown in terms of AOD and also CO retrievals derived from TROPOMI (Figure III-11 and Figure III-15), the densest smoke plumes were seen over the eastern part of Australia (east of 140°E). The 3D distribution of biomass burning aerosols on 20 December is presented in terms of iso-aerosol extinction surface in Figure III-14a. It shows that this large BB aerosols plume located at 25–37°S 140–150°E mainly extended vertically from the surface up to 3–4 km of altitude, with the core plume section reaching 5 km of altitude (seen for some pixels, see the aerosol top heights in Figure III-14b). According to estimations of the atmospheric stability in terms of Brunt–Vaisala

frequency N^2 (calculated at 0.25° degree resolution from ERA5 reanalysis) meteorological conditions at these locations led to a relatively stable atmosphere (values of $1.5 \times 10^{-4} \text{ s}^{-2}$). Winds at the northern sector of this aerosol plume blew towards the northwestern direction (Figure III-14b). The southern and western parts of this aerosol plume reached higher altitudes (5–7 km, seen for some pixels between clouds), co-located with a less stable atmosphere (lower values of N^2). Strong winds at 600 hPa (Figure III-14d) were expected to transport these aerosols in the southeastern direction over the Pacific.

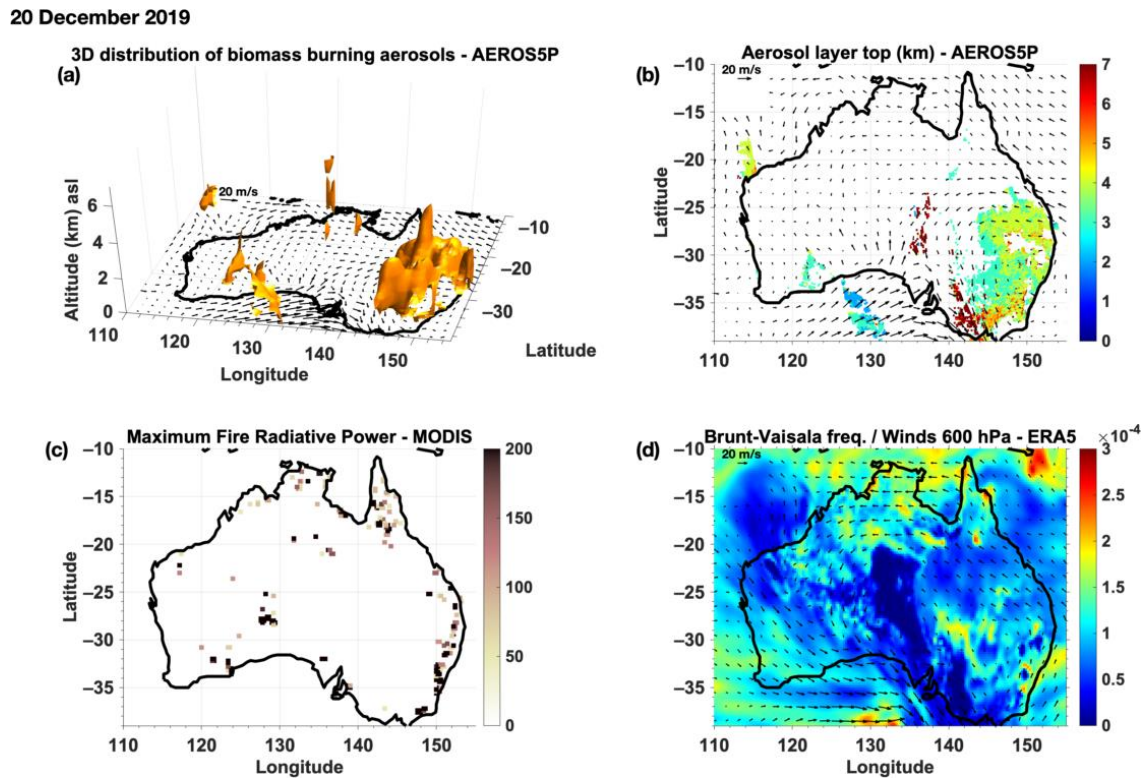


Figure III-14 (a) Three-dimensional distribution of biomass burning aerosols on 20 December 2019 derived from TROPOMI measurements using the AEROS5P method, depicted as iso-aerosol extinction surface of 0.1 km^{-1} . (b) Aerosol layer top derived from AEROS5P as the aerosol highest layer with aerosol extinction larger than 0.1 km^{-1} . (c) Maximum fire radiative power [megawatt] derived from MODIS sensors onboard Aqua and Terra within each $0.2 \times 0.2^\circ$ grid pixel. (d) Static atmospheric stability depicted by the Brunt–Vaisala frequency [rad/s] derived from ERA5 reanalysis. Winds [$\text{m}\cdot\text{s}^{-1}$] from ERA5 reanalysis are overlaid on panels (a,b) at 850 hPa and (d) at 600 hPa. For clarity, wind speed scales are shown in the upper left corners and wind vectors are displayed for each $1.5 \times 1.5^\circ$ grid point.

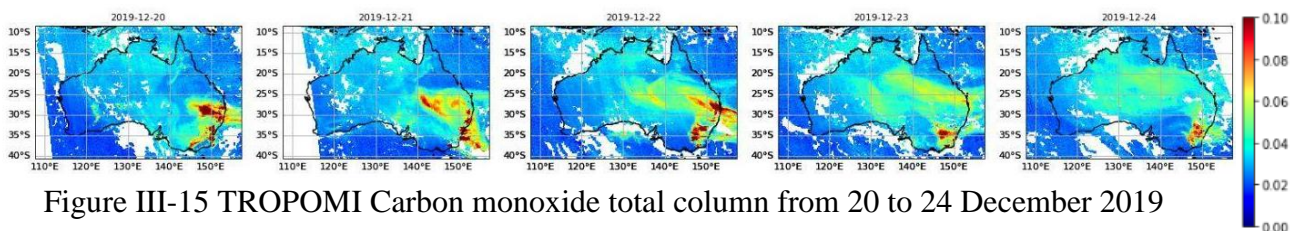


Figure III-15 TROPOMI Carbon monoxide total column from 20 to 24 December 2019
 [10^{24} molec.m⁻²].

On 21 December, the density of intense fires at the southeast hotspot (35°S 150°E) enhanced while it diminished elsewhere (Figure III-16c). Atmospheric stability was rather low above this region and within a band in the northwestern direction, according to low values of N^2 (Figure III-16d). As clearly depicted by AEROS5P, the 3D distribution of injections of biomass burning aerosol plumes formed a consistent structure extending from the surface near the fire region and ascending up to 7 km altitude at the northwestern fringe of the plume. The slanted shape of the aerosol plume is consistent with a northwestern direction of the wind at 600 hPa. The eastern part of this large aerosol plume (east of 145°E) extended vertically up to 3–4 km at maximum, which is concomitant with a clearly more vertically stable atmosphere (see higher N^2 values in Figure III-16d). Other aerosol plumes extending vertically up to 3–4 km (at the south-central part of the continent) were also co-located with stable conditions.

21 December 2019

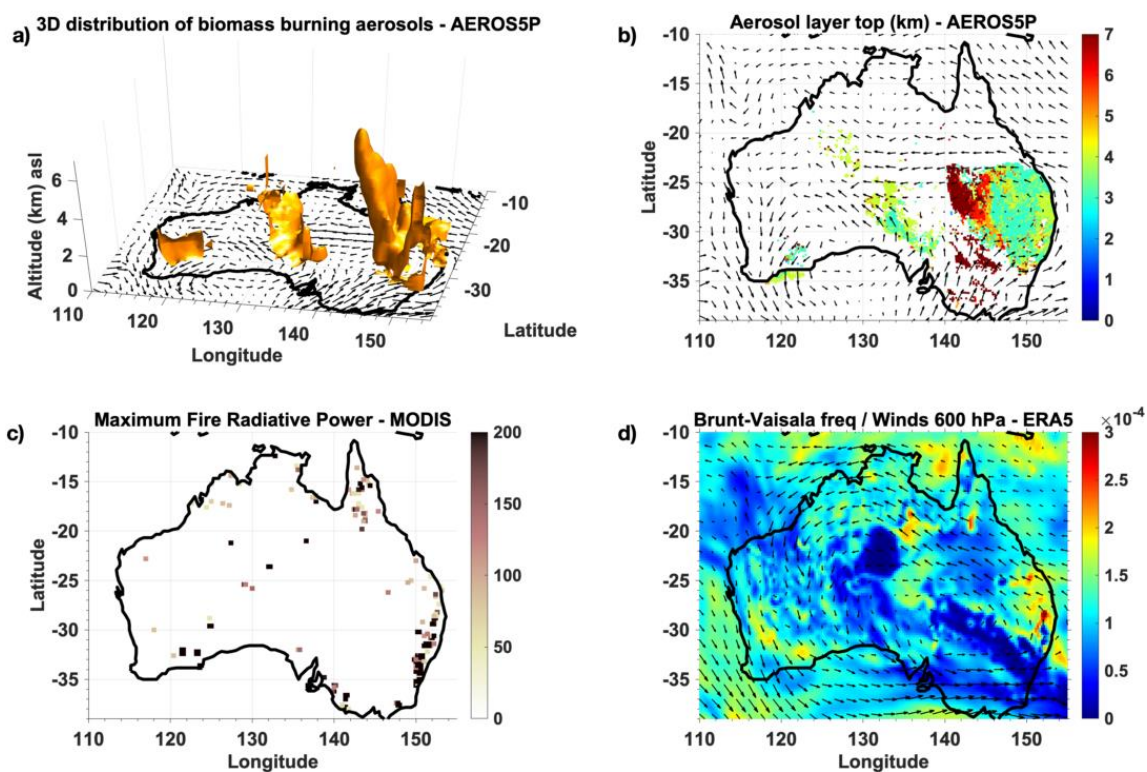


Figure III-16 Same as Figure III-14 but for 21 December 2019.

On the following day (22 December), the fire activity diminished over the Eastern coast of Australia (Figure III-17c). Two aerosol plumes were observed by AEROS5P (Figure III-17b,d). The aerosols located at central Australia were likely those emitted the previous day over eastern Australia and then transported northeastward by the winds (see clear and steady wind direction at 850 and 600 hPa in Figure III-17b,d). This aerosol plume distributed vertically from the surface up 4–5 km, while being located far from pyro-convective activity of the fire hotspots. The other aerosol plume over eastern Australia was likely freshly emitted by fires. It was injected up to 6 km of altitude (Figure III-17a,b), in a region with mostly low atmospheric stability.

22 December 2019

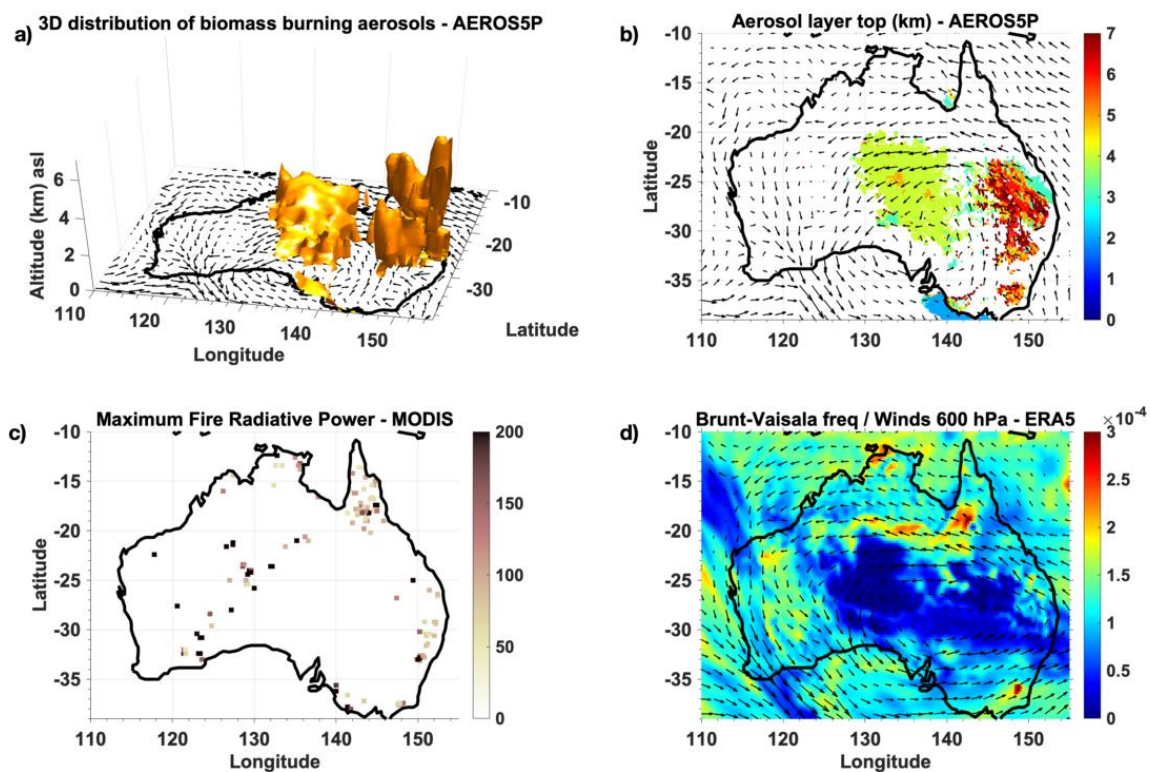


Figure III-17 Same as Figure III-14 but for 22 December 2019.

On 23 December, wildfire activity was almost absent over the eastern coast of Australia. The two aerosol plumes observed in central Australia by AEROS5P extended vertically up to 4 km of altitude. This limited vertical extent and their geographical location west from the fire regions suggest that they were likely smoke plumes emitted in the previous days. This is consistent with the reduction of new fire emissions on this day. In the absence of a direct link with fire emissions, low atmospheric stability (Figure III-18d) did not seem to favor the uplift of these aerosol plumes into higher atmospheric layers. A similar situation is noted on the following day (24 December 2019).

Additionally, we notice that the fire activity detected by MODIS in some regions did not necessarily imply the emission of large amounts of aerosols. This is the case for the example of the Cape York Peninsula located at the northeast of the continent, where fires were detected consistently every day (20–23 December) but dense aerosol plumes were not clearly seen as being emitted from this area. This might be linked to the type of vegetation in this area (Figure III-19). Grasslands and savanna are the main vegetation over the Cape York Peninsula, where fire events are more frequent but less intense [204,205]. Over the southeastern coast of the continent, the climate is more temperate and evergreen forests are present. This may explain the emission of large amounts of biomass burning aerosols (e.g., Figure III-9) and carbon monoxide (Figure III-15) from this area.

23 December 2019

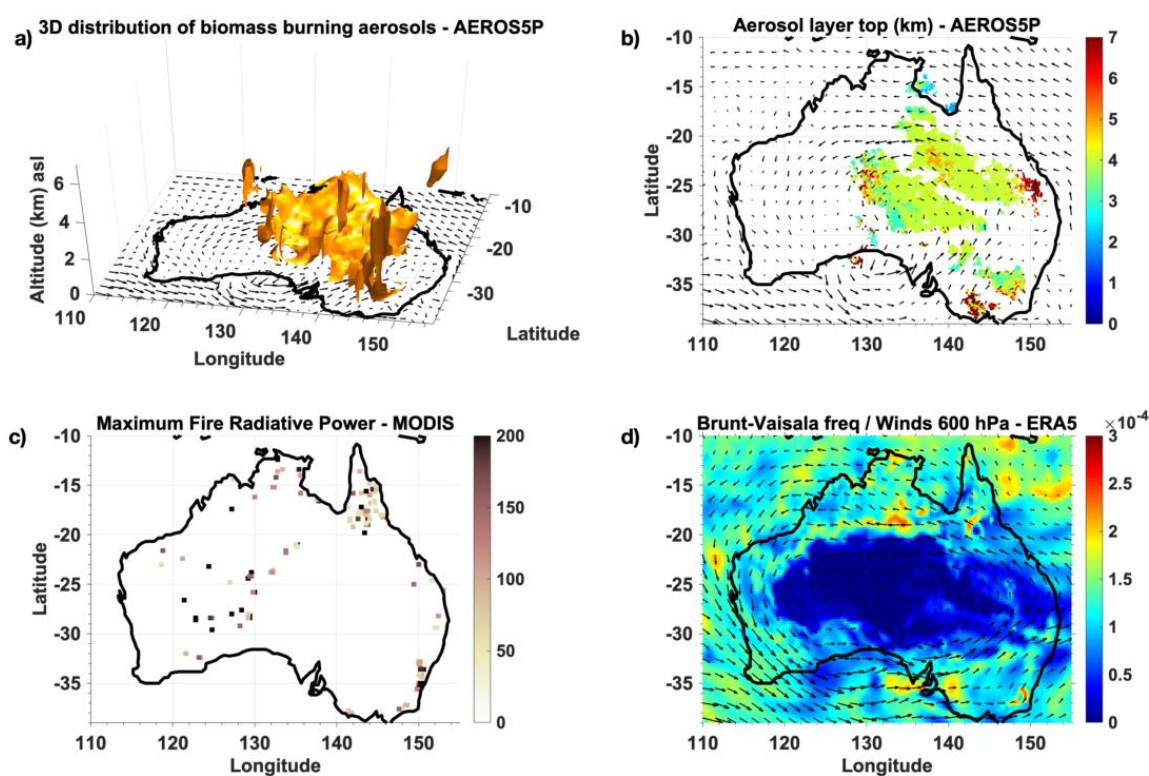


Figure III-18 Same as Figure III-14 but for 23 December 2019.

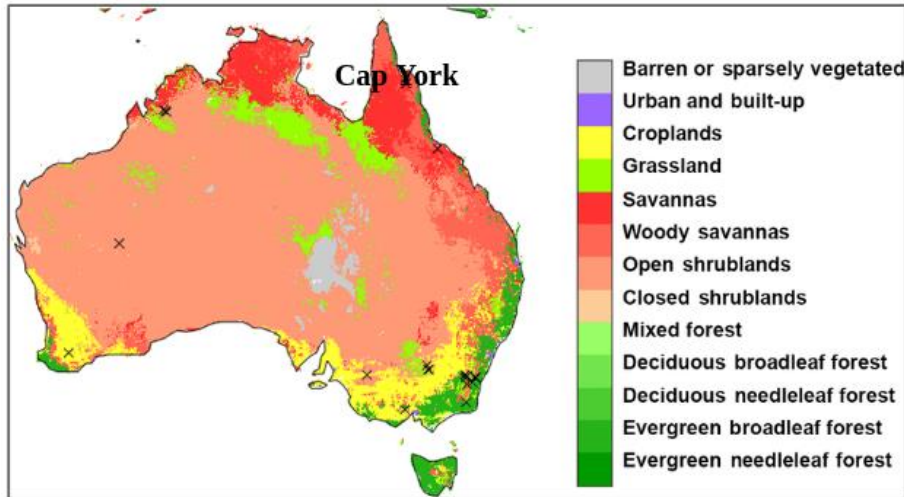


Figure III-19 Land cover type map for Australia [206].

4. Conclusion

We presented the new satellite-based AEROS5P approach, which enables the observation of the daily evolution of the 3D distribution of biomass burning aerosols. It was used here to characterize the 3D structure of aerosol plumes emitted by Australian wildfires during late December 2019. This approach derives vertical profiles of aerosol extinction at 550 nm in cloud-free conditions, by analyzing the reflectance spectra in the near infrared and visible measured by TROPOMI at $3.5 \times 5.5 \text{ km}^2$ resolution. The spectral signature at the oxygen A-band near 760 nm provides information on the altitude of the aerosols, while that in the visible spectrum from 406 to 674 nm provides information about their total column abundance. These two quantities are derived simultaneously.

Aerosol optical depths retrieved by AEROS5P show good agreement with those derived from MODIS and VIIRS, with similar horizontal distribution. These datasets present correlations of up to 0.87 and root-mean-squared differences of 0.3. Transects of vertical profiles of aerosol extinction derived by AEROS5P show similar structures of aerosol plumes to those derived from measurements of the CALIOP spaceborne lidar. Mean altitudes of aerosol layers derived by AEROS5P and CALIOP are also in good agreement. These two datasets show a small mean bias (185 m), root-mean-squared differences of 500 m and a correlation coefficient of 0.83, for the mean plume altitude, for this event.

Moreover, AEROS5P provides an unprecedented daily description of the 3D structure of the vertical injection of smoke plumes. The 3D distribution of these plumes is consistent

with fire radiative power hotspots detected by MODIS and atmospheric stability described by a buoyancy frequency diagnostic. For the period of analysis (20–24 December 2019), we remark that aerosol plumes linked with intense fire activities and relatively low atmospheric stability were injected into the atmosphere up to 6 and 7 km of altitude. On the other hand, aerosol plumes extended vertically up to 3–4 km for relatively higher atmospheric stability even near the emission regions. Aerosol plumes reach these altitudes once they are transported far away from sources.

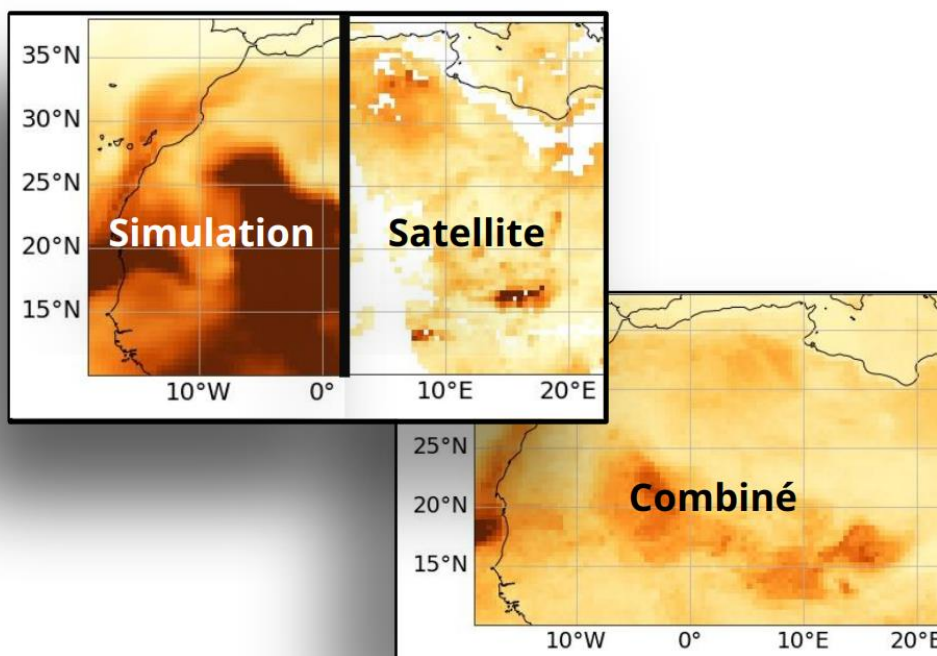
The present chapter presents the first application of the AEROS5P approach for a major outbreak of mainly biomass burning aerosols, assuming their microphysical properties for this event. Preliminary tests suggested reasonable performances of the current version of the method for wildfire events in other regions (not shown), for situations without mixing with other aerosol types. As typically done for other satellite aerosol products, we expect better performances using sets of aerosol properties (mainly size and refractive index) adapted for different regions. Future work will be dedicated to consolidating and extending the approach to more diverse situations, such as other periods when BB aerosols from the Australian fires reached the stratosphere or other large fire events in other regions (e.g., Central and South Africa, California, Siberia, etc.). Additional tests on the advantages of jointly adjusting aerosol properties such as size will be considered. Moreover, these studies will include sensitivity analyses on the impact of assumptions and uncertainties of the inputs of the method. More accurate cloud screening is also expected to improve the performance of the method. In addition, future versions of the AEROS5P will also aim at characterizing other types of aerosols such as those from anthropogenic pollution like that of nitrates.

Data Availability Statement: The data used in this study was from the following platforms: MODIS AOD data (1 September 2021), MODIS Fire Power (4 October 2021), VIIRS (19 September 2021) data from EarthData portal from NASA (<https://search.earthdata.nasa.gov/>); CALIOP measurements (13 September 2021) from the AERIS/ICARE datacentre (<https://www.aeris-data.fr/> and <http://www.icare.univ-lille1.fr/>); TROPOMI L1B data (13 July 2020) from the Sentinel 5P hub (<https://s5phub.copernicus.eu/>); TROPOMI Instrument Spectral Response Function (20 May 2020) from: <http://www.tropomi.eu>; AERONET AOD measurements (16 March 2021) from the AERONET portal (<https://aeronet.gsfc.nasa.gov/>); TROPOMI surface DLER database from Sentinel 5p+ Innovation project (https://www.grasp-sas.com/projects/aod-brdf_sentinel-5p-

innovation/); ERA5 meteorological reanalyses (21 September 2021) from the ESPRI IPSL mesocentre (<https://mesocentre.ipsl.fr/>). The AEROS5P data generated within this work are publicly available on the following repository link: <https://trng-b2share.eudat.eu/records/c5ce76e0247547a3b1d276738b9b17de> (accessed on 20 May 2022) of the collaborative data infrastructure EUDAT/B2SHARE (<https://b2share.eudat.eu>).

IV. Chapitre 4 : Synergie des observations satellitaires et la simulation de chimie-transport

Résumé.....	125
1. Synergy of satellite observations and chemistry-transport simulation	129
2. AOD correction approach.....	130
3. Inputs.....	131
3.1. MODIS satellite observations	131
3.2. CHIMERE simulations	132
3.3. Dataset preparation	136
4. Bias correction ML models construction.....	138
4.1. Multiple linear regression	139
4.2. Feed-forward neural networks	140
4.3. Random forest.....	142
4.4. Gradient boosting.....	143
5. Results and discussion	143
5.1. Comparison against independent MODIS observations	144
5.2. Comparison with AERONET ground based measurements.....	154
6. Conclusions.....	156



Résumé

La distribution spatiale des aérosols peut être caractériser en termes d'AOD à partir des observations satellitaires ou avec la modélisation de chimie-transport. Ces deux approches ont des avantages et des limitations. Les observations satellitaires tel que celles de MODIS sont assez précises, cependant elles ont une couverture spatiale limitée à l'angle de vue, et aux situations de ciel clair, en plus de la faible fréquence d'observation (un passage en journée pour les orbites polaires). Quant aux simulations de chimie-transport, la couverture spatio-temporelle est complète (sur la région simulée 24 h sur 24) mais elles sont limitées par leurs précisions. Ces dernières sont autant plus faibles en zones qui manquent de stations de mesure in situ pour ajuster les paramètres de la simulation et assimiler les mesures. L'étude développée dans ce chapitre sur l'évaluation de la précision d'un modèle de chimie-transport a montré que la simulation de l'AOD est 3.5 fois moins précise que l'observation satellitaire. Les incertitudes peuvent être attribuée au approximations physiques et numériques, mais surtout aux erreurs dans les données utilisées tel que les flux des émissions, la météorologie, les propriétés microphysiques et optiques des particules simulées, etc.

Il est possible d'estimer la distribution spatiale des aérosols avec une bonne précision, comme celle des produits satellitaires, avec une couverture spatio-temporelle complète, à l'instar des modèles. Les méthodes traditionnelles de synergie modèles-observations reposent sur l'assimilation de données. Cette dernière est très couteuse en termes de coût computationnel. Dans ce chapitre, je développe une approche d'apprentissage automatique pour corriger les simulations d'un modèle de chimie-transport avec des observations satellitaires d'AOD. Ce couplage novateur est implémenté particulièrement pour améliorer la précision de la distribution horizontale des aérosols simulée par CHIMERE en terme d'AOD. Ceci est réalisé à l'aide des observations de MODIS utilisées comme référence lors de l'apprentissage supervisé.

Les observations d'AOD de MODIS sont utilisées pour la supervision car elles sont plus précises par rapport à la simulation, ceci est vérifié en comparant avec les observations du réseau des photomètres solaires AERONET. L'écart quadratique moyen (RMSE) entre l'AOD de MODIS et celle d'AERONET étant égale à 0.13 seulement avec un coefficient de corrélation égale à 0.88. À l'opposé, la simulation CHIMERE montre une estimation moins précise (coefficient de corrélation est seulement 0.63 et l'RMSE égale à 0.46). Nous choisissons

l’Afrique du Nord comme région pour une première analyse, étant donné la forte variabilité de l’AOD et les incertitudes élevées des modèles de chimie-transport quant aux simulations des tempêtes de poussières désertiques et ses propriétés optiques.

Quatre types de modèles d’apprentissage supervisé sont mis en œuvre : la régression linéaire multiple (MLR), les forêts aléatoires (RF), l’amplification de gradient (XGB, pour *extreme gradient boosting*) et les réseaux de neurones artificiels (NN, pour *neural networks*). Ces modèles prennent comme entrée l’AOD simulée par CHIMERE, quelques profils verticaux de concentrations (poussière, carbone organique, ozone, PM...), des données de spécificité de surface (albédo, humidité au sol...), la hauteur de la couche limite et des conditions météorologiques (pression, humidité relative...). En totale la taille du vecteur d’entrée (pour un pixel à prédire) est de 96 spécificités, sachant qu’on prend uniquement quatre valeurs de chaque profil vertical. Ces données sont tirées des sorties de la simulation CHIMERE pour l’Afrique du Nord pour l’année 2021. L’entraînement a été fait en utilisant un tiers des données disponibles, les deux autres tiers sont gardés pour l’ajustement des paramètres internes des modèles et pour la prédiction.

Nous avons effectué des comparaisons des résultats de la prédiction (correction) de l’AOD de CHIMERE avec des observations satellitaires et celles issue de 8 stations d’AERONET. Ce sont des données qui n’ont pas été utilisé pour l’apprentissage des modèles. Les comparaisons suggèrent que tous les modèles ont des performances comparables. Le modèle d’apprentissage automatique le moins performant en termes de correction de l’AOD est MLR, les trois autres modèles ont des performances statistiquement comparables, avec un léger avantage du modèle RF qui exprime moins d’artefacts spatiaux. Ces derniers sont de occasionnelles zones de très faibles AOD prédit par NN et MLR. En plus, les prédictions d’AOD du modèle XGB sont légèrement bruitées par rapport aux prédictions des autres modèles, ceci pourrait être expliqué par le sur-apprentissage du modèle XGB des AODs de MODIS.

Une analyse de l’importance des variables géophysiques a été réalisé. Cette étude de la contribution marginale des composant du vecteur d’entrée a montré statistiquement que l’humidité, et la concentration des PM sont des grandeurs conséquentes pour les quatre modèles de correction de l’AOD. On arrive aussi à attribuer certains aspects régionaux de l’AOD prédite a la variable responsable, tel que le cas des nitrates pour la correction de l’AOD en péninsule Arabique, et le carbone organique qui engendre les quelques artéfacts spatiaux.

Bien que la correction sous-estime légèrement les AOD très élevées (principalement au Sahara), elle réduit considérablement l'erreur absolue moyenne (0.12), tout en améliorant clairement les corrélations linéaires par rapport aux observations de MODIS (0.71). Cette approche constitue un premier pas vers la synergie entre le modèle et l'observation satellitaire pour corriger le profil vertical de concentration d'aérosols en utilisant des produits de télédétection d'aérosols comme AEROS5P, ou des lidars.

Les prochaines sections du présent chapitre sont rédigées en anglais, car ce sont des travaux publiés dans un journal international.

1. Synergy of satellite observations and chemistry-transport simulation

Satellite measurements, such as those derived from the Moderate Resolution Imaging Spectroradiometer (MODIS) spaceborne sensor [207], play a fundamental role in observing the spatial distribution of aerosols at the global scale. These remote sensing observations are mainly derived in terms of the AOD, which describes the total column extinction that is integrated over the whole atmosphere. They therefore provide maps of the horizontal distribution of aerosol abundance. However, they are only available for cloud-free conditions and are limited by the overpass time of the (polar orbiting) satellite. On the other hand, chemistry transport models (CTMs), such as CHIMERE [208], numerically simulate the hourly 3D evolution of aerosol plumes in the atmosphere, independent of the cloudiness. The accuracy of the CTM simulations depends on the precision of the inputs, e.g., the emissions of atmospheric constituents, wind, vertical velocity fields, and on the assumed aerosol properties (e.g., microphysical and optical properties). Due to assumptions and inaccuracies in these datasets, simulations of the aerosol spatial distribution are prone to bias compared to observations [146,209]. These errors are related to uncertainties in the physical parameterizations of the model, input data, and numerical approximations [9]. Modeling errors are more pronounced in regions lacking ground-based stations to validate and constrain the simulations, which is the case for the African continent and the Middle East.

Most of the techniques used to constrain chemistry transport models use ground-based in situ measurements and, more recently, satellite data such as MODIS AOD, e.g., for the North African mineral dust emissions inversion [210,211] and for the Copernicus Atmosphere Monitoring Service (CAMS) [212]. They are mainly based on data assimilation techniques such as variational or filtering approaches [213], which are computationally expensive. On the other hand, fast and computationally efficient approaches use machine learning (ML) techniques to correct the modeling systematic biases. These methods are increasingly being used due to the progress in the development of ML hardware and technology. ML bias correction techniques are mainly used for chemistry transport model simulations of trace gases, e.g., [214], or used in situ surface data for aerosols, e.g., [215,216], or for post-processing weather forecasts, e.g., [217,218]. ML is used in the context of the monthly mean AOD estimation as well [219].

In this work, we develop a new ML-based bias correction for post-processing the CHIMERE-simulated AOD maps at 550 nm wavelength for 13:00 LT (local time). The method improves the regional simulated AOD at the scale and the resolution of the simulation ($0.45^\circ \times 0.45^\circ$) and for all sky conditions. The correction relies on the good accuracy of AOD satellite measurements from MODIS taken at the same wavelength and approximately the same local time. Observations are used only in the training phase, and the method provides full-coverage and daily maps of the corrected AODs over the entire CHIMERE simulation domain. The method is developed for the North African region using data from the year 2021, which we introduce in Section 2. In the sections 3 and 4, we describe the inputs and the implementation of the four different ML models : multiple linear regression, neural networks, random forest, and gradient boost model against independent, ground-based, and satellite observations. A test case is presented in Section 5 to assess the effectiveness of the correction and discuss the relative and regional influence of some geophysical inputs on the inferred AOD.

2. AOD correction approach

The principle of the AOD bias correction method relies on the use of the AOD and other atmospheric composition variables simulated by CHIMERE, in addition to certain meteorological fields. These constitute an input for a trained ML regression model. The latter derives a posteriori bias corrected AODs that are consistent with the AOD observations that were used for training the ML model beforehand, which are MODIS AODs. This inference, also called prediction in other contexts, is performed on a pixel-by-pixel basis for all ground pixels, i.e., no surrounding regional information is used for the AOD correction. A schematic description of this process is shown in Figure IV-1. We assume that the AOD is dependent on certain inputs, and we derive an approximate relationship that maps the inputs to the correct AOD using statistical modeling. This type of modeling is data-driven, meaning that it requires a large database of known solutions encompassing a broad range of possible scenarios.

While the method is applicable for any region, we choose North Africa and the Arabian Peninsula for this study because desert dust and anthropogenic emissions in this area are highly uncertain. Therefore, chemistry transport simulations over this region are particularly imprecise and require correction. In addition, this region provides an interesting AOD training dataset, since clear sky conditions allow greater availability of satellite data over large areas with a wide range of AOD values.

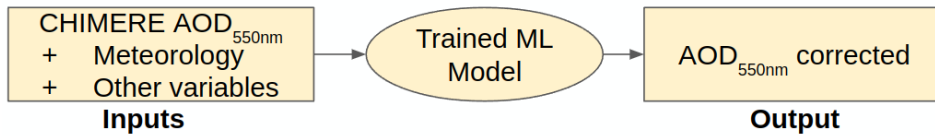


Figure IV-1 Flowchart of the pixel-by-pixel bias correction approach of the CHIMERE AOD estimation.

3. Inputs

3.1. MODIS satellite observations

The satellite observations used for training the ML model are derived from the MODIS spaceborne multi-wavelength radiometer onboard the AQUA platform with an overpass time around 13:30 LT. This satellite is in a near-polar sun-synchronous low Earth orbit, hovering at an altitude of 705 km and within the A-Train constellation. MODIS observes the Earth's backscattered radiation in 36 spectral bands with a horizontal resolution ranging from 250 to 1000 m and a nadir swath 2330 km wide, covering most of the Earth's surface on a daily basis [87]. The MODIS AOD product used in this work is the collection 6.1 MYD04_L2 with a resolution of 10 km [220,221]. This product is a combination of AOD products derived using the Dark Target [222] and Deep Blue [223] algorithms. The Dark Target algorithm is suitable over ocean and dark land (e.g., vegetation), while the Deep Blue algorithm covers the entire land area, including both dark and bright surfaces. This MODIS dataset is interpolated at the horizontal resolution of the CHIMERE model ($0.45^\circ \times 0.45^\circ$) using the mean value method. All MODIS AODs mentioned hereafter refer to the observations from the AQUA platform at the wavelength of 550 nm, unless otherwise noted.

- **MODIS AOD estimation accuracy assessment**

The quality of the reference dataset for supervision is critical for accurate ML modeling. Therefore, we filter out low-quality observations by keeping only the retrievals with the highest quality assurance flag. We verify the quality of the MODIS AODs used here by comparing them (Figure IV-2a) with daily averages of direct measurements of the AOD from ground-based sun photometers of the AEROSOL ROBOTIC NETWORK [224]. This network enforces standardization of instrumentation, calibration, and processing to ensure the best quality. The comparison (Figure IV-2a) shows that the AOD observations from MODIS for the year 2021 correlate well with the measurements from the eight AERONET stations (the location of each station in Figure IV-3). The Pearson correlation coefficient (r) is 0.88 and the root mean

squared error (RMSE) is quite low (0.13). The AERONET data being used for the comparison are of version 3, level 1.5. We convert the AOD from 500 nm to 550 nm using the 440–675 nm Angstrom exponent provided by the same station.

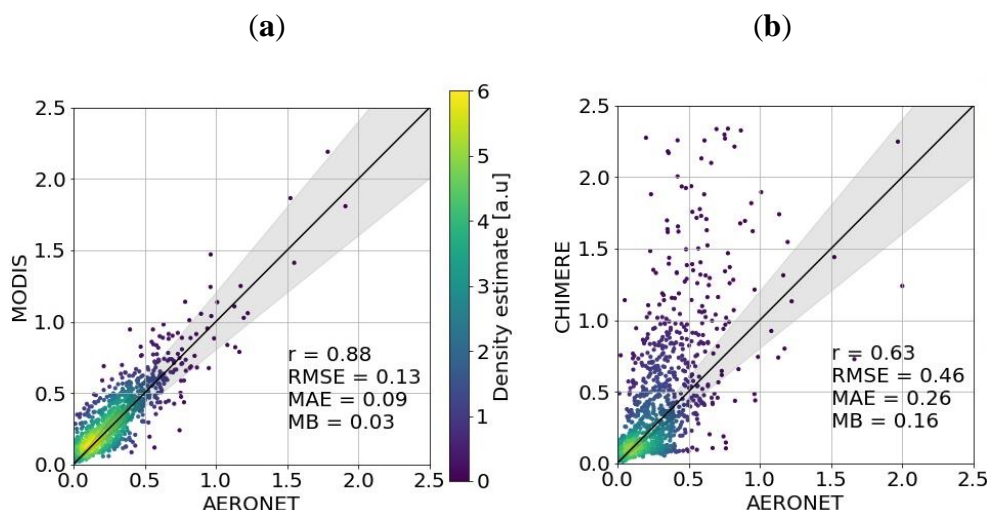


Figure IV-2 Scatter plot of 966 collocated data points comparing the AOD at 550 nm measured by 8 AERONET sun photometers for the year 2021 with (a) MODIS satellite observations and (b) simulated AODs by the CHIMERE chemistry transport model. The black line is the $y = x$ line with the gray shaded areas representing the $\pm 20\%$ interval of its slope.

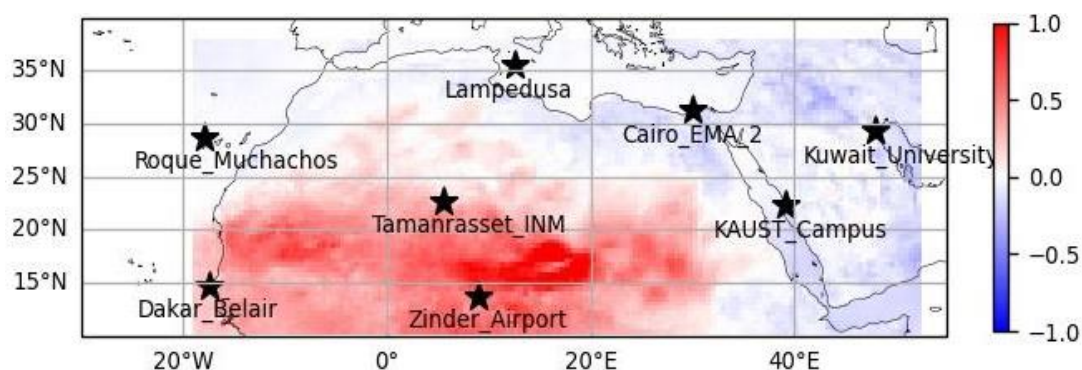


Figure IV-3 Annual median of the difference between CHIMERE and MODIS AOD at 550 nm for the year 2021 over the simulation domain of North Africa and the Arabian Peninsula. The stars indicate the 8 AERONET stations used for the comparisons in Figure IV-2.

3.2. CHIMERE simulations

CHIMERE is an Eulerian CTM that simulates the formation, deposition, and transport of aerosols and other atmospheric species [225]. It is capable of simulating phenomena ranging from the local scale, such as urban heat islands, to the hemispherical scale. The principle is to

use the available information of the Earth's atmospheric composition, and its interactions with the surface, including source emissions, and then simulate the evolution of the atmospheric species, taking into account the internal forcing, e.g., wind, and the external forcing, such as the incoming shortwave radiation. The evolution of the atmospheric constituents of the plumes is calculated by numerically solving the transport equation and using a chemical interaction scheme. CHIMERE is used as a tool to forecast and analyze the daily air quality in terms of particulate and gaseous pollution [124]. It is widely used for aerosol, and aerosol precursor research, e.g., [34,120–123].

The CHIMERE simulations used as input for the ML model are derived for the 12 months of the year 2021 using CHIMERE-2017 version [118] over the region of North Africa and the Arabian Peninsula (10 to 38°N 19°W to 53°E). The horizontal resolution is $0.45^\circ \times 0.45^\circ$ on a regular grid, while the vertical resolution consists of 20 layers of increasing thickness, from about 30 m at the surface level to 675 m (upper limit about 500 hPa). The aerosols are distributed into 10 size classes (also called bins), ranging from 0.05 μm to 40 μm . The AOD is calculated for all the aerosol species considered in the simulation, assuming external mixing. This is calculated with the aerosols concentrations and using online Fast-JX (version 7.0b) photolysis calculations [138], which provide the optical properties [226]. The calculation is performed at five wavelengths (200, 300, 400, 600, and 999 nm). In this work, we use the 400 nm AOD interpolated to 550 nm using the 400–600 nm Angstrom exponent.

The simulation boundary and initial conditions are taken from the *Laboratoire de Météorologie Dynamique general circulation* model coupled with the interaction with chemistry and aerosols (LMDz-INCA) [126]. The CHIMERE simulation is run in offline mode using National Oceanic and Atmospheric Administration (NOAA) meteorological final analysis data and the Weather Research and Forecasting Model (WRF), version 4.1.1 [227]. MELCHIOR2 (*Modèle Lagrangien de Chimie de l'Ozone à l'échelle Régionale*) [134] is used as chemistry interaction scheme. An emission inventory derived for 2015 from the Emissions Database for Global Atmospheric Research EDGARv5.0 [127] is used for anthropogenic gaseous and particulate matter emissions. The dust emission scheme implemented in CHIMERE requires the knowledge of soil properties and wind conditions. This module allows for the calculation of dust aerosol emissions and their size distribution, by modeling the processes of sandblasting and saltation [125,228]. It takes into account the uplift of both silty and sandy soils, which are emitted from the northern Algerian region and the Sahara,

respectively [229]. It should be noted that non-negligible uncertainties remain in the modeling of Saharan dust emission and transport [230].

- **CHIMERE AOD estimation accuracy assessment**

An evaluation of the accuracy of the AOD simulated by CHIMERE with respect to AERONET reference measurements is shown in Figure IV-2b. For high aerosol abundance cases, the CHIMERE AOD is clearly overestimated with respect to the AOD measured by the AERONET sun photometers. The mean bias (MB) is positive and equal to 0.16, but negative biases are observed for background AODs. The correlation coefficient is significantly lower (0.63) than its value for the comparison of MODIS with AERONET AOD (0.88). The RMSE is much larger, being 0.46 for CHIMERE AOD against 0.13 for MODIS. The same behavior is observed for the mean absolute error (MAE). Other chemistry transport models have also shown biases in the AOD simulated over North Africa, with comparable overestimation, e.g., [231].

The map of the annual median difference between CHIMERE and MODIS AOD (Figure IV-3) shows that the major overestimation is located over most of the Sahara Desert. The positive bias of the CHIMERE AOD is most pronounced over a region northeast of the Lake Tchad named the Bodélé Depression, where a bias greater than unity is found for half of the days of the year. This region represents an important natural emission source of mineral dust (Figure IV-4b), with a frequency of about 100 days per year [232] and an estimated abundance of 40 million tons of dust per year. The large positive biases of the CHIMERE simulation's AOD may be related to the uncertainties in the inputs of the dust emission model, e.g., near-surface wind speed and friction velocity, but may also be due to the dust refractive index and size distribution used for the AOD calculation.

Otherwise, the AOD over the Arabian Peninsula and northern Egypt is slightly underestimated by CHIMERE, as compared to MODIS. We see a low correlation (0.53) and a high standard deviation between the two estimates, leading to an RMSE of 0.62 and an MAE of 0.35.

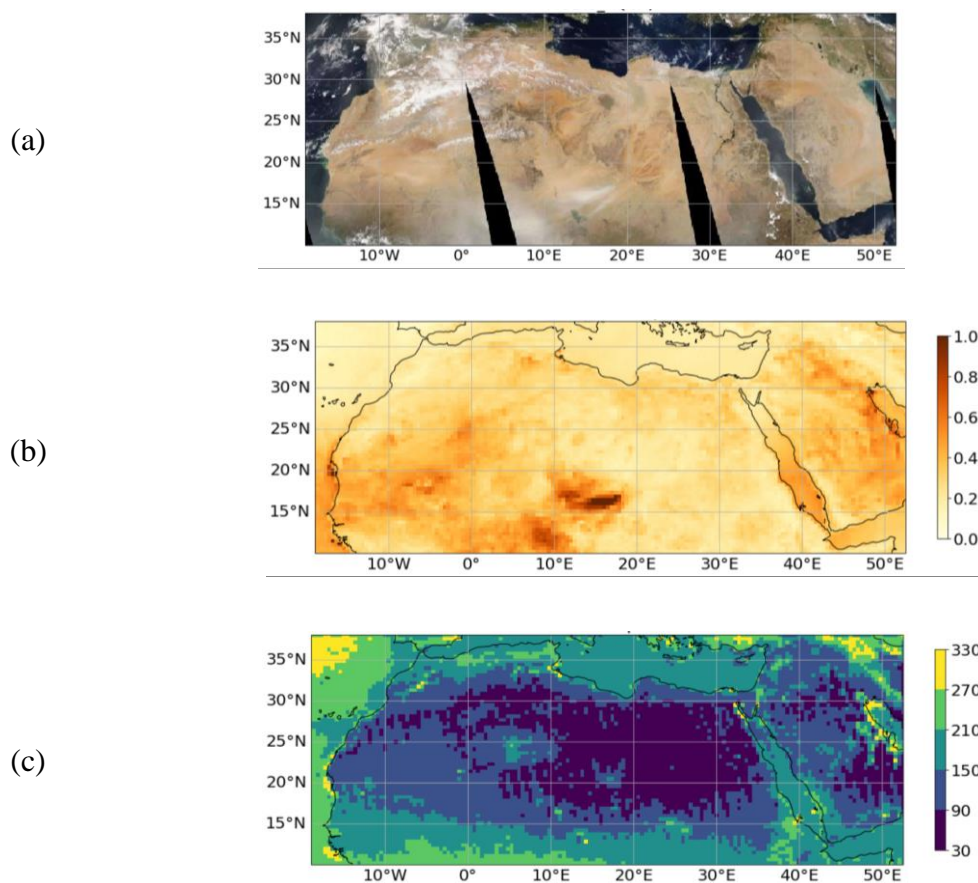


Figure IV-4 (a) MODIS true color showing the mineral dust emitted from Bodélé depression region on 23 March 2021. (b) Average MODIS AOD₅₅₀ values for 2021. (c) Number of observations used in the computation of the average AOD₅₅₀ of panel (b)

The bias of CHIMERE raw AOD simulation output with respect to MODIS AOD is assessed for different AOD intensities (Figure IV-5) for the whole year. We clearly that the bias is more important and tends toward positive values at higher CHIMERE AODs than lower ones; the linear correlation has a slope of 0.79 and the intercept value is low (-0.2). The linear correlation of the bias with the raw AOD is among the highest correlations with the other CHIMERE variables like PM₂₅, PM₁₀ (not shown here). The exact reason for the linear relationship between the bias and the raw AOD is unknown, yet our method that we describe in the next section is a ML model designed for improving these biases a posteriori.

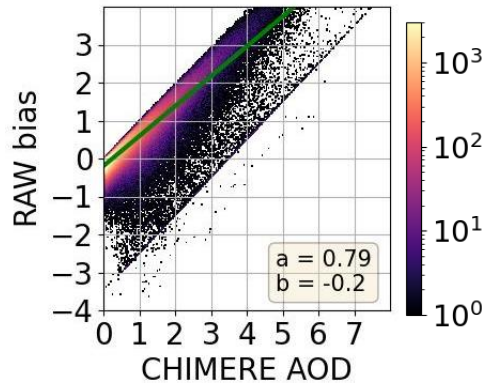


Figure IV-5 Scatterplot showing the linear relationship between CHIMERE AOD₅₅₀ bias toward MODIS observations for different AOD₅₅₀ levels represented by the CHIMERE AOD₅₅₀. The colorbar represents the occurrence in the year 2021. The green line represents the linear fit with the correlation coefficient a and the intercept b .

3.3. Dataset preparation

We chose the days used for training following the common convention in the field of machine learning, i.e., dividing the whole 12 months of 2021 into two parts : 66% is used in the training and validation phase, and the remaining 33% is reserved for testing the performance of the models.

Adequate size and diversity of the training data are important for a good generalization of an ML model. In order to avoid the bias caused by seasonal variability, we take two-thirds of each month in 2021 to build the models. Specifically, we use consecutive days from the 1st to the 20th of each month for the training and validation stages, making 240 days from 2021. We call this dataset D_{Train} and it consists of about 1.4 million cloud-free ground pixels. According to several tests performed during the design of the ML models, this training data sampling strategy allows for good performance at any season and at diverse pollution levels, from low background pollution to intense dust outbreaks. Even though D_{Train} is made of sequences of shuffled pixels of consecutive images, the correlation between a training subset and a validation subset is limited due to the decorrelation time scale of the atmospheric processes. This is a key point to ensure that the test dataset is independent and that the computed scores are robust to evaluate the generalization ability of the ML model.

The period from the 21st to the end of each month are reserved to test the ML models, which amounts to 122 days not used in the training process. We call this dataset D_{Test} and it

consists of about 0.7 million ground pixels of the CHIMERE simulation. We had to exclude the days 23, 24, and 25 of September 2021 due to the unavailability of MODIS observations.

Besides the CHIMERE-simulated AOD, which we will refer to as the “RAW AOD”, additional atmospheric composition variables from CHIMERE are used as inputs of the ML models. We consider the following ones: vertical concentration profiles of desert dust, sea salt, organic carbon particles, water droplets, ammonium particles (NH_4^+), sulfate particles (SO_4^{2-}), nitrate particles (NO_3^-), particulate matter (PM_{10} and $\text{PM}_{2.5}$ with diameters smaller than, respectively, 10 and 2.5 μm), carbon monoxide (CO), ammonia (NH_3), toluene, ozone (O_3), nitrogen dioxide (NO_2), sulfur dioxide (SO_2), nitrous acid (HNO_2), specific humidity, hydroperoxide (ROOH), non-methane hydrocarbons (NMHC), hydroxyl radical (OH), and the sum of nitrogen monoxide and dioxide (NO_x) and reactive nitrogen compounds (NO_y). For these vertical profiles, we consider 4 simulation levels out of the 20 available, corresponding to approximately 967, 920, 797, and 560 hPa pressure levels. The other CHIMERE outputs used as inputs for the ML are related to surface properties and NOAA reanalysis meteorological variables: surface albedo, shortwave radiation flux, soil moisture, surface relative humidity, boundary layer height, surface latent heat flux, surface sensible heat flux, pressure profile, and relative humidity profile. We do not include wind fields because they do not improve the AOD prediction residuals. All CHIMERE and meteorological variables are taken at 13:00 LT, which is the closest to the AQUA platform spacecraft overpass time. Therefore, D_{Train} and D_{Test} have a total of 96 features (images). The terms variable and feature could be used interchangeably, but to be specific, the first one is used to denote any measurable geophysical quantity, while the second one can be used to refer to a variable or a subset of a variable, such as the values of a variable at a particular altitude.

In general, the ML input variables are expected to be directly or indirectly related to the AOD or to aerosol concentrations. Selecting only a subset of the available variables accelerates the training and inference time, mitigates the overfitting, and reduces the curse of dimensionality [233]. The selection of the input variables for the ML models was based on numerous empirical tests. After considering all the variables and performing several modeling trials, we excluded variables that showed a tendency to lower performance of the modeling (lower correlation with reference data and higher modeling residuals) and kept those that gave better performance (higher correlation and lower modeling residuals). Knowing that exhaustive trials were not possible, these tests were conducted using the simplest and most straightforward

ML model (i.e., the multiple linear regression, described in Section 4.1) to keep the computation time relatively short. We kept correlated variables such as PM_{10} and $PM_{2.5}$ (see Annexe A) when they improve the performance of the AOD correction.

Although most features of D_{Train} as well as D_{Test} follow a log-normal distribution, e.g., AOD (Figure IV-6) and PMs, using the normal logarithm of these variables as inputs to the ML models results in underestimations of the predicted AODs. Our experiments show that preprocessing techniques such as standardization/rescaling, and/or outlier removal do not significantly improve the accuracy of the ML approach either. Therefore, we keep the input data, without any preprocessing transformation. We use the same input variables for all ML models developed in this study (see Section 3.3), to ensure consistency and comparability during performance assessment in Section 5.

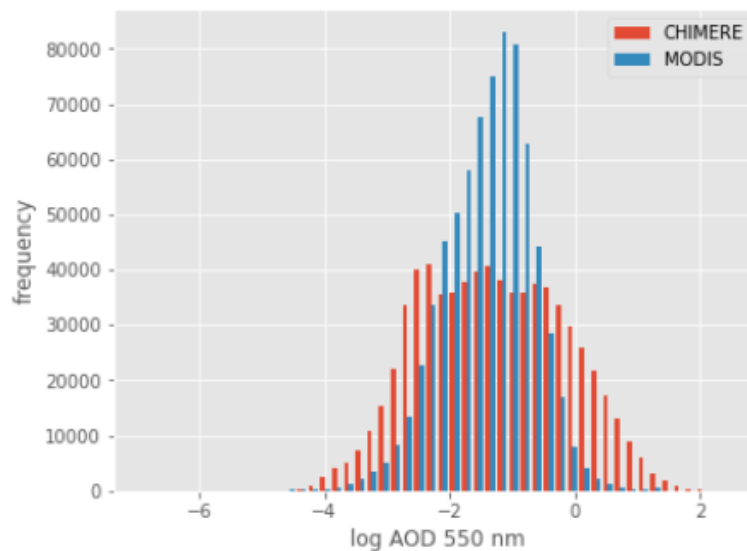


Figure IV-6 Histogram showing the logarithmic distribution of the AODs for both CHIMERE and MODIS over North Africa for the year 2021.

4. Bias correction ML models construction

The fitting problem we are trying to solve is overdetermined and the high dimensional input variables are not independent. We implement several machine learning approaches, tree based, a neural network model, and the more classical linear regression model. In the following sections, we describe how the four models are built (Section 4) using the Python programming language version 3.8 [234] and the web-based interactive computing environment Jupyter notebook [235]. While training is partially performed on a Quadro P620 GPU, inference runs

only on a single thread of a CPU (i9 2.30GHz). The time required for training and inference of the ML models is an important indicator of the usability of the developed method. The training phases are relatively short ; in fact, the multiple linear regression takes a few seconds, and the training of the other ML models does not exceed one hour each (not counting the hyperparameter search step, which is much longer). Regarding the inference time, all approaches perform the daily correction of the AOD in less than half a second (Table IV-1), which allows for a potential use of these models in real-time applications.

Table IV-1 Average performance of each AOD bias corrector and raw CHIMERE on D_{Test} with respect to MODIS AOD. t: daily mean inference time cost; r: Pearson correlation coefficient; RMSE: root mean squared error; MAE: mean absolute error; Skp: Pearson’s coefficient of skewness; and μ : mean; percentages show the percentiles of the error. The metrics are calculated for the 737129 pixels making D_{Test} .

	t(s)	r	RMSE	MAE	Skp	μ	min	25%	50%	75%	max
RAW	N/A	0.56	0.65	0.37	2.55	0.24	-3.57	-0.09	0.03	0.39	6.95
MLR	0.19	0.62	0.21	0.13	-3.9	0	-4.15	-0.06	0.03	0.1	2.49
NN	0.35	0.69	0.19	0.12	-3.18	0	-4.04	-0.06	0.02	0.09	5.09
RF	0.22	0.71	0.19	0.12	-3.45	0.01	-4.21	-0.05	0.03	0.1	2
XGB	0.3	0.71	0.19	0.12	-2.93	0.01	-3.96	-0.06	0.02	0.09	2.47

4.1. Multiple linear regression

Multiple linear regression modeling is widely used in several fields because of its simplicity and its ease of use and interpretation. While it is more appropriate when the variables are independent, we tested it here to serve as a performance baseline and compare it to the three other more sophisticated ML models. Training is performed on half of randomly sampled D_{Train} data to keep the comparison fair with respect to the other ML models, which also use sampled data during a tuning phase.

If F is the function that maps the input features X to the output denoted by y , our goal is to find a function G that approximates F using a set of known solutions $\{X_i, y_i\}_i^N$. Typically, the function is learned successively by minimizing the expected value of a loss function $L(y, G(X))$. The residuals are assumed to be normally distributed with constant variance. Given these assumptions and the fact that F is linear, we approximate it as the following equation :

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon \quad (IV.1)$$

where ε represents the error associated with the approximation, β_0 is the intercept, β_i is the regression coefficient corresponding to the feature X_i , where i is the feature order, and n the total number of the input features, which is 96 in our case.

The cost function L (Equations (IV.2 and (IV.3) is below, which represents the residuals from the model and the observation is solved using the least squares method.

$$L = \sum_{j=1}^N (y_j - y'_j)^2 \quad (\text{IV.2})$$

$$= \sum_{j=1}^N (y_j - \beta_0 - \sum_{i=1}^n \beta_i X_{i,j})^2 \quad (\text{IV.3})$$

where N is the number of the observations available, and y'_i is the prediction of the multiple linear regression model for the feature of order i . For this work, we use its implementation in the sklearn v1.0.2 library [236].

4.2. Feed-forward neural networks

Neural networks regression is a modeling technique that is increasingly being used in many fields thanks to the advancement of big data and parallelized hardware. The principle is to use mathematical function (equation (IV.4)) represented in what is called perceptrons (inspired by the biological neuron) stacked in layers (Figure IV-7) to approximate a nonlinear relationship [237], b is a bias added to the input weighted sum, the weight w_i of each perception input i is learned from a reference dataset by gradient backpropagation [233,238,239]. The technique can model non-linear relationships by using some activation functions (σ) [240]. Optimizing multidimensional functions with neural networks is hard because of the proliferation of saddle points [241], but stochastic gradient descent algorithms are good at finding the global minimum [242–244].

$$y = \sigma \left(\sum_{i=1}^n x_i w_i + b \right) \quad (\text{IV.4})$$

Neural network regression is a modeling technique that is increasingly being used in many fields, thanks to the advancement of big data and parallelized hardware. The principle is to use perceptrons that are stacked in layers to approximate a non-linear relationship [237],

where the weight of each perceptron is learned from a reference dataset by gradient backpropagation [233,238,239]. The technique can model non-linear relationships by using some activation functions [240]. Optimizing multi-dimensional functions with neural networks is difficult because of the proliferation of saddle points [241], but stochastic gradient descent algorithms are good at finding the global minimum [242–244].

Training of the NN AOD bias corrector is performed using the Tensorflow library v2.5.0 [245]. First, we search for a suitable network architecture for the problem by running 100 trials of feed-forward neural networks with dense layers.

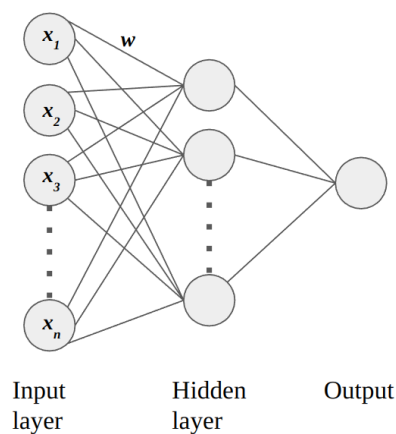


Figure IV-7 A Neural network graph representaion.

The architectures are sampled from an arbitrary reasonable hyperparameter search space (Table IV-2). We use the random search tool of the Keras-tuner 1.0.3 library [246] to find the best network. All the tested networks use batch-normalized input feed [247] and a rectified linear unit [248] as the activation function. The use of a dropout layer [249,250] degrades the accuracy so we do not use it here. The weights of the network are optimized using the Adam optimizer [251], which is a variant of stochastic gradient descent, with respect to the mean squared error. The training is performed for 100 epochs, we set early stopping in the conditions when the validation loss either exceeds 0.1 or stabilizes at 0.01. We notice that increasing in the number of neurons (perceptrons) does not improve the accuracy of the estimator. Knowing that the architecture search is not exhaustive, we manage to obtain the same performance as the best randomly generated NN model using a bottleneck architecture, with less complexity and fewer parameters (Table IV-3). The network has 3943 trainable parameters.

Table IV-2 The network architecture search space.

Hyperparameter	Minimum value	Maximum value	Increment
Number of layers	10	50	10
Number of perceptrons	10	50	10
Batch size	5000	20000	1000

Table IV-3 The best architecture for the NN regressor.

Layer type	Output shape	Number of parameters
Batch normalization	96	384
Dense	30	2910
Dense	20	620
Dense	10	210
Dense	1	11

4.3. Random forest

A random forest regressor [252] is an ensemble model that is widely used in both classification and regression problems. Each member of the ensemble (Figure IV-8), called a tree [253], is constructed using different randomized parameters. The final prediction is the average of the predictions of all the trees (forest).

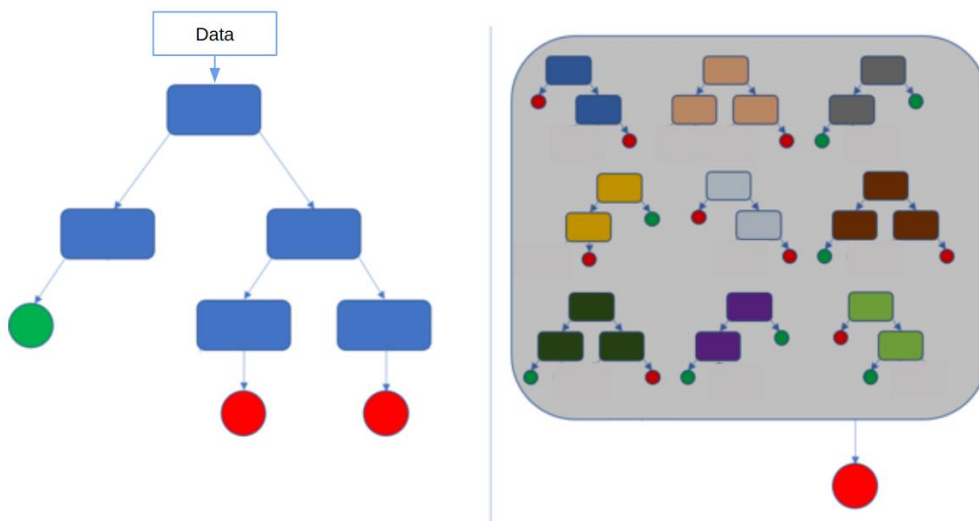


Figure IV-8 Illustration of a decision tree (left) and the random forest regressor (right panel). Boxes represent split nodes, and circles represent the prediction, adapted from [254].

The model is tuned for three key hyperparameters: number of estimators (trees), minimum samples in a leaf, and maximum number of features. The tuning is performed using the brute-force search method with two cross-fold validations on 10,000 randomly selected ground pixel data from D_{Train} . Using only a subset of the available training dataset speeds up the search process and saves memory. We find out that the best performing RF candidate has a maximum of 20 features, a minimum of 4 samples in leaf, and 100 estimators. We then train the best model using half of D_{Train} . The training was performed using scikit-learn v1.0.2.

4.4. Gradient boosting

A gradient boosting regressor [255,256] is also a tree-based modeling technique, but it constructs trees successively in such a way that each tree minimizes the residuals of the preceding trees. The final estimate is the sum of predictions of all the trees.

We use the Tree-based Pipeline Optimization Tool (TPOT) library v0.11.7. It is a genetic programming-based automated ML system that optimizes a set of feature preprocessors and ML models to build a better data processing pipeline for supervised tasks [257]. We let TPOT determine the best pipeline for correcting the CHIMERE AOD using 10,000 randomly selected data from D_{Train} . The number of generations is set empirically to 20 and the population size is set to 20, because further increasing these numbers slows down the search process. The best pipeline suggested by TPOT added the zeros and non-zero counts to the input features, and uses the XGBoost model (XGB) [256], which is a variation of the gradient boost model that uses L1 and L2 regularization. The XGB post-processor consists of 100 estimators, with a maximum depth of 9 and a minimum child weight of 17 (see the documentation [258] for more details on the meaning of these hyperparameters). The best pipeline proposed by TPOT is retrained using half of the D_{Train} dataset and is randomly sampled.

5. Results and discussion

In this section, we analyze the performance of the trained AOD bias correctors on D_{Test} against the MODIS observations. First, we detail a case study (Section 5.1a) where we investigate the influence of the input features on the predicted AOD. We discuss the statistical performance of the ML models in Section 5.1b, then present a use case of the AOD correction at a time of the day that is different from the one used for training the ML models (Section 5.1c). Finally, we present a comparison against ground-based AERONET measurements (Section 5.2).

5.1. Comparison against independent MODIS observations

a. Case study of 30 September 2021

The bias-corrected AODs predicted with the four ML models (Figure IV-9c–f) are compared with the MODIS AOD measurements (Figure IV-9a) and with the CHIMERE raw (uncorrected) AOD simulations (Figure IV-9b) for the case study of 30 September 2021. This event is characterized by a large desert dust outbreak originating from the southern border of the Sahara (as typically observed during summertime in this region, e.g., [165]), and anthropogenic pollution fine particles located over the southeastern part of the Arabian Peninsula. This is consistently shown by CHIMERE simulation and also by the aerosol Ångström exponent values derived from MODIS AQUA Deep Blue (maps available in the Worldview portal [259]). As expected, the latter are close to zero over the Sahara, which is typical for coarse particles such as desert dust, but they are significantly higher (about 1.8) over the AOD patches over the Arabian Peninsula, thus suggesting the presence of fine particles as those of anthropogenic origin.

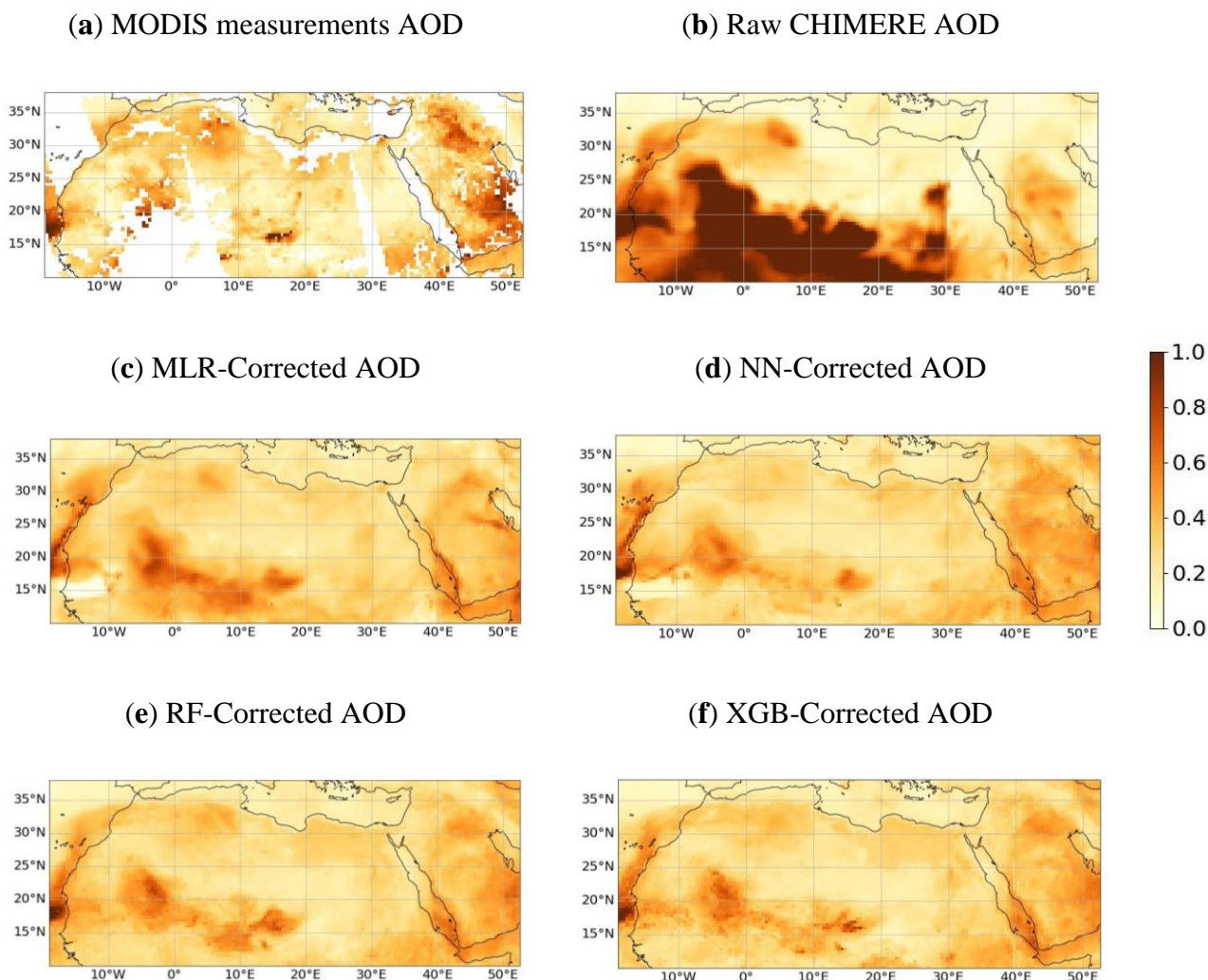


Figure IV-9 Horizontal distribution of AOD at 550 nm on 30 September 2021 (not used for training) from (a) MODIS satellite observations, (b) raw CHIMERE simulation, and bias-corrected AODs using the (c) MLR, (d) NN, (e) RF, and (f) XGB models. The AOD predictions of the remaining days are available at the link provided in the data availability clause.

Compared to the satellite data, the corrected AOD maps predicted by the ML models provide a clear added value with their full geographic coverage. They are not affected by the cloud cover or the satellite swath size, while offering better accuracy than the raw CHIMERE simulations. The ML models successfully corrected the AOD by reducing the raw CHIMERE values in the high desert dust load regions such as the Southern, Central, and Western Sahara. Additionally, they increase the low AOD values over the northern and southern Arabian Peninsula that are underestimated by CHIMERE. This significantly improves the agreement with respect to MODIS measurements. Despite the independent pixel-by-pixel processing of

the ML approaches, the horizontal structures of the corrected AOD features are continuous and homogeneous in most of the regions. This is true for most of the dates of the D_{Test} dates. There are exceptions, such as some horizontal discontinuities seen near the southwest coast of North Africa on 30 September 2021 from the NN and MLR predictions (Figure IV-9c,d). According to MODIS observations, these discontinuities in the AOD do not reflect the natural concentration gradient found in these dust plumes. In the case of MLR (Figure IV-9c), the issue could be associated with the use of negative regression coefficients. For the NN model, the problem might be explained by the fact that multi-layer perceptrons may not extrapolate well the non-linear relationships outside the training set [260]. In fact, its low performance has also been observed in another model's inter-comparison work [219]. The XGB and RF bias correctors do not show such artifacts (Figure IV-9e,f). However, we note that XGB is slightly noisier, suggesting an overfit to the MODIS AOD pixel noise. In fact, Breiman suggests that using a random selection of features to split each node in tree-based models yields error rates that are comparable to adaptive boosting (Adaboost) but more robust to noise [252].

Additional analyses of the performance of the ML approaches are provided by feature importance values (Figure IV-10, Figure IV-11 and Figure IV-12) computed using the Shapley Additive exPlanations (SHAP) framework [261]. Shapley values are commonly used in cooperative game theory to fairly determine each player's contribution to the total payoff, taking into account all the possible coalitions of players. In our study, they allow us to investigate the source of the spatial artifacts and help us understand which input features are the most important and how they spatially affect the prediction of the ML models. These values are computed considering all the possible combinations of the input features while estimating the marginal contribution (either positive or negative) of each one of them to the model output. Input features of 3D variable fields are vertically summed up to obtain the contribution of the actual geophysical variable, rather than a unique feature (for each individual altitude). All of these calculations are quite computationally costly. Therefore, the following variable importance analysis is limited to the test case of 30 September 2021.

Figure IV-10 shows the horizontally averaged absolute contribution of each input variable for the four ML models in terms of Shapley values. It shows that a larger number of variables contribute relatively evenly to the prediction of the AOD values for the RF and XGB models than those of MLR and NN. For the latter, some variables, such as the abundance of ammonium particles and OH radicals, show almost zero importance. For all ML models, as

expected, the raw CHIMERE AOD is important for the predictions, and in the case of RF, this is also true for other variables that also represent the abundance of aerosols, such as particle mass concentrations in terms of PM_{2.5}, PM₁₀, and total dust concentration. Another important variable for all the ML models is the specific humidity, especially for the MLR model, where three humidity-related variables share half of the total prediction importance. As noted in the following paragraph, the ML models appear to use humidity to numerically adjust the background values of AODs in different ways over either dry or humid regions. In the case of RF and XGB, moderate importance is also seen for meteorological variables (such as atmospheric pressure, shortwave radiation, sensible heat flux, and water droplet precipitation) and also for atmospheric species related to anthropogenic activities and chemically active species, such as nitrate, sulfate particles, and OH. The latter suggests a potential ability of the two ML models mentioned above to account for contrasting meteorological conditions on aerosol evolution (e.g., clear skies during high-pressure situations or the influence of low-pressure systems with increasing cloudiness) and the influence of anthropogenic emissions on particle abundances (the geographical influence of the latter is discussed in the following paragraphs).

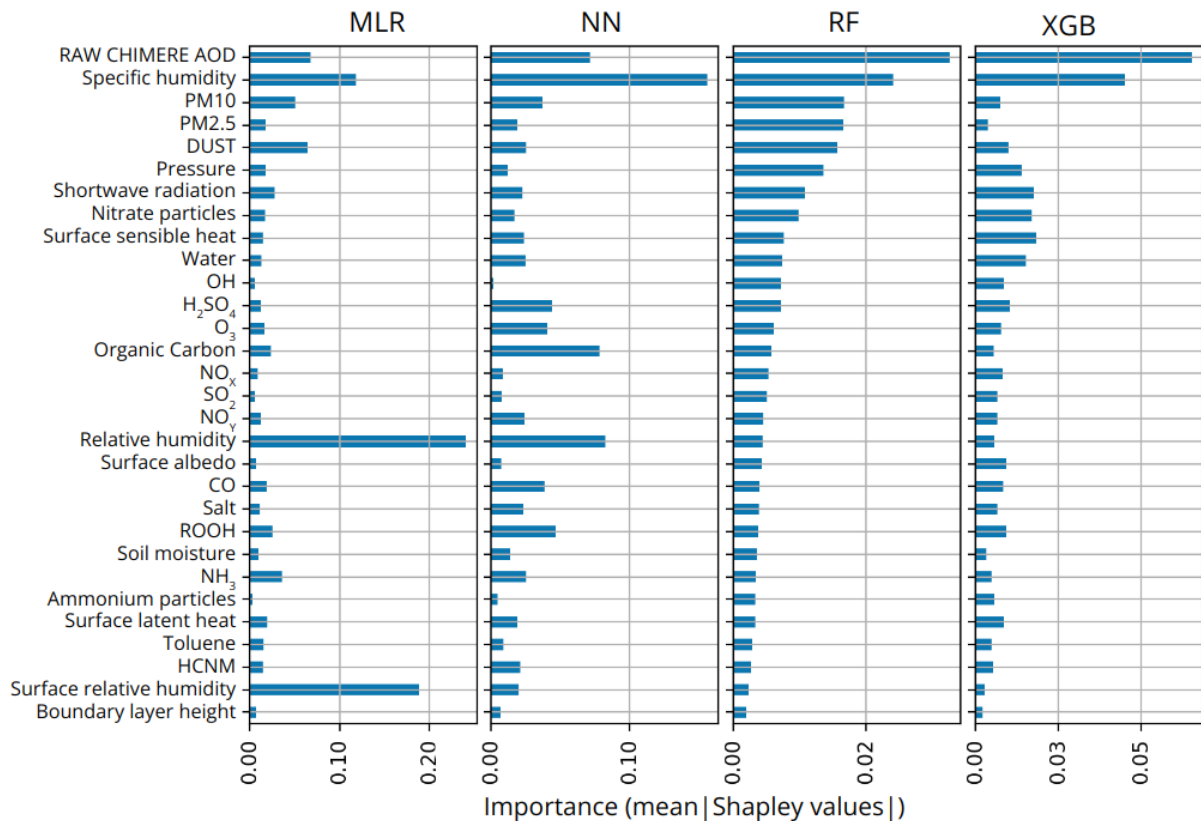


Figure IV-10 Mean absolute Shapley values for the input variables of the ML models for the 30 September 2021.

The horizontal distribution of the Shapley values is very valuable for explaining the role of each input feature in correcting the AOD (see Figure IV-11 and Figure IV-12 for, respectively, RF and MLR models). Shapely maps for the CHIMERE raw AODs show that this feature is used to modulate the AOD in the regions affected by the desert dust outbreak in the central Sahara (see the very intense reddish values in Figure IV-11a). A similar use is expected for the PM_{10} variable, but with some local differences (such as over northern part of the Sahara) and a more moderate intensity (Figure IV-11c). The map of Shapley values for the specific humidity shows a quite different structure, with significant and negative values north of $20^{\circ}N$. Over northeast Africa, this region is dry and shows rather low values of desert dust AODs, which are underestimated by CHIMERE (see Figure IV-9a,b). We suggest that the ML models may use these co-located conditions as a corrector for these background values of desert dust AOD (see the correction of this bias in Figure IV-9c–f).

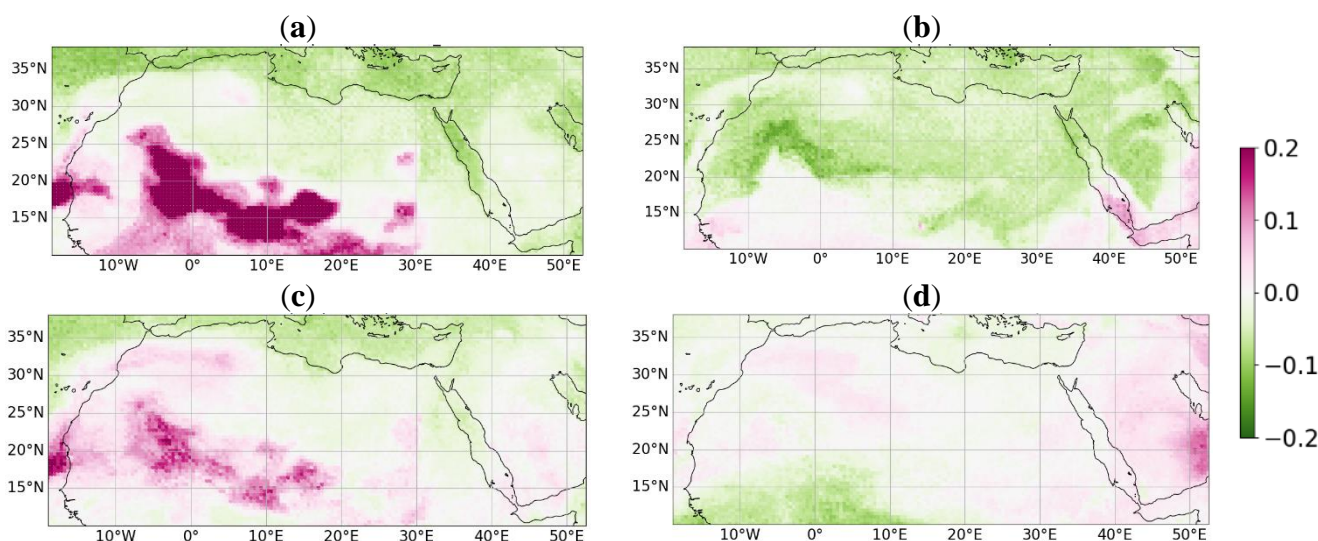


Figure IV-11 Shapley values for (a) raw CHIMERE AOD, (b) specific humidity, (c) PM_{10} , and (d) nitrate particles for the RF model for the 30 September 2021.

Another very interesting aspect is highlighted by the Shapley values for nitrate particles (Figure IV-11d). We notice that their most important influence is located over the Arabian Peninsula, collocated with the anthropogenic particle event prevailing over this region. Interestingly, we clearly see that the ML models (especially RF and XGB, Figure IV-9c–f) can adequately correct this negative bias of the raw simulation’s AOD, which is related to particles of a different origin (anthropogenic activity) as over North Africa, and appropriately use the nitrate particle concentration for its prediction.

Furthermore, the artifacts of sharp variations of the AOD with very low values over southwestern part of the Sahara for the MLR and NN model predictions (Figure IV-9c,d) can

also be explained by the SHAP analysis (Figure IV-12). It shows that two variables are responsible for this artifact, the hydroperoxide radical and the non-methane hydrocarbon concentrations (Figure IV-12 for MLR). The negative contribution of these two variables in the region around 10°W, 15°N dominates the overall prediction of the AOD, leading to the abnormally low values (Figure IV-9c,d). This is not observed for RF and XGB.

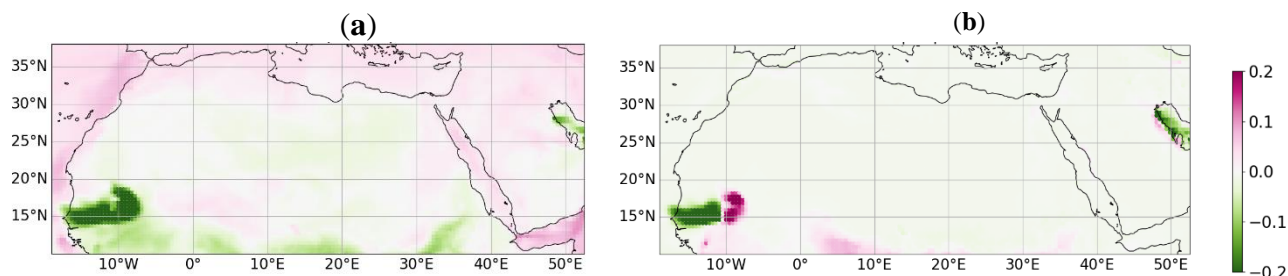


Figure IV-12 Shapley values for (a) hydroperoxide (ROOH) and (b) non-methane hydrocarbons (NMHC) for MLR model for the 30 September 2021.

b. Statistical analysis on the testing dataset

For the entire D_{Test} dataset in 2021, we compare the estimates of the raw AOD CHIMERE simulation and the inferences from the ML models with respect to the MODIS observations that are not used in the training phase (Table IV-1). The table shows that the ML models successfully correct the raw CHIMERE AODs in a statistically consistent and stable manner throughout the year 2021. They all show a comparable improvement in AOD correction, with both larger correlation coefficients r with MODIS AOD (r between 0.62 and 0.71) and smaller RMSE (from 0.19 to 0.21) and MAE. The improvement in AOD estimation is clearly evident when comparing to the raw CHIMERE AODs performance (r value of 0.56 and RMSE of 0.65). We note that the XGB and RF are the best-performing bias correctors in terms of the correlation coefficient (0.71) and RMSE (0.19). The NN model follows with the same RMSE but a slightly lower correlation coefficient (0.69). Finally, the MLR corrector offers a slightly lower correlation coefficient (0.62) and a higher RMSE (0.21).

The AOD correction biases (with respect to MODIS observations) show frequency histograms that are bell-shaped and centered around a small positive value. This suggests a small overall positive overcorrection of the raw AOD (Figure IV-13). This is consistent with the positive values of the medians (50% percentile) and the negative skewness in Table IV-1. On the other hand, the mean bias values (μ) are close to or equal to zero, which is expected since the models are trained by minimizing the mean squared error.

The model bias ranges are also quite similar. The minimum bias is rather similar for all ML models (between -4.21 and -3.96). The maximum bias values on the other hand, which correspond to the extreme underestimation of the AOD, come from the NN model with a maximum bias of 5.09 ; while the best model in terms of underestimation is the RF model with a maximum bias of 2 .

We observe that the distributions of the AOD bias corrections of CHIMERE (Figure IV-13) are skewed toward higher values. This is because there are more background AOD pixels that are underestimated by CHIMERE's raw AODs, while they are overestimated for the high AOD episodes. The overestimation occurs in most of the dust emission episodes with very large bias values, as shown in Figure IV-8, which is why the bias RMSE is high, and the median bias is positive (0.03) in Table IV-1.

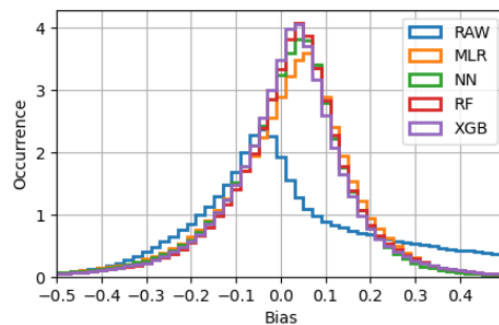


Figure IV-13 Histogram showing the frequency per 0.02 bin size bias, which is the difference between modeled (raw CHIMERE) or corrected (ML models) AOD_{550} and satellite observed AOD_{550} from MODIS for D_{Test} .

According to the RF and XGB models, the most important input feature is the CHIMERE raw AOD (Figure IV-10). Therefore, it is interesting to compare the correction versus the a priori (CHIMERE raw) AOD, which indicates the relationship between the a priori AOD in terms of intensities on the predicted AOD. The bias in the AOD correction of the different ML models is found to be only slightly dependent on the a priori AOD values (Figure IV-14). For all of them, we observe that for high a priori AODs (above a 0.5 value), the correction can be significant. While for low a priori values, the AODs correction is more limited, suggesting low bias for raw CHIMERE AODs typically under 0.3 (in the majority of the pixels). This may be attributed to the fact that there are more pixels available for training with a typically low background AOD level. The same model behavior was observed in previous research [219].

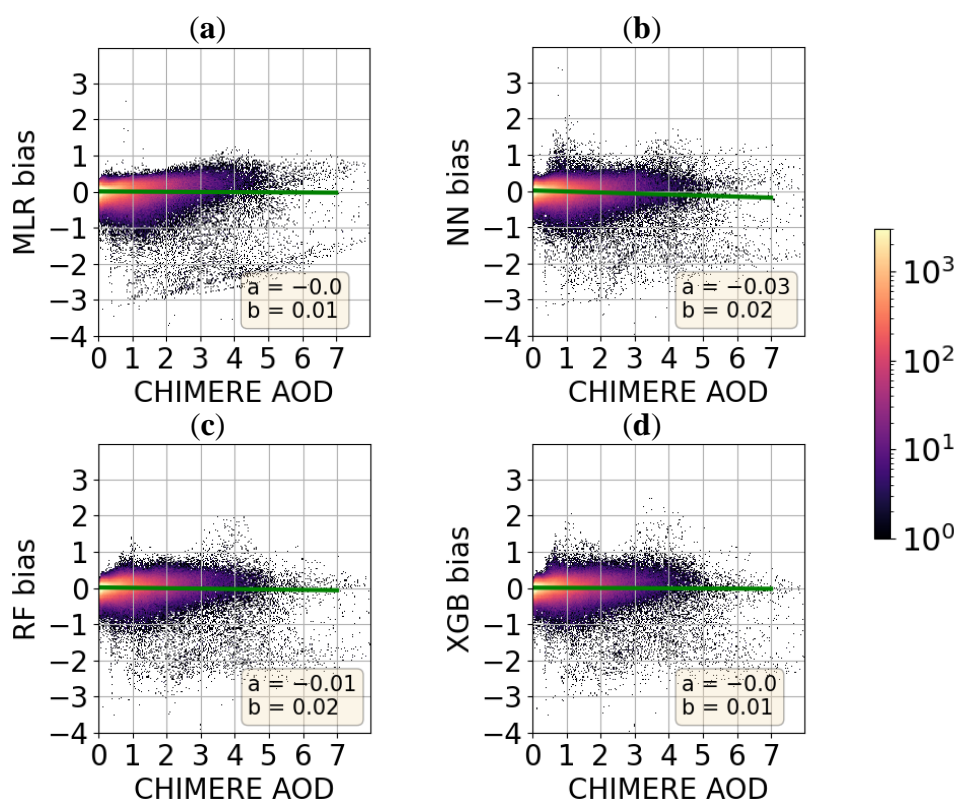


Figure IV-14 Scatterplots showing the AOD biases amplitudes corrected by the following ML models: (a) MLR, (b) NN, (c) RF, and (d) MLR, with respect to the a priori AODs (raw CHIMERE AOD₅₅₀ estimates) for D_{Test}. The green line represents the linear fit with a slope a and the intercept b . Colormap represents the occurrence.

The analysis of the predicted AOD of the different ML models varies according to the AOD level (Figure IV-15). For low AOD levels (below 1), the models mostly correct the AOD of CHIMERE to an acceptable degree (yellow colors), with a slightly insufficient correction for higher AOD levels. The MLR model (Figure IV-15a) infers negative AODs for some pixels. This is an artifact (see also Figure IV-9c) that is associated with the use of negative regression coefficients, which improves the overall performance (r) of the MLR model. In these terms, the best performing ML model is the XGB (Figure IV-15d) with a linear fit slope of 0.52, followed by the RF model.

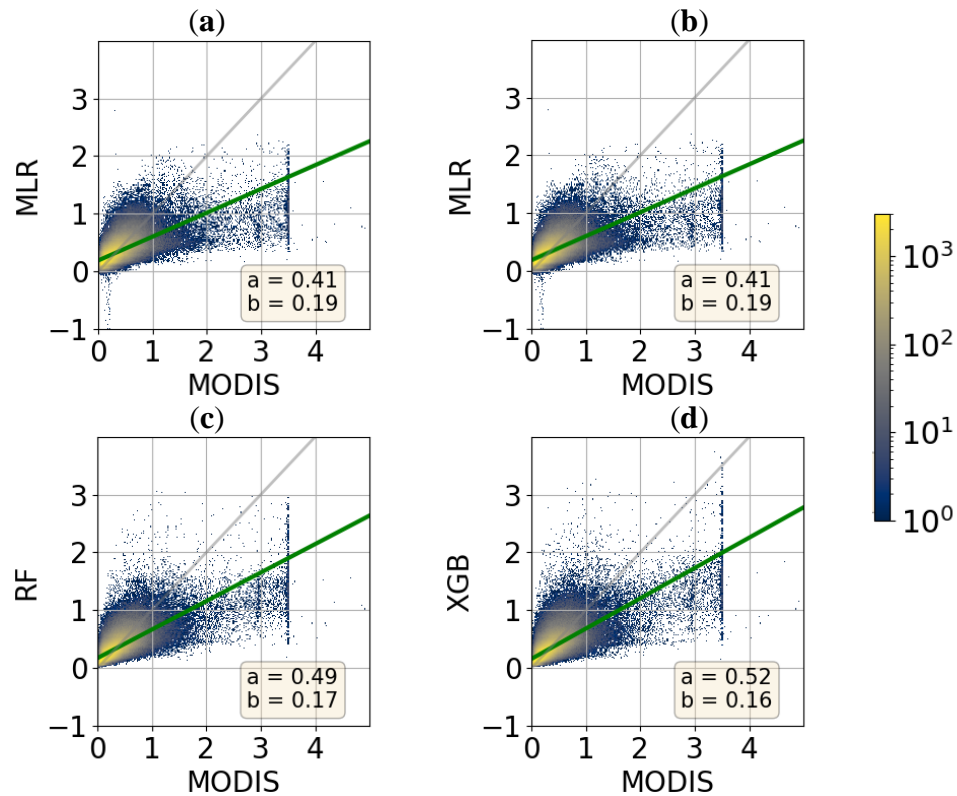


Figure IV-15 Scatterplots of corrected AOD₅₅₀ inferred by the (a) MLR, (b) NN, (c) RF, and (d) MLR models, with respect to AOD measurements from MODIS for D_{Test} . The green line represents the linear fit with a slope a and the intercept b . The gray lines represent the lines $y = x$. Colormap represents the occurrence. The apparent limit of the MODIS AOD at the value 3.5 is an arbitrary upper limit fixed by the original Deep Blue retrieval algorithm.

Figure IV-16 shows an evaluation of the performance of the four ML-based bias correctors in terms of the median bias of the AOD correction over the entire test dataset, D_{Test} . The median bias does not exceed 0.3 in most of the domain, and it is significantly smaller than that of the raw CHIMERE simulations (Figure IV-3). Similar biases are seen for the four correctors, with some positive and some negative biases for several regions characterized by high annual mean AOD values (Figure IV-4b), and they are smaller in absolute terms for RF and XGB, especially over the Arabian Peninsula and the southwestern part of the Sahara. A common behavior of all the ML models is that they strongly underestimate the AOD over a zone (10° to 20°E, 14° to 20°N) corresponding to the Bodélé Depression (Figure IV-16).

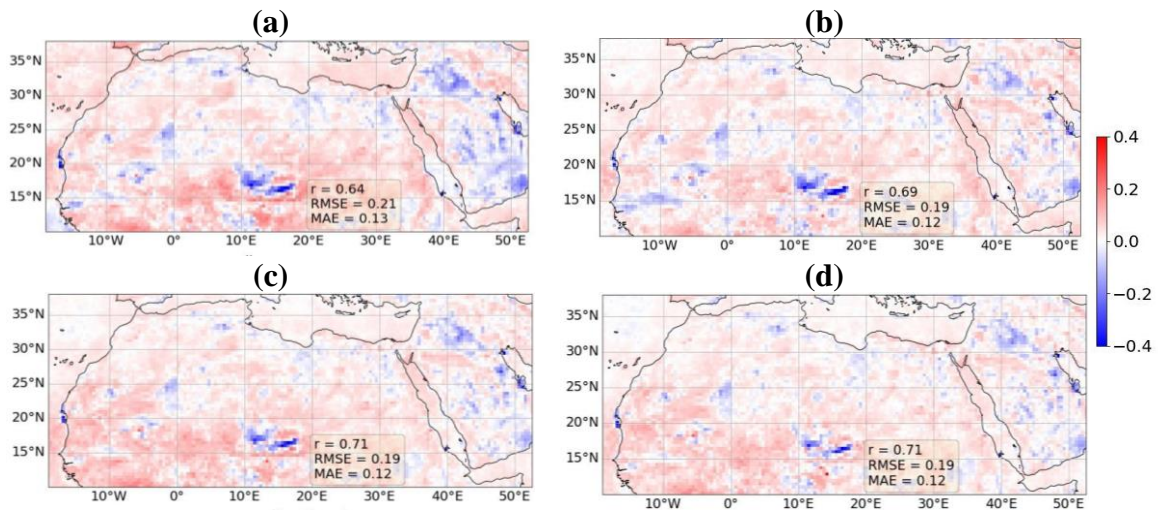


Figure IV-16 Median values of the difference between CHIMERE bias corrected AOD₅₅₀ and MODIS AOD₅₅₀ for the dates of D_{Test} for the (a) MLR, (b) NN, (c) RF, and (d) XGB ML models.

The Bodélé Depression region is known to be a major source of desert dust [232]. The region is part of the former paleolake Megachad, which was the largest lake on the planet 7000 years ago [262]. Mineral dust emissions from the Bodélé Depression region are generally relatively fine-grained coarse dusts dominated by quartz, with admixtures of clay minerals and Fe oxyhydrates [263]. Therefore, they are brighter (higher reflectance) and whiter (flat reflectance spectrum), compared to the rest of the Saharan plumes [264] (Figure IV-4a). Indeed, Algerian Saharan dust [265] and the Bodélé Depression region dust [2] have different refractive indices. In contrast to the CHIMERE CTM model, the aerosol retrieval algorithm Deep Blue uses two different single scattering albedos for the dust AOD retrieval [264], one for the redder dust, and one for the whiter dust. Furthermore, an upper limit of 3.5 is set for the AOD values derived from the MODIS AQUA operational algorithm [220].

These remarks suggest that potential improvements in the representation of the particle optical properties and the size within CHIMERE, along with the accuracy of the near-surface wind speeds responsible for the dust uplift, could reduce the bias in the simulated AOD in North Africa. However, this is beyond the scope of this paper.

c. Prediction of bias corrected AODs at the different daytimes

The AOD correction models are trained using MODIS AQUA satellite observations, which has an overpass time of (13:30 LT). It is necessary to ensure that the models are temporally stable and usable outside the AQUA overpass time. For this purpose, we correct the raw CHIMERE AOD outputs at 10:00 LT using the RF model, and then compare the inferred AODs to MODIS observations from the Terra satellite, which has a local overpass time of 10:30 LT. Thus, the daytime of the training database (13:30 LT) is different from that of the test (10:30 LT). We chose only the RF model to perform the upcoming analysis because the statistical performance of the other ML models is similar. The comparison shows that improvement in correlation and residuals is comparable to the corrected AODs at 13:00 LT. From Table IV-4, we can see that the residuals (RMSE, MAE, and MB) are improved. The RMSE is reduced by 0.68% and the correlation is increased to 0.68 (from 0.52 for the raw CHIMERE AODs). This good AOD correction for a morning hour implies the possibility of using the correction method for different simulation hours of the day. However, we remain conservative in using the AOD correction method during night hours as the ML model will more often have to extrapolate unseen situations, e.g., low boundary layer heights, low temperatures, etc.

Table IV-4 Correlation and residuals of AOD estimations for the year 2021 against MODIS Terra observation. The raw AODs are CHIMERE outputs at 10:00 LT and for the corrected AOD with the RF model (trained with MODIS AQUA observations at 13:30 LT). Number of data points is 3.7 million.

	r	RMSE	MAE	MB
RAW	0.52	0.59	0.34	-0.23
RF-corrected	0.68	0.19	0.12	-0.03

5.2. Comparison with AERONET ground based measurements

The performance of the RF model together with that of MODIS is assessed against the collocated AERONET measurements at the eight stations for test dates (Table IV-5). We see an improvement in the correlation coefficient r to 0.73 (from 0.54 for the raw AOD). The residuals also improve as the RMSE and MAE decrease from 0.45 and 0.27 to about 0.16 and 0.12, respectively. The MB is significantly reduced from 0.18 to about 0.6 for the corrected AOD.

Figure IV-17 illustrates the daily evolution and temporal consistency of the performance of the RF model. It is worth noting that for visual completeness, the figure also shows comparisons for the days used in the training phase (in the gray background), but we limit the analysis to the test dates (in the white background). We chose the measurement time series from the station of Cairo for this purpose because of its good data coverage of the period and because it shows situations of positive and negative corrections of the biases, with desert dust and anthropogenic aerosol events (the same kind of figures for other sites can be found at the link in the data availability clause). A number of high AOD episodes are seen in March and April and from 24 to 29 October 2021. In these cases, the raw AODs simulated by CHIMERE (orange lines) are clearly overestimated with respect to the AERONET measurements (black lines). The ML-based corrector reduces the AOD values (blue lines) to clearly approach the sun photometer measurements, while still depicting the AOD peaks.

On the other hand, CHIMERE simulations underestimate the AOD during the periods of 20–27 June and 21 to 30 September. These periods are characterized by background anthropogenic pollution, as suggested by co-located relatively high aerosol Angstrom exponents (around 1.8) measured by MODIS AQUA (seen in the Worldview portal, not shown). In this case, the ML-corrected AODs are successfully improved to reach values that are closer to both AERONET and MODIS measurements (red triangles). Moreover, in cases of rather correct raw CHIMERE AODs (e.g., on 20–21 May), the ML-derived bias corrector does not modify the AOD, which remains around 0.24.

Table IV-5 Correlation and residuals of 330 AOD estimations on testing dates that are collocated with the 8 AERONET stations.

	r	RMSE	MAE	MB
MODIS	0.85	0.12	0.09	0.03
RAW	0.54	0.45	0.27	0.18
RF	0.73	0.16	0.12	0.06

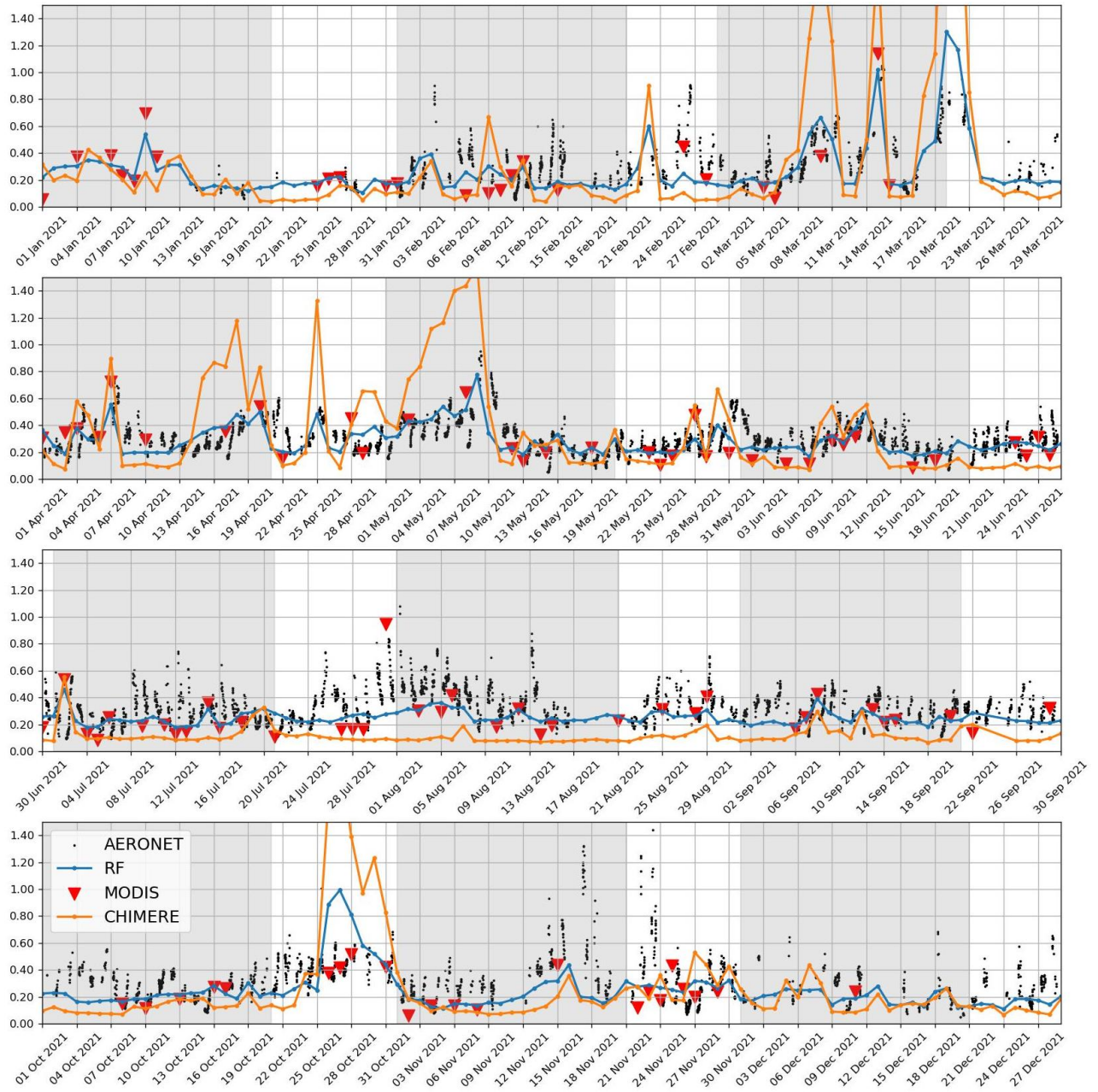


Figure IV-17 Time series of AOD₅₅₀ in Cairo during the year 2021 from AERONET ground-based sun photometer measurements, raw CHIMERE simulations, ML-corrected AODs, and MODIS. The gray areas represent dates used for training the models.

6. Conclusions

In this work, we develop a new ML-based model to significantly improve the accuracy of AOD maps, initially derived from CHIMERE chemistry transport model simulations. The ML model is trained with MODIS satellite measurements, which are not used in the inference stage. This approach provides full-coverage daily maps of AODs over North Africa and the Arabian Peninsula with a clearly better agreement with respect to satellite and ground-based

observations as compared to AODs from raw, uncorrected CHIMERE simulations. The AODs corrected by the ML models show substantially higher correlations and lower errors (RMSE and MAE) than the raw AOD simulations, as compared to MODIS satellite measurements. The RMSE of the AODs is reduced from 0.65 (raw CHIMERE) to 0.19 (RF bias corrector), and the correlation coefficient r is increased from 0.56 to 0.71 (respectively, for raw CHIMERE and the RF corrector). The bias corrector reduces the overestimation in the AOD maps over the Saharan desert and the underestimation over the Arabian Peninsula. This is performed not only for desert dust during outbreaks and background conditions, but also for anthropogenic pollution aerosols. However, a slight overestimation of the correction is observed at low aerosol loads.

The spatially continuous AOD correction could potentially be used as a gap-filling system for the global MODIS AOD observations. Additionally, this current accomplishment suggests the possibility of using advanced retrievals of the 3D distribution of aerosols such as AEROIASI [96], and AEROS5P [139], which utilize Infrared Atmospheric Sounding Interferometer (IASI) thermal infrared measurements for coarse particles and TROPospheric Ozone Monitoring Instrument (TROPOMI) measurements for fine particles, respectively, to correct the CHIMERE vertical profile of aerosol concentrations.

Among the four ML-based bias correctors, the best results are found for the RF regressor, which gives a spatially smooth AOD in good agreement with MODIS AODs. The other bias correctors used here (MLR, XGB, and NN) also improve the accuracy of the AOD, but show spatial artifacts, some of them related to overfitting. The AOD correction was tested for a different time of the day than the one used for the training. The comparison of the CHIMERE AOD correction at 10:00 LT with that of MODIS Terra, which has a close overpass time (10:30 LT), shows that the RF model also provides good results even for this time of the day, which is different from that of the training dataset. This is a step toward validating the AOD correction for any daytime application of the CHIMERE simulation.

Evaluation of the daily consistency of the corrected AOD shows that the overestimation of peaks and the underestimation of background values at a given location (shown for Cairo) have been successfully improved, for both desert dust and anthropogenic aerosol loads. These results are consistent with analyses of the influence of the input features, showing that the RF model adequately uses different atmospheric constituents to deal with different situation (e.g., desert dust, anthropogenic pollution, and background concentrations). We also find that the

ML-derived AODs may still underestimate very high AOD values compared to MODIS observations.

In this study, ML models are implemented with CHIMERE simulations that are performed with meteorological analysis data. Therefore, the AOD correction applied to CHIMERE simulations using meteorological forecasts may be less accurate. Adjustments to the ML models may be required. Furthermore, a deep understanding of the contribution of different features will help to improve the temporal stability and reduce the spatial artifacts.

V. Conclusion générale et perspectives

Cette thèse constitue une contribution à l'amélioration de l'estimation de l'abondance des aérosols dans l'atmosphère. En effet, grâce aux données satellitaires, on a pu introduire une autre dimension (l'altitude) aux estimations passives de la distribution spatiale des aérosols. Des images tridimensionnelles des aérosols fins ont été obtenues pour la première fois utilisant un capteur passif et qui montrent l'altitude d'injection des fumées. L'algorithme décrit au Chapitre III qu'on appelle AEROS5P se base sur la technique d'inversion des observations multi-spectrales de l'instrument satellitaire TROPOMI de $3,5 \times 5,5 \text{ km}^2$ de résolution. L'approche développée est basée sur la modélisation de transfert radiatif dans l'atmosphère, et utilise un modèle microphysique des aérosols (indice de réfraction et distribution de taille). La signature spectrale dans la bande A de l'oxygène (vers 760 nm) fournit l'informations sur l'altitude, tandis que le spectre visible de 406 à 674 nm fournit des informations sur l'abondance en colonne totale des aérosols. AEROS5P nous a permis d'observer les différentes couches d'aérosols fins de feux émis lors de l'épisode d'incendies de forêts en Australie de 2019/2020 avec une résolution verticale de 1 km. Les restitutions d'AOD (distribution intégrée verticalement) des aérosols de feux montrent un bon accord avec les observations de VIIRS avec un coefficient de corrélation de 0,87 et une erreur absolue moyenne de 0,3. La comparaison des profils d'extinctions d'aérosols avec les mesures du lidar CALIOP montrent un accord acceptable, avec un faible biais moyen (185 m) et un coefficient de corrélation de 0,83 sur les altitudes moyennes. La distribution 3D de ces panaches est cohérente avec les observations de la puissance radiative détectée par MODIS et la stabilité atmosphérique pour la période d'analyse (20-24 décembre 2019). Grâce aux restitutions AEROS5P, nous avons pu constater l'altitude exceptionnel à laquelle les incendies ont injecté les fumées dans l'atmosphère atteignant les 7 km.

La méthode AEROS5P est développée exclusivement pour les aérosols de feux mais on vise à observer d'autres espèces d'aérosols, en adaptant le modèle d'aérosols supposé présent dans la scène observée, en plus, un ajustement itératif des caractéristiques des aérosols a priori pourra théoriquement améliorer la restitution. Des tests supplémentaires sur la sensibilité de la restitution par rapport aux différents paramètres sont envisagés. D'autre part, comme la restitution est assez exigeante en termes de calcul numérique, un remplacement de la simulation du transfert radiatif par un modèle rapide basé sur l'apprentissage profond (*deep learning*) ou

sur un tableau de recherches (*lookup table*) est envisagé afin d'assurer une couverture globale en temps réelle. En terme de résolution temporelle, l'adaptation de l'algorithme à d'autres capteurs satellitaires en orbites polaires ayant des heures de passages différentes serait utile pour combler le manque d'observations dans les différentes heures de la journée. Enfin, des capteurs multi-spectral en orbite géostationnaire tel que MTG (Météosat Troisième Génération), GEMS (*Geostationary Environment Monitoring Spectrometer*), et TEMPO (*Tropospheric Emissions: Monitoring of Pollution*) seront idéals pour améliorer la résolution temporelle de l'algorithme, car la fréquence d'acquisition de ces instruments est de l'ordre de l'heure.

La deuxième partie de ma thèse est focalisée sur l'amélioration de l'estimation de l'épaisseur optique d'un modèle de chimie-transport. L'inconvénient des simulations de l'AOD bien qu'elles sont mieux résolues dans l'espace et dans le temps est qu'elles sont sujet de déviation de la réalité à cause des incertitudes dans les données d'entrée et les paramètres physique du modèle. Afin d'améliorer la précision de la distribution spatiale de la simulation des aérosols, j'ai développé une correction basée sur l'intelligence artificielle (voir Chapitre IV), celle-ci améliore la qualité de l'estimation 2D de l'AOD simulée par le modèle CHIMERE dans la région du nord d'Afrique et la péninsule Arabique. J'ai testé les performances de quatre types de modèles de correction basés sur les forêts aléatoires, amplification de gradient, réseaux de neurones artificielles, et la régression linéaire multiple. Après l'entraînement des modèles sur les observations du capteur MODIS, j'ai analysé les résultats de la prédiction de l'AOD corrigée. Ces derniers montrent que les performances des quatre prédictions sont comparables, néanmoins, on constate une supériorité du modèle de forêts aléatoires quant à l'homogénéité spatiale de l'AOD prédite, suivi du modèle basé sur l'amplification de gradient. Les deux autres modèles basés sur les neurones artificiels et la régression multiple ont montré des artefacts traduits par des discontinuités occasionnelles dans les valeurs de l'AOD prédite. La comparaison de l'AOD corrigée avec les observations de MODIS qui n'ont pas été utilisées lors de l'entraînement montre une amélioration marquée par rapport aux valeurs non corrigées. En effet, pour le modèle des forêts aléatoires, l'erreur quadratique moyenne est améliorée de presque 30 % et le coefficient de corrélation est augmenté de 0,56 à 0,71. Bien que la correction a réduit drastiquement le biais sur l'AOD dans la zone du Sahara, une légère surestimation est introduite à la région de la péninsule Arabique, mais cette dernière reste largement inférieure au biais initial sur les AOD non corrigées. Une analyse de l'importance de chaque variable utilisée comme entrée pour la correction de l'AOD a été réalisé afin d'identifier la cause des

artefacts spatiaux. En effet, on constate que les particules de carbone organique étaient la cause des mal prédictions de l'AOD pour le cas d'étude analysé. En plus, cette analyse nous a rendu compte de l'importance de l'humidité de l'atmosphère quant à la prédiction de l'AOD, c'est la variable qui a une influence importante sur la prédiction, partagé avec l'AOD a priori et les PMs.

Ce travail est prometteur dans le sens de l'amélioration de la distribution verticale des aérosols simulées par CHIMERE. Ceci pourrait être réalisé en utilisant comme référence pour la supervision un produit d'observation de profil vertical tel que le lidar CALIOP, ou encore le nouveau produit développé AEROS5P. D'autre part, les corrections de l'AOD sont obtenues a posteriori et sont applicables uniquement pendant la journée, néanmoins elles sont très favorables pour être assimilées avec la simulation, ce qui rend l'estimation de CHIMERE plus précise. Enfin, la nouvelle méthode de correction de biais pourrait potentiellement être utilisée comme méthode de remplissage des lacunes pour les cartes globales d'AOD de MODIS en adaptant les entrées avec un modèle de circulation générale.

Annexe A

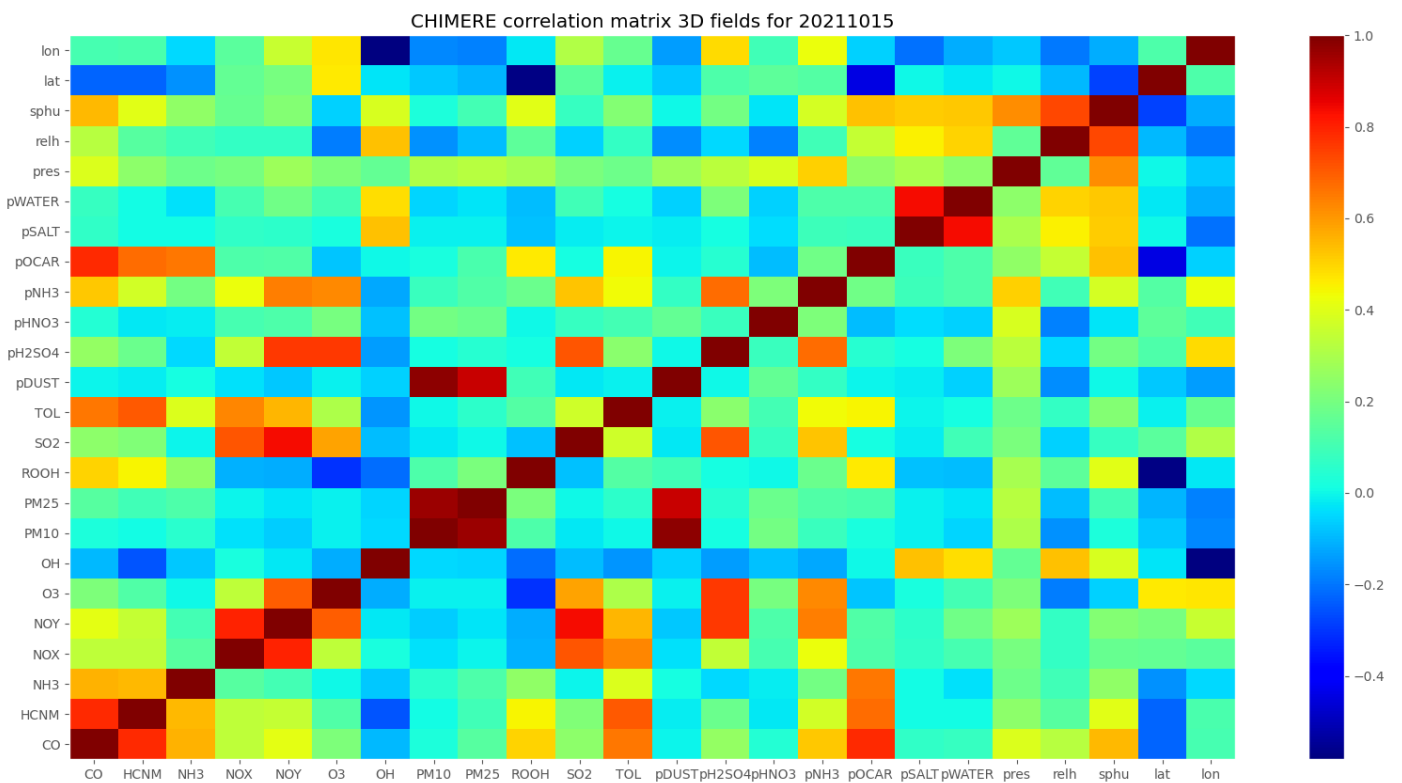
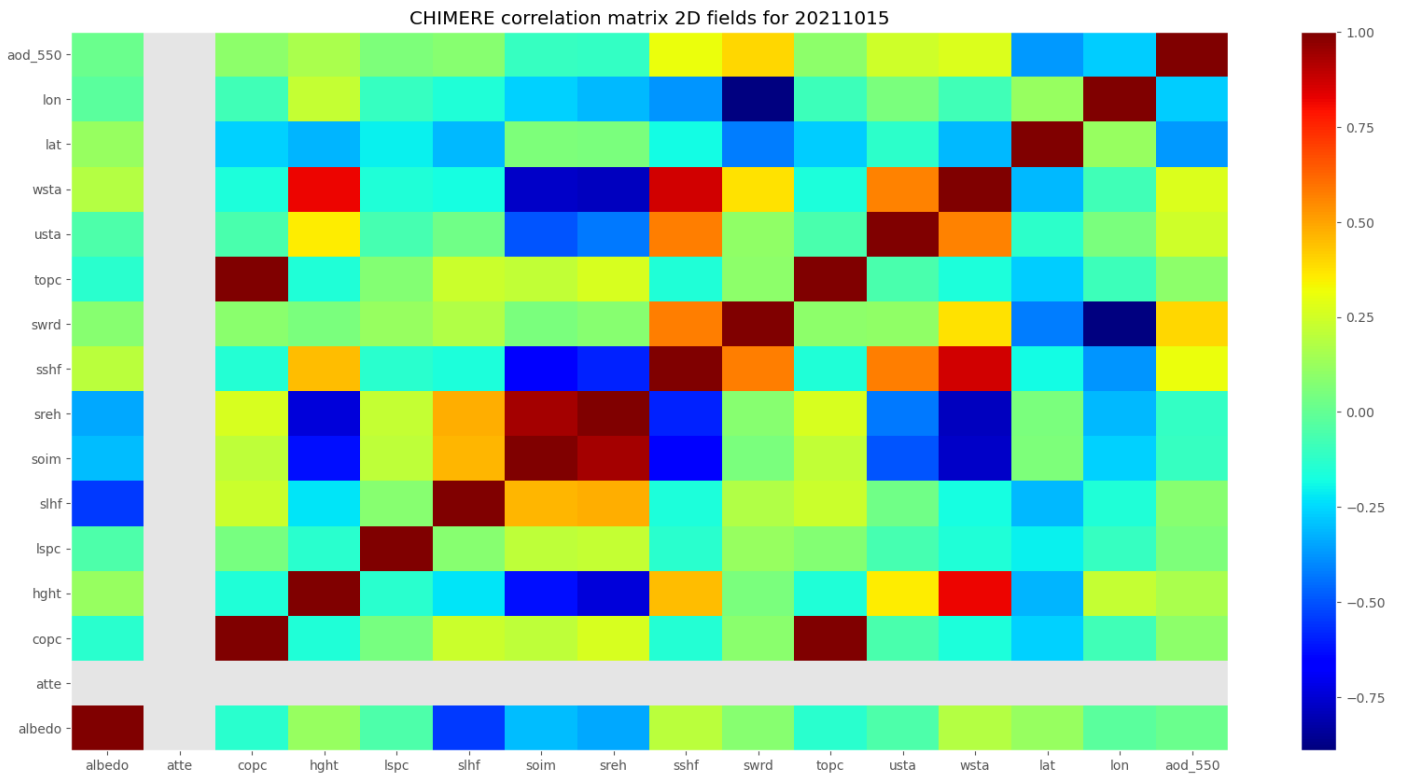
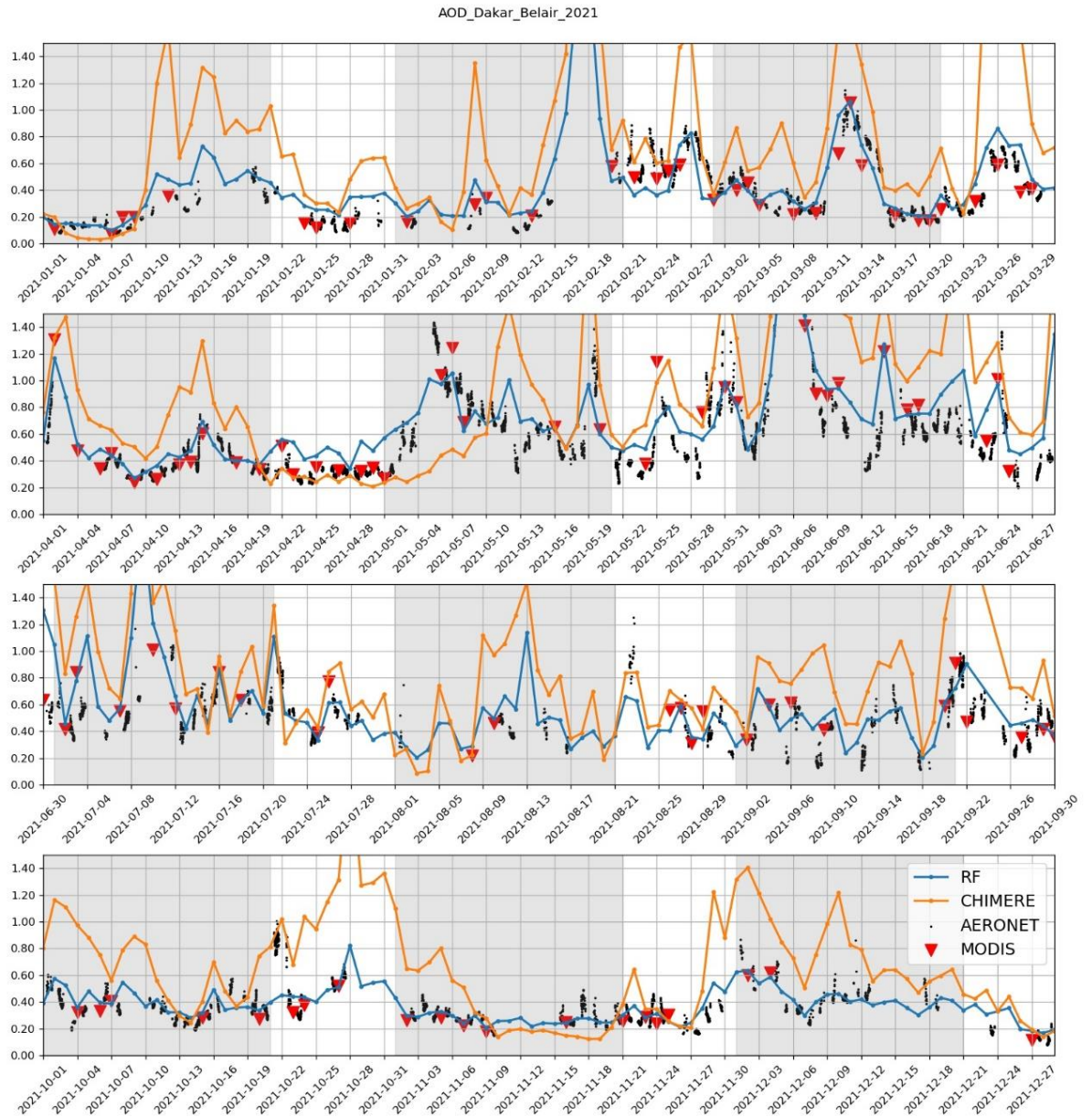


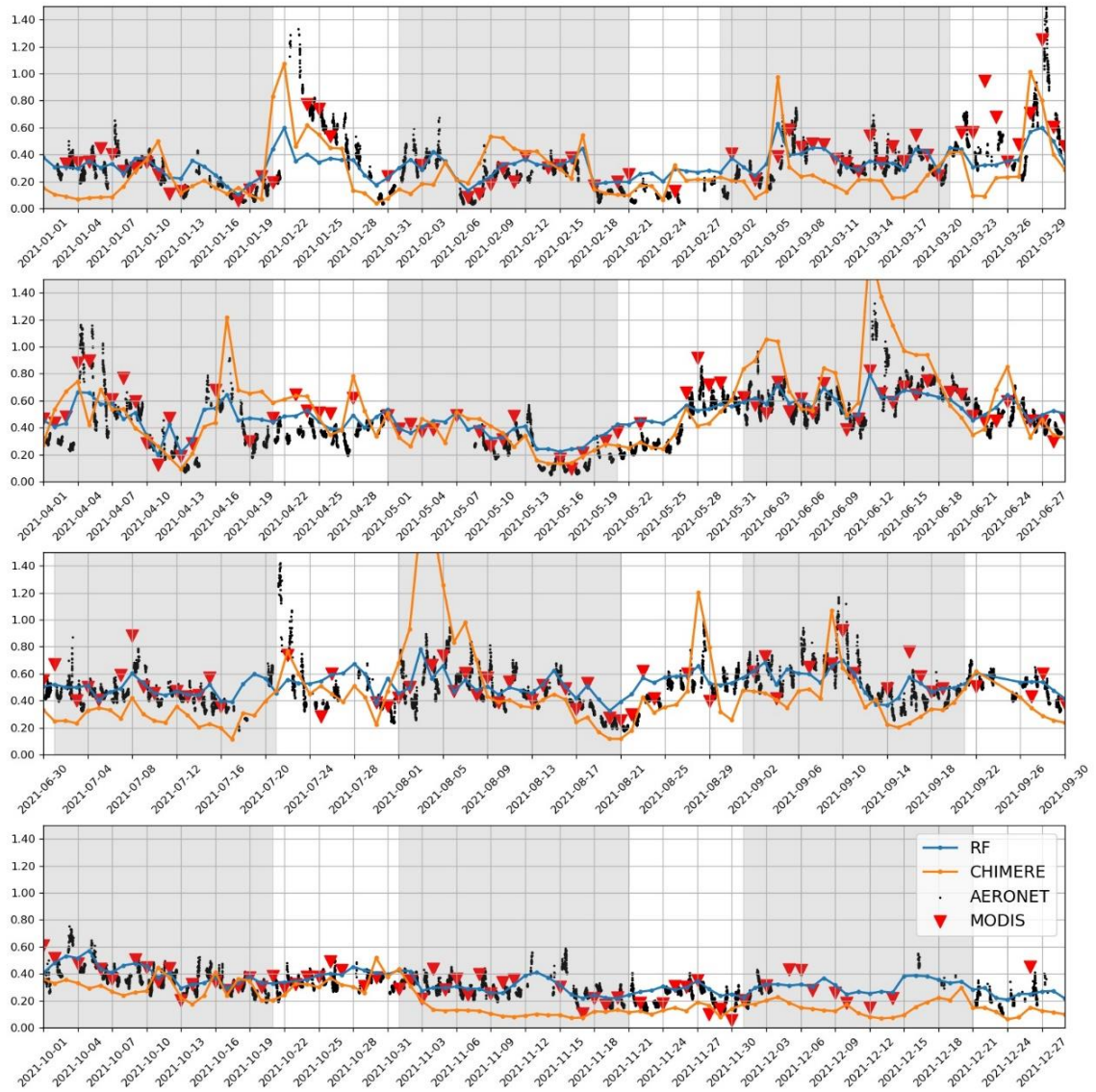
Figure S2. Correlation matrix of some variables from CHIMERE simulation output for a typical day (15 octobre 2021). The following list shows the meaning of the variable. (lon:Longitude, usta:Friction velocity, Lat:Latitude, wsta:Convection velocity, relh:Relative humidity, soim:Soil moisture, temp:Temperature, sreH:Surface relative humidity, winz:Zonal wind, pres:Pressure, winm:Meridional wind, sshf:Surface sensible heat flux, winw:Vertical wind, slhf:Surface latent heat flux, sphu:Specific humidity, copc:Convective precipitation, pNH3:Ammonium particles, lspc:Large scale precipitation, pHNO3:Nitrates particles, topc:Precipitation, TOL:Toluene, aerr:Aerodynamic resistance, albedo:Surface albedo, atte:Radiation attenuation factor, hght:Boundary layer height, swrd:Shortwave radiation flux).

Annexe B

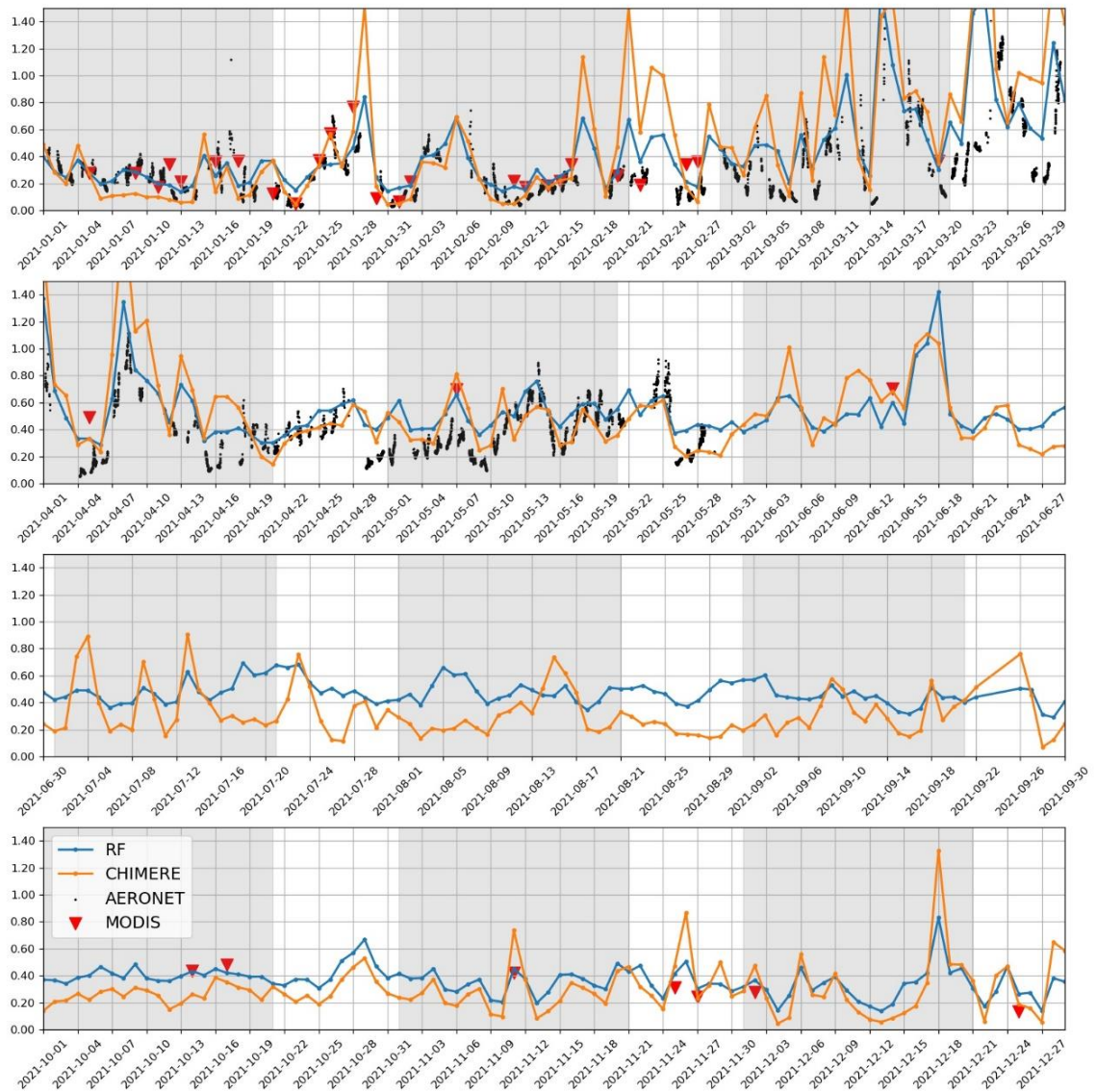
Séries temporelles de l'AOD₅₅₀ pour l'année 2021 depuis les mesures d'une station AERONET, l'instrument satellitaire MODIS, la simulation de CHIMERE non-corrigée, et corrigée avec le modèle de forêts aléatoires.



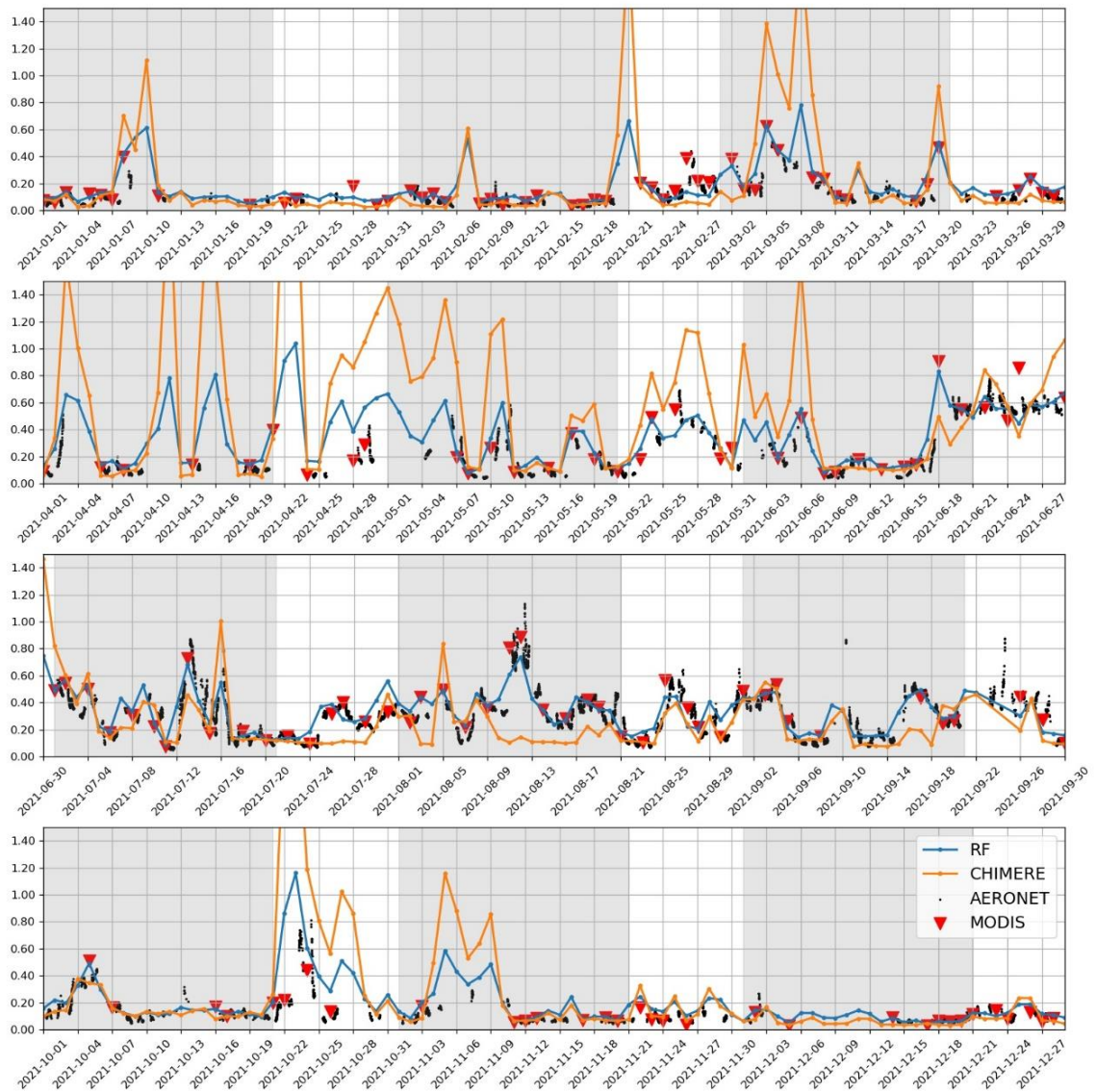
AOD_KAUST_Campus_2021



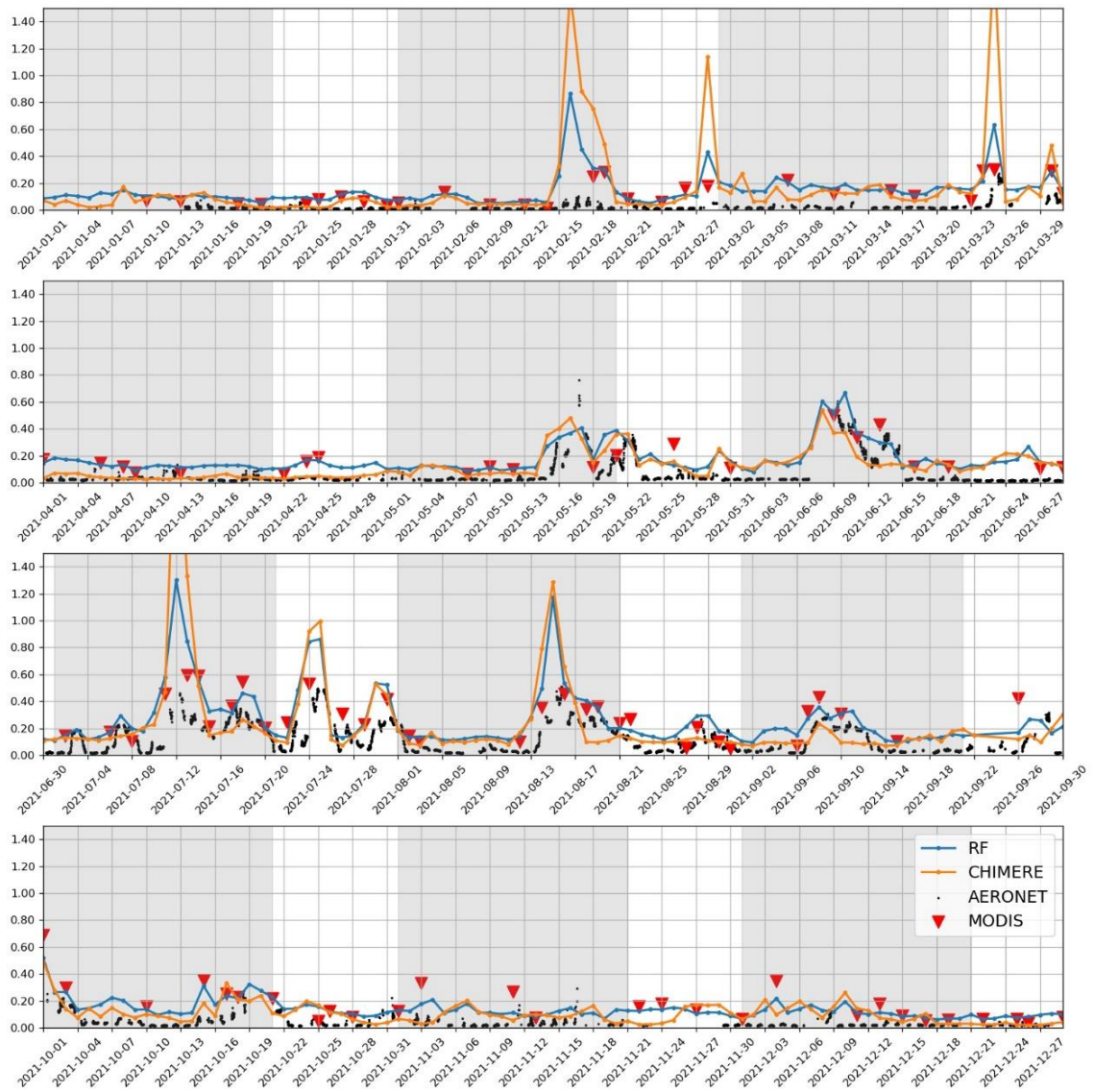
AOD_Kuwait_University_2021



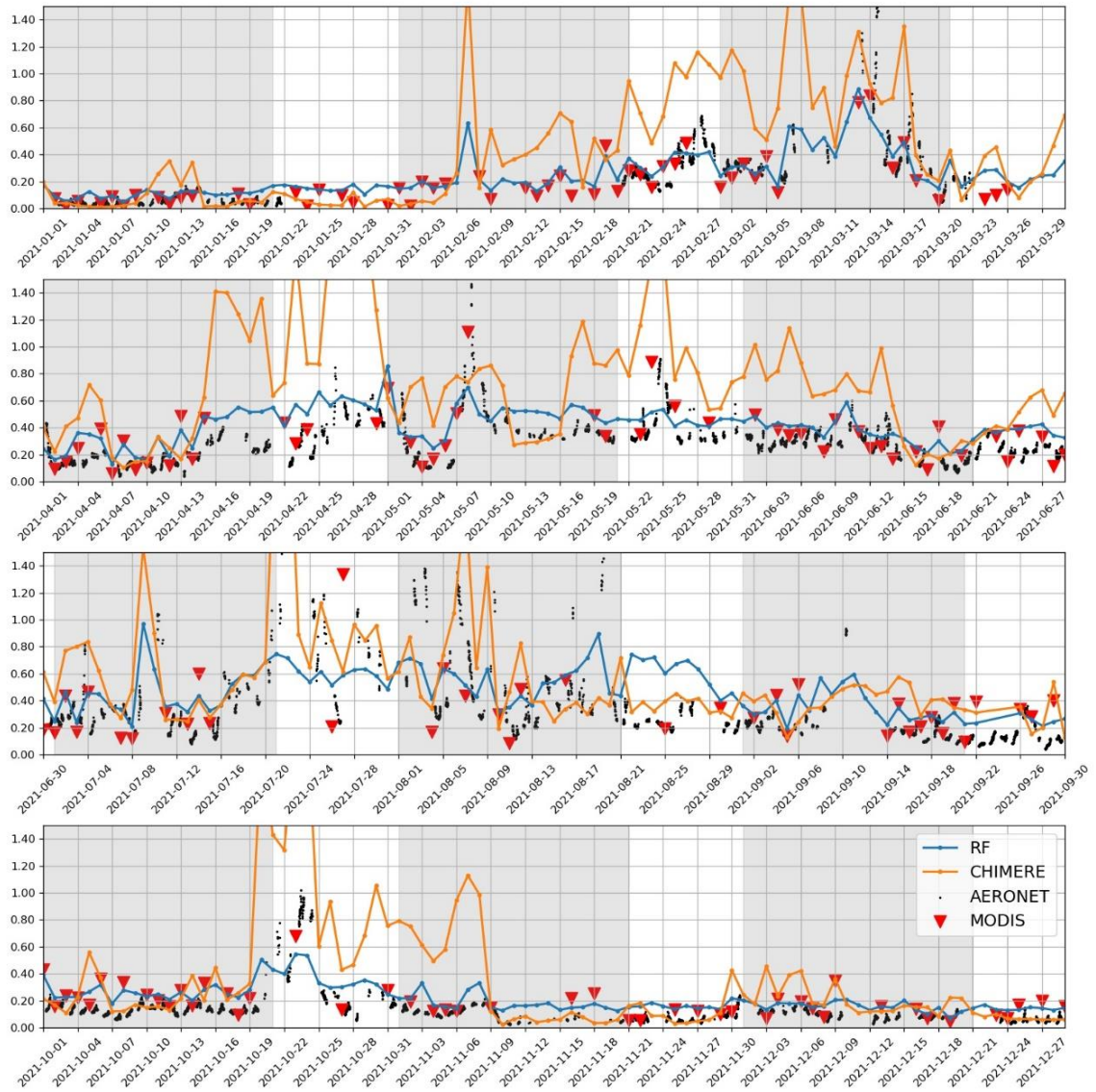
AOD_Lampedusa_2021



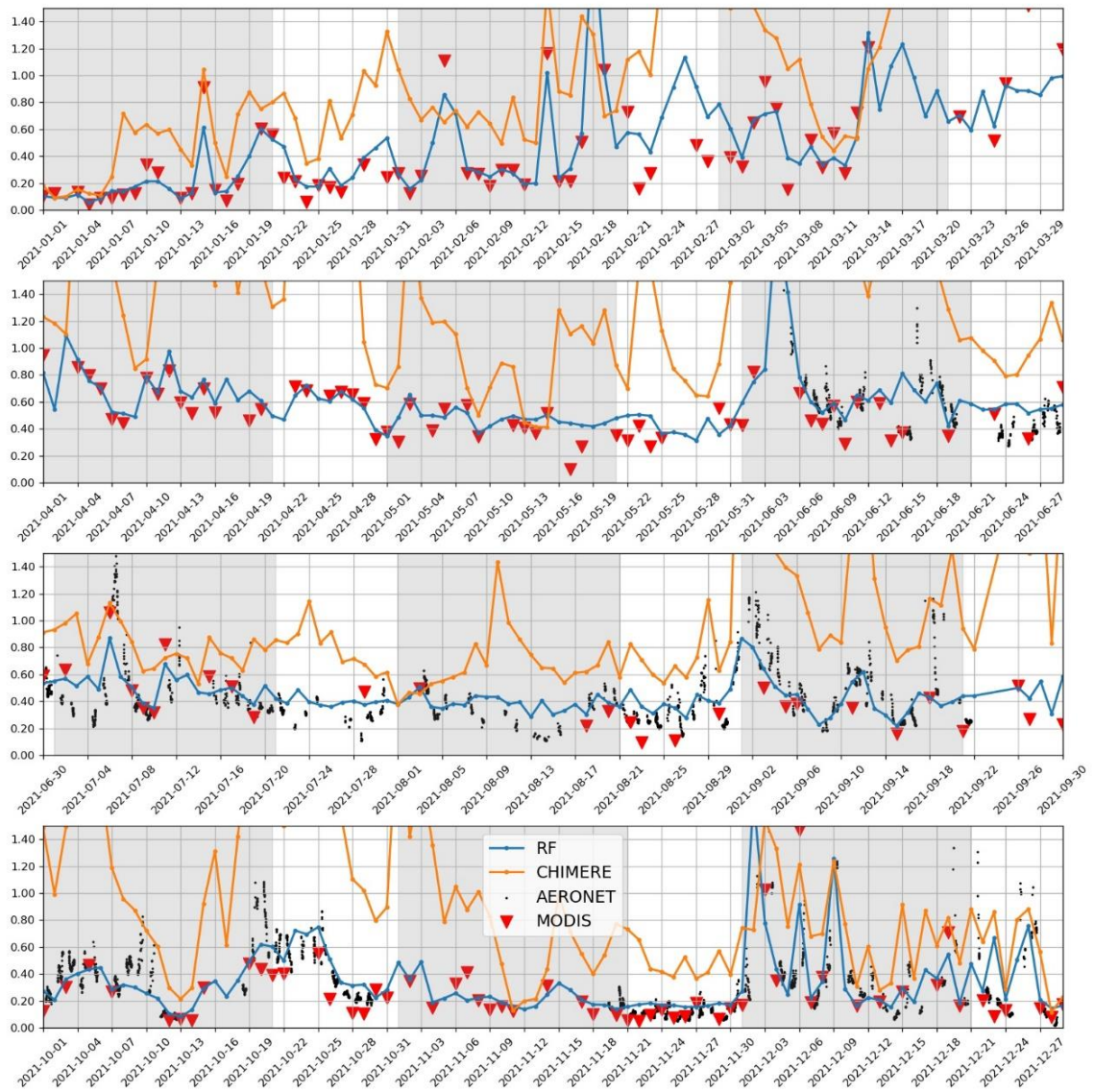
AOD_Roque_Muchachos_2021



AOD_Tamanrasset_INM_2021



AOD_Zinder_Airport_2021



Annexe C

Participation aux formations et aux évènements

- Formations

Valorisation professionnelle et création d'entreprises/projets innovants (07 février 2022 - 11 février 2022). Formation à distance via meet.jit.si. 24 heures.

Travailler dans le conseil après son doctorat. Formation à distance via zoom (17 mars 2021). 14 heures.

Ist join training course on atmospheric composition (04 novembre 2019 - 8 novembre 2019) ESA/EUMETSAT/CAMS Cluj-Napoca (Roumanie). 40 heures.

CHIMERE 2021 Training Course (08 mars 2021). Laboratoire de Météorologie Dynamique, Paris. Formation à distance. 12 heures.

Intégrité scientifique et éthique de la recherche (09 juin 2021 au 7 juin 2021). Formation à distance. 6 heures.

URFIST - Règles et usages de publication d'un article scientifique en sciences, techniques et médecine (05 février 2020 - 5 février 2020). Paris. 6 heures.

- Conférences

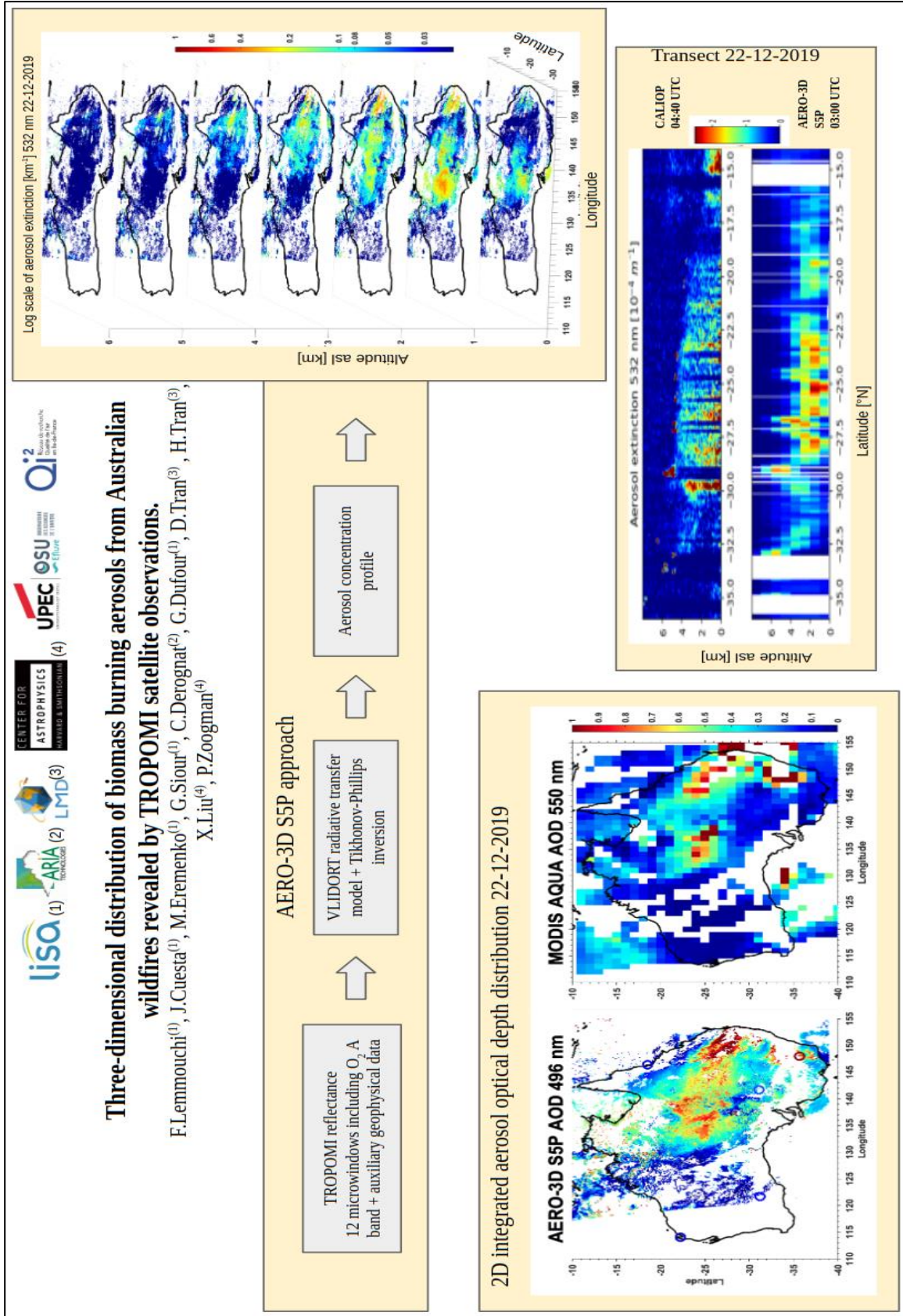
Présentation orale à la conférence *EUMETSAT 2022*, Bruxelles, Belgique.

Présentation orale à la conférence *Living Planet Symposium 2022*, Bonn, Allemagne.

Journée scientifique ED SIE. Présentation de poster - 29 JUIN 2022 (29 juin 2022 - 29 juin 2022) Champs-sur-Marne. 4 heures.

Présentation orale à la conférence *European Geoscience Union 2021* (vEGU21), En ligne à cause de la pandémie.

Présentation vEGU21



Poster journée scientifique ED SIE



Machine learning-based constraint of the aerosol optical depth maps simulated by the CHIMERE chemistry-transport model

Farouk Lemmouchi¹, Juan Cuesta¹, Mathieu Lachatre², Julien Brajard³, Adriana Coman¹, Claude derognat², and Matthias Beekmann¹

¹ LISA, UMR 7583 CNRS, Université Paris-Est Créteil, France

² ARIA Technologies, Boulogne-Billancourt, France

³ Nansen Environmental and Remote Sensing Center (NERSC), Bergen, Norway

Introduction

Aerosols are suspended particles in the air like salts, dust, soots, etc with a radius size usually smaller than 10 μm . They are **harmful** for health and they can be transported several thousands of km.

The abundance of aerosols vertically integrated in the atmosphere can be estimated by the **Aerosol Optical Depth (AOD)**. We can estimate the AOD by **remote aerosol retrieval algorithms** e.g. MODIS Dark Target/Deep Blue [2], or by **chemistry-transport modeling (CTM)** e.g. CHIMERE [1].

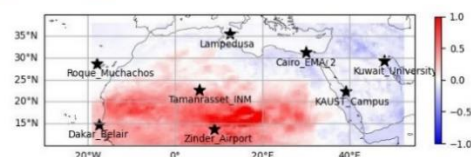


Figure.1 CHIMERE AOD simulation bias compared to MODIS Aqua satellite observations. The stars show AERONET stations location used in the comparison in Figure 2.

Materials and method

The Goal is to develop a post processor of CHIMERE 01:00 PM AOD 550 nm estimation to correct it toward values similar to that of MODIS Aqua.

We use the **AOD** and **other variables** from CHIMERE outputs, in addition to some **meteorological fields**. We feed them to a trained artificial intelligence (AI) agent which is a **random forest** regressor. The latter transforms the inputs into an AOD that is better than the raw CHIMERE AOD output when compared to MODIS AOD. This is because the AI is trained using MODIS AOD as reference values during the supervised training stage. The inference of the AOD is done pixel-by-pixel, i.e. no surrounding pixels information is needed.



Inputs

- CHIMERE simulation at 1:00 PM : AOD₅₅₀, PM₁₀, PM_{2.5}, pDUST, pOCAR, pWATER, pSALT, ROOH, HCNM, pH₂SO₄, pHNO₃, pNH₃, sphu, O₃, SO₂, CO, OH, NH₃, TOL, NO_x, NO_y
- ERA5 at 1:00 PM: pres, relh, soil moisture, boundary layer height, albedo, shortwave radiation, surface relative humidity, surface latent heat flux, surface sensible heat flux.

Reference for supervision

- MODIS/aqua AOD_{550nm} overpass at 01:30 PM.

Training + testing dataset : The first 20 days of each month = 240 days
=> Table of ~1.4M simulation ground pixel and 96 features

Results and conclusion

We apply our post processing correction method for the 122 validation days. The corrected AOD is similar to the MODIS observations (Figure.4 (b) & (c)). We see an improvement in the correlation coefficient with MODIS AOD of 27% and 71% reduction in the root mean squared error. The comparison with AERONET stations measurements also show a substantial improvement.

Table.2 Performance of random forest correction of CHIMERE AOD, values for uncorrected AOD between brackets.

Reference estimation	r	RMSE
MODIS	0.71 (0.56)	0.19 (0.65)
AERONET	0.85 (0.54)	0.12 (0.45)

Perspectives

- Develop the AOD correction post-processor using MODIS AOD from the Terra platform which has an overpass time of 10:30 AM.

Table.1 Comparison pros and cons of AOD estimation using CHIMERE simulation vs MODIS observations.

	CHIMERE AOD	MODIS AOD
Temporal resolution	1h	24h
spatial coverage	Full coverage	Cloud-free only
Accuracy	good	Better

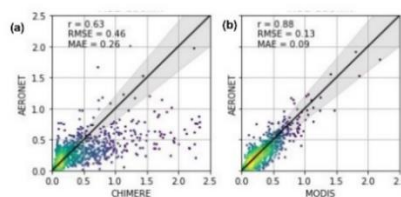


Figure.2 Scatter plot showing the correlation with AERONET station measurements for MODIS (a) and for CHIMERE (b).

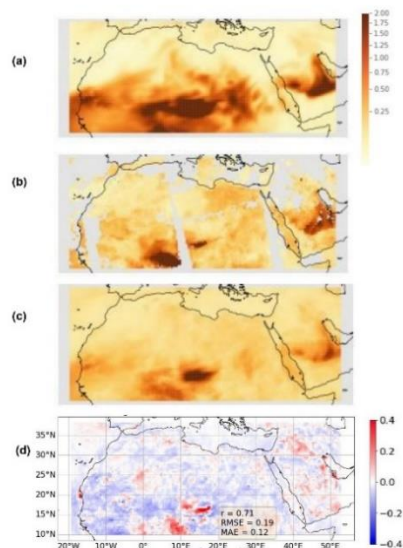


Figure.4 Map showing the AOD over the CHIMERE simulation domain. (a) CHIMERE simulation AOD₅₅₀, (b) MODIS Aqua, (c) CHIMERE corrected AOD, (d) Average AOD correction bias with respect to MODIS on the testing dates.

References :

- [1] Maillet, S., Menut, L., Khvorostyanov, D., Valari, M., Couvidat, F., Slour, G., Turquetly, S., Briant, R., Tuccella, P., Bessagnet, B., Colette, A., Létinois, L., Markakis, K., and Meleux, F.: CHIMERE-2017: from urban to hemispheric chemistry-transport modeling, *Geosci. Model Dev.*, 10, 2397–2423
[2] https://ladsweb.modaps.eosdis.nasa.gov/missions-and-measurements/products/MYD04_L2

farouk.lemmouchi@lisa.ipsl.fr

Références

1. Lauer, A.; Hendricks, J. Simulating Aerosol Microphysics with the ECHAM4/MADE GCM – Part II: Results from a First Multiannual Simulation of the Submicrometer Aerosol. *Atmos. Chem. Phys.* 2006, 6, 5495–5513.
2. Todd, M.C.; Washington, R.; Martins, J.V.; Dubovik, O.; Lizcano, G.; M'Bainayel, S.; Engelstaedter, S. Mineral Dust Emission from the Bodélé Depression, Northern Chad, during BoDEx 2005. *J. Geophys. Res.* 2007, 112, doi:10.1029/2006jd007170.
3. Mack, S.M.; Madl, A.K.; Pinkerton, K.E. Respiratory Health Effects of Exposure to Ambient Particulate Matter and Bioaerosols. *Compr. Physiol.* 2019, 10, 1–20.
4. Raizenne, M.; Neas, L.M.; Damokosh, A.I.; Dockery, D.W.; Spengler, J.D.; Koutrakis, P.; Ware, J.H.; Speizer, F.E. Health Effects of Acid Aerosols on North American Children: Pulmonary Function. *Environ. Health Perspect.* 1996, 104, 506–514.
5. Fussell, J.C.; Kelly, F.J. Mechanisms Underlying the Health Effects of Desert Sand Dust. *Environ. Int.* 2021, 157, 106790.
6. Maheshwarkar, P.; Sunder Raman, R. Population Exposure across Central India to PM_{2.5} Derived Using Remotely Sensed Products in a Three-Stage Statistical Model. *Sci. Rep.* 2021, 11, 544.
7. Lin, G.; Penner, J.E.; Flanner, M.G.; Sillman, S.; Xu, L.; Zhou, C. Radiative Forcing of Organic Aerosol in the Atmosphere and on Snow: Effects of SOA and Brown Carbon. *J. Geophys. Res.* 2014, 119, 7453–7476.
8. Sellitto, P.; Belhadji, R.; Kloss, C.; Legras, B. Radiative Impacts of the Australian Bushfires 2019–2020 – Part 1: Large-Scale Radiative Forcing 2022.
9. Mallet, V.; Sportisse, B. Uncertainty in a Chemistry-Transport Model Due to Physical Parameterizations and Numerical Approximations: An Ensemble Approach Applied to Ozone Modeling. *Journal of Geophysical Research* 2006, 111.
10. Hodzic, A.; Chepfer, H.; Vautard, R.; Chazette, P.; Beekmann, M.; Bessagnet, B.; Chatenet, B.; Cuesta, J.; Drobinski, P.; Goloub, P.; et al. Comparison of Aerosol Chemistry Transport Model Simulations with Lidar and Sun Photometer Observations at a Site near Paris. *J. Geophys. Res.* 2004, 109, doi:10.1029/2004jd004735.
11. Kalita, T.L.; Titlyanov, E.A. Influence of Temperature on the Infradian Growth Rhythm in *Ulva Lactuca* (Chlorophyta). *Eur. J. Phycol.* 2013, 48, 210–220.
12. Delmas, R.; Mégie, G.; Peuch, V.-H. *Physique et chimie de l'atmosphère*; Belin, 2005; ISBN 9782701137001.
13. Helbig, M.; Gerken, T.; Beamesderfer, E.R.; Baldocchi, D.D.; Banerjee, T.; Biraud, S.C.; Brown, W.O.J.; Brunsell, N.A.; Burakowski, E.A.; Burns, S.P.; et al. Integrating

- Continuous Atmospheric Boundary Layer and Tower-Based Flux Measurements to Advance Understanding of Land-Atmosphere Interactions. *Agricultural and Forest Meteorology* 2021, 307, 108509.
14. Cuesta, J.; Marsham, J.H.; Parker, D.J.; Flamant, C. Dynamical Mechanisms Controlling the Vertical Redistribution of Dust and the Thermodynamic Structure of the West Saharan Atmospheric Boundary Layer during Summer. *Atmos. Sci. Lett.* 2009, 10, 34–42.
 15. Wouters; Petrova A Novel Framework to Investigate Atmospheric Boundary Layer Dynamics from Balloon Soundings Worldwide: CLASS4GL. *Geophys. Astrophys. Fluid Dyn.* 2019.
 16. Stull, R.B. *An Introduction to Boundary Layer Meteorology*; Springer Science & Business Media, 1988; ISBN 9789027727688.
 17. Seinfeld, J.H.; Pandis, S.N.; Noone, K. Atmospheric Chemistry and Physics: From Air Pollution to Climate Change. *Phys. Today* 1998, 51, 88–90.
 18. Radiation and Climate. In *International Geophysics*; International geophysics series; Elsevier, 2002; pp. 442–521 ISBN 9780124514515.
 19. La circulation générale de l’atmosphère - Cours météo Available online: <https://www.meteocontact.fr/pour-aller-plus-loin/la-circulation-generale-de-l-atmosphere> (accessed on 7 March 2023).
 20. *Dust Transported in the N Orth-East and South-East Trade Winds in the Atlantic Ocean*;
 21. Francis, D.; Eayrs, C.; Chaboureau, J.-P.; Mote, T.; Holland, D.M. A Meandering Polar Jet Caused the Development of a Saharan Cyclone and the Transport of Dust toward Greenland. *Adv. Sci. Res.* 2019, 16, 49–56.
 22. A Satellite’s View of the Climate Energy Budget Available online: <https://www.e-education.psu.edu/earth103/node/1007> (accessed on 27 December 2022).
 23. Grainger, R.G. *Some Useful Formulae for Particle Size Distributions and Optical Properties*; 2022;
 24. Watson, J.G.; Chow, J.C.; Chen, L.; Wang, X.; Merrifield, T.M.; Fine, P.M.; Barker, K. Measurement System Evaluation for Upwind/Downwind Sampling of Fugitive Dust Emissions. *Aerosol and Air Quality Research* 2011, 11, 331–350.
 25. Schumann, U. *Atmospheric Physics*; Springer Berlin Heidelberg, 2012;
 26. Lang-Yona, N.; Abo-Riziq, A.; Erlick, C.; Segre, E.; Trainic, M.; Rudich, Y. Interaction of Internally Mixed Aerosols with Light. *Phys. Chem. Chem. Phys.* 2010, 12, 21–31.

27. Li, W.; Ge, P.; Chen, M.; Tang, J.; Cao, M.; Cui, Y.; Hu, K.; Nie, D. Tracers from Biomass Burning Emissions and Identification of Biomass Burning. *Atmosphere (Basel)* 2021, *12*, 1401.
28. Reid; Koppmann; Eck A Review of Biomass Burning Emissions Part II: Intensive Physical Properties of Biomass Burning Particles. *Chem. Phys. Lipids* 2005.
29. Us Epa, O. Climate Change Indicators: Wildfires. 2016.
30. Westerling, A.L. Increasing Western U.S. Forest Wildfire Activity: Sensitivity to Changes in the Timing of Spring. *Phil. Trans. R. Soc. B* 2016, *371*.
31. Cammas; Brioude; Chaboureau Injection in the Lower Stratosphere of Biomass Fire Emissions Followed by Long-Range Transport: A MOZAIC Case Study. *Atmos. Clim. Sci.* 2009.
32. Ohneiser, K.; Ansmann, A.; Witthuhn, J.; Deneke, H.; Chudnovsky, A.; Walter, G. Self-Lofting of Wildfire Smoke in the Troposphere and Stratosphere Caused by Radiative Heating: Simulations vs Space Lidar Observations 2022.
33. Nowak, S.; Lafon, S.; Caquineau, S.; Journet, E.; Laurent, B. Quantitative Study of the Mineralogical Composition of Mineral Dust Aerosols by X-Ray Diffraction. *Talanta* 2018, *186*, 133–139.
34. Lachatre, M.; Foret, G.; Laurent, B.; Siour, G.; Cuesta, J.; Dufour, G.; Meng, F.; Tang, W.; Zhang, Q.; Beekmann, M. Air Quality Degradation by Mineral Dust over Beijing, Chengdu and Shanghai Chinese Megacities. *Atmosphere* 2020, *11*, 708.
35. Chen, W.; Meng, H.; Song, H.; Zheng, H. Progress in Dust Modelling, Global Dust Budgets, and Soil Organic Carbon Dynamics. *Land* 2022, *11*, 176.
36. Xiao, H.-W.; Xiao, H.-Y.; Shen, C.-Y.; Zhang, Z.-Y.; Long, A.-M. Chemical Composition and Sources of Marine Aerosol over the Western North Pacific Ocean in Winter. *Atmosphere (Basel)* 2018, *9*, 298.
37. Ding, A.; Wang, T.; Zhao, M.; Wang, T.; Li, Z. Simulation of Sea-Land Breezes and a Discussion of Their Implications on the Transport of Air Pollution during a Multi-Day Ozone Episode in the Pearl River Delta of China. *Atmos. Environ. (1994)* 2004, *38*, 6737–6750.
38. Finlayson-Pitts, B.J.; Wingen, L.M.; Perraud, V.; Ezell, M.J. Open Questions on the Chemical Composition of Airborne Particles. *Communications Chemistry* 2020, *3*, 1–5.
39. Curtius, J. Nucleation of Atmospheric Particles. *EPJ Web of Conferences* 2009, *1*, 199–209.
40. Boy, M.; Kulmala, M. Nucleation Events in the Continental Boundary Layer: Influence of Physical and Meteorological Parameters. *Atmos. Chem. Phys.* 2002, *2*, 1–16.

41. Srivastava, D.; Vu, T.V.; Tong, S.; Shi, Z.; Harrison, R.M. Formation of Secondary Organic Aerosols from Anthropogenic Precursors in Laboratory Studies. *npj Climate and Atmospheric Science* 2022, 5, 1–30.
42. Toossi, R.; Novakov, T. The Lifetime of Aerosols in Ambient Air: Consideration of the Effects of Surfactants and Chemical Reactions. *Atmos. Environ.* 1985, 19, 127–133.
43. Prasad, A.K.; El-Askary, H.; Kafatos, M. Implications of High Altitude Desert Dust Transport from Western Sahara to Nile Delta during Biomass Burning Season. *Environ. Pollut.* 2010, 158, 3385–3391.
44. Kok, J.F.; Adebisi, A.A.; Albani, S.; Balkanski, Y.; Checa-Garcia, R.; Chin, M.; Colarco, P.R.; Hamilton, D.S.; Huang, Y.; Ito, A.; et al. Contribution of the World's Main Dust Source Regions to the Global Cycle of Desert Dust. *Atmos. Chem. Phys.* 2021, 21, 8169–8193.
45. Chin, M.; Ginoux, P.; Kinne, S.; Torres, O.; Holben, B.N.; Duncan, B.N.; Martin, R.V.; Logan, J.A.; Higurashi, A.; Nakajima, T. Tropospheric Aerosol Optical Thickness from the GOCART Model and Comparisons with Satellite and Sun Photometer Measurements. *J. Atmos. Sci.* 2002, 59, 461–483.
46. Cohen, A.J.; Brauer, M.; Burnett, R.; Anderson, H.R.; Frostad, J.; Estep, K.; Balakrishnan, K.; Brunekreef, B.; Dandona, L.; Dandona, R.; et al. Estimates and 25-Year Trends of the Global Burden of Disease Attributable to Ambient Air Pollution: An Analysis of Data from the Global Burden of Diseases Study 2015. *Lancet* 2017, 389, 1907–1918.
47. Shah, S.; Wei, Y.; Chan, C. Effects of Particulate Matter from Air Pollution on Cardiovascular System of the Body. *Int. J. Appl. Res.* 2020, 6, 281–285.
48. Darquenne, C. Deposition Mechanisms. *J. Aerosol Med. Pulm. Drug Deliv.* 2020, 33, 181–185.
49. Smith, D.J.; Griffin, D.W.; Jaffe, D.A. The High Life: Transport of Microbes in the Atmosphere. *Eos (Washington DC)* 2011, 92, 249–250.
50. Griffin, D.W.; Kellogg, C.A.; Shinn, E.A. Dust in the Wind: Long Range Transport of Dust in the Atmosphere and Its Implications for Global Public and Ecosystem Health. *Glob. Chang. Hum. Health* 2001, 2, 20–33.
51. Du, Y.; Xu, X.; Chu, M.; Guo, Y.; Wang, J. Air Particulate Matter and Cardiovascular Disease: The Epidemiological, Biomedical and Clinical Evidence. *J. Thorac. Dis.* 2016, 8, E8–E19.
52. Meskhidze, N. Dust and Pollution: A Recipe for Enhanced Ocean Fertilization? *J. Geophys. Res.* 2005, 110, doi:10.1029/2004jd005082.
53. Remer, L.A. Dust, Fertilization and Sources. *Environ. Res. Lett.* 2006, 1, 011001.
54. Yu, H.; Chin, M.; Yuan, T.; Bian, H.; Remer, L.A.; Prospero, J.M.; Omar, A.; Winker, D.; Yang, Y.; Zhang, Y.; et al. The Fertilizing Role of African Dust in the Amazon

- Rainforest: A First Multiyear Assessment Based on Data from Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations. *Geophys. Res. Lett.* 2015, *42*, 1984–1991.
55. Li, F.; Luo, Q.; Lin, W.; Li, J.; Jiang, B. Simulating the Impact of Biomass Burning Aerosols on an Intensive Precipitation Event in Urban Areas of the Pearl River Delta. *Atmos. Res.* 2022, *266*, 105966.
56. Novakov, T.; Penner, J.E. Large Contribution of Organic Aerosols to Cloud-Condensation-Nuclei Concentrations. *Nature* 1993, *365*, 823–826.
57. Singh, A.; Agrawal, M. Acid Rain and Its Ecological Consequences. *J. Environ. Biol.* 2008, *29*, 15–24.
58. van Breemen, N.; van Dijk, H.F. Ecosystem Effects of Atmospheric Deposition of Nitrogen in The Netherlands. *Environ. Pollut.* 1988, *54*, 249–274.
59. Huang, H.; Gu, Y.; Xue, Y.; Jiang, J.; Zhao, B. Assessing Aerosol Indirect Effect on Clouds and Regional Climate of East/South Asia and West Africa Using NCEP GFS. *Clim. Dyn.* 2019, *52*, 5759–5774.
60. Penner, J.E.; Andreae, M.; Annegarn, H.; Barrie, L.; Feichter, J.; Hegg, D.; Jayaraman, A.; Leaitch, R.; Murphy, D.; Nganga, J.; et al. Aerosols, Their Direct and Indirect Effects Co-Ordinating Lead Author. In; 2001.
61. Short-Lived Climate Forcers. In *Climate change 2021: The physical science basis*; 2021.
62. Hansen, J.E.; Travis, L.D. Light Scattering in Planetary Atmospheres. *Space Science Reviews* 1974, *16*, 527–610.
63. Gearhart, C. Black-Body Radiation. In *Compendium of Quantum Physics*; Greenberger, D., Hentschel, K., Weinert, F., Eds.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2009; pp. 39–42 ISBN 9783540706267.
64. The Laws of Infrared Heating Available online: <https://www.ceramicx.com/information/support/why-infrared-laws-of-infrared-heating/> (accessed on 25 October 2022).
65. Popa, D.; Udrea, F. Towards Integrated Mid-Infrared Gas Sensors. *Sensors* 2019, *19*, doi:10.3390/s19092076.
66. Www.e-education.psu.edu Available online: <https://www.e-education.psu.edu/meteo300/node/785> (accessed on 25 July 2022).
67. Deuzé, J.L.; Goloub, P.; Herman, M.; Marchand, A.; Perry, G.; Susana, S.; Tanré, D. Estimate of the Aerosol Properties over the Ocean with POLDER. *J. Geophys. Res.* 2000, *105*, 15329–15346.
68. Nishita, T.; Sirai, T.; Tadamura, K.; Nakamae, E. Display of the Earth Taking into Account Atmospheric Scattering. In Proceedings of the Proceedings of the 20th annual

conference on Computer graphics and interactive techniques - SIGGRAPH '93; ACM Press: New York, New York, USA, 1993.

69. Haiducek, J. EXPERIMENTAL VALIDATION TECHNIQUES FOR THE HELEEOS OFF-AXIS LASER PROPAGATION MODEL, AIR FORCE INSTITUTE OF TECHNOLOGY: Ohio, 2010.
70. Lord, R. XXXIV. On the Transmission of Light through an Atmosphere Containing Small Particles in Suspension, and on the Origin of the Blue of the Sky. *Lond. Edinb. Dublin Philos. Mag. J. Sci.* 1899, 47, 375–384.
71. Cornette, W.M.; Shanks, J.G. Physically Reasonable Analytic Expression for the Single-Scattering Phase Function. *Appl. Opt.* 1992, 31, 3152–3160.
72. Jensen, K.D. Flow Measurements. *J. Brazil. Soc. Mech. Sci. Eng.* 2004, 26, 400–419.
73. Aerosol Robotic Network (AERONET) Homepage Available online: <https://aeronet.gsfc.nasa.gov/> (accessed on 6 October 2021).
74. Dubovik, O.; Holben, B.; Eck, T.F.; Smirnov, A.; Kaufman, Y.J.; King, M.D.; Tanré, D.; Slutsker, I. Variability of Absorption and Optical Properties of Key Aerosol Types Observed in Worldwide Locations. *J. Atmos. Sci.* 2002, 59, 590–608.
75. www.thermofisher.com Available online: www.thermofisher.com (accessed on 12 October 2022).
76. Minikin, A.; Petzold, A.; Weinzierl, B.; Gayet, J.-F. In Situ Measurement Methods for Atmospheric Aerosol Particles and Cloud Elements. In *Atmospheric Physics: Background – Methods – Trends*; Schumann, U., Ed.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2012; pp. 297–315 ISBN 9783642301834.
77. Winker, D.M.; Hunt, W.H.; McGill, M.J. Initial Performance Assessment of CALIOP. *Geophys. Res. Lett.* 2007, 34, doi:10.1029/2007gl030135.
78. <https://www.eoportal.org> Available online: <https://www.eoportal.org/satellite-missions/calipso#mission-status> (accessed on 10 October 2022).
79. Giles, D.M.; Sinyuk, A.; Sorokin, M.G.; Schafer, J.S.; Smirnov, A.; Slutsker, I.; Eck, T.F.; Holben, B.N.; Lewis, J.R.; Campbell, J.R.; et al. Advancements in the Aerosol Robotic Network (AERONET) Version 3 Database – Automated near-Real-Time Quality Control Algorithm with Improved Cloud Screening for Sun Photometer Aerosol Optical Depth (AOD) Measurements. *Atmospheric Measurement Techniques* 2019, 12, 169–209.
80. <https://aeronet.gsfc.nasa.gov> Available online: <https://aeronet.gsfc.nasa.gov> (accessed on 6 October 2021).
81. <https://airbornescience.nasa.gov> Available online: https://airbornescience.nasa.gov/sunsat/content/SunSat_AERONET (accessed on 16 October 2022).

82. Barreto, Á.; Cuevas, E.; Granados-Muñoz, M.-J.; Alados-Arboledas, L.; Romero, P.M.; Gröbner, J.; Kouremeti, N.; Almansa, A.F.; Stone, T.; Toledano, C.; et al. The New Sun-Sky-Lunar Cimel CE318-T Multiband Photometer – a Comprehensive Performance Evaluation. *Atmospheric Measurement Techniques* 2016, 9, 631–654.
83. Hodgkins, L. Schoolphysics ::Welcome:: Available online: https://www.schoolphysics.co.uk/age11-14/Astronomy/text/Artificial_satellites/index.html (accessed on 26 December 2022).
84. Lee, K.H.; Li, Z.; Kim, Y.J.; Kokhanovsky, A. Atmospheric Aerosol Monitoring from Satellite Observations: A History of Three Decades. In *Atmospheric and Biological Environmental Monitoring*; Springer Netherlands: Dordrecht, 2009; pp. 13–38 ISBN 9781402096730.
85. <https://commons.wikimedia.org> Available online: https://commons.wikimedia.org/wiki/File:Solar_Spectrum.png (accessed on 27 August 2022).
86. <https://aqua.nasa.gov/modis> Available online: <https://aqua.nasa.gov/modis> (accessed on 27 August 2022).
87. <https://modis.gsfc.nasa.gov> Available online: <https://modis.gsfc.nasa.gov/about/specifications.php> (accessed on 27 July 2022).
88. <https://crisp.nus.edu.sg> Available online: <https://crisp.nus.edu.sg/~research/tutorial/modis.htm> (accessed on 27 August 2022).
89. <https://earthobservatory.nasa.gov> Available online: https://earthobservatory.nasa.gov/global-maps/MODAL2_M_AER_OD (accessed on 10 October 2022).
90. Veefkind, J.P.; Aben, I.; McMullan, K.; Förster, H.; de Vries, J.; Otter, G.; Claas, J.; Eskes, H.J.; de Haan, J.F.; Kleipool, Q.; et al. TROPOMI on the ESA Sentinel-5 Precursor: A GMES Mission for Global Observations of the Atmospheric Composition for Climate, Air Quality and Ozone Layer Applications. *Remote Sens. Environ.* 2012, *Complete*, 70–83.
91. Algorithm Theoretical Basis Document for the TROPOMI L01b Data Processor Available online: <https://sentinels.copernicus.eu/documents/247904/2476257/Sentinel-5P-TROPOMI-Level-1B-ATBD> (accessed on 23 December 2022).
92. <https://sentinel.esa.int> Available online: <https://sentinel.esa.int/web/sentinel/user-guides/sentinel-5p-tropomi> (accessed on 27 August 2022).
93. Home Available online: <http://www.tropomi.eu/> (accessed on 23 December 2022).
94. Georgoulias, A.K.; Boersma, K.F.; van Vliet, J.; Zhang, X.; van der A, R.; Zanis, P.; de Laat, J. Detection of NO₂ Pollution Plumes from Individual Ships with the TROPOMI/S5P Satellite Sensor. *Environ. Res. Lett.* 2020, 15, 124037.

95. Van Geffen, J.H.G.M.; Eskes, H.J.; Boersma, K.F.; Veefkind, J.P. TROPOMI ATBD of the Total and Tropospheric NO₂ Data Products Available online: <https://sentinel.esa.int/documents/247904/2476257/Sentinel-5P-TROPOMI-ATBD-NO2-data-products> (accessed on 23 December 2022).
96. Cuesta, J.; Eremenko, M.; Flamant, C.; Dufour, G.; Laurent, B.; Bergametti, G.; Höpfner, M.; Orphal, J.; Zhou, D. Three-Dimensional Distribution of a Major Desert Dust Outbreak over East Asia in March 2008 Derived from IASI Satellite Observations. *Journal of Geophysical Research: Atmospheres* 2015, *120*, 7099–7127.
97. Zawada, D.; Franssens, G.; Loughman, R.; Mikkonen, A.; Rozanov, A.; Emde, C.; Bourassa, A.; Dueck, S.; Lindqvist, H.; Ramon, D.; et al. Systematic Comparison of Vectorial Spherical Radiative Transfer Models in Limb Scattering Geometry. *Atmos. Meas. Tech.* 2021, *14*, 3953–3972.
98. Cheng, A.F.; Domingue, D.L. Radiative Transfer Models for Light Scattering from Planetary Surfaces. *J. Geophys. Res.* 2000, *105*, 9477–9482.
99. Mishchenko, M.; Travis, L.; Lacis, A. RADIATIVE TRANSFER | Scattering. *Encyclopedia of Atmospheric Sciences* 2003, 1882–1891.
100. Chandrasekhar, S. *Radiative Transfer*; Courier Corporation, 2013; ISBN 9780486318455.
101. Stamnes, K. The Theory of Multiple Scattering of Radiation in Plane Parallel Atmospheres. *Rev. Geophys.* 1986, *24*, 299.
102. Dahlback, A.; Stamnes, K. A New Spherical Model for Computing the Radiation Field Available for Photolysis and Heating at Twilight. *Planet. Space Sci.* 1991, *39*, 671–683.
103. Stamnes, K.; Tsay, S.C.; Wiscombe, W.; Jayaweera, K. Numerically Stable Algorithm for Discrete-Ordinate-Method Radiative Transfer in Multiple Scattering and Emitting Layered Media. *Appl. Opt.* 1988, *27*, 2502–2509.
104. Korokin, S.; Sayer, A.M.; Ibrahim, A.; Lyapustin, A. A Practical Guide to Writing a Radiative Transfer Code. *Comput. Phys. Commun.* 2022, *271*, 108198.
105. Efremenko, D.S.; Loyola, D.G.; Doicu, A.; Spurr, R.J.D. Multi-Core-CPU and GPU-Accelerated Radiative Transfer Models Based on the Discrete Ordinate Method. *Comput. Phys. Commun.* 2014, *185*, 3079–3089.
106. Ramon, D.; Steinmetz, F.; Jolivet, D.; Compiègne, M.; Frouin, R. Modeling Polarized Radiative Transfer in the Ocean-Atmosphere System with the GPU-Accelerated SMART-G Monte Carlo Code. *J. Quant. Spectrosc. Radiat. Transf.* 2019, *222–223*, 89–107.
107. Ramon, D. HYGEOS - Smartg Available online: <https://www.hygeos.com/smartg> (accessed on 15 February 2023).

108. Spurr, R.J.D. VLIDORT: A Linearized Pseudo-Spherical Vector Discrete Ordinate Radiative Transfer Code for Forward Model and Retrieval Studies in Multilayer Multiple Scattering Media. *J. Quant. Spectrosc. Radiat. Transf.* 2006, 102, 316–342.
109. Saunders, R.; Hocking, J.; Turner, E.; Rayner, P.; Rundle, D.; Brunel, P.; Vidot, J.; Roquet, P.; Matricardi, M.; Geer, A.; et al. An Update on the RTTOV Fast Radiative Transfer Model (Currently at Version 12). *Geosci. Model Dev.* 2018, 11, 2717–2737.
110. Berk, A.; Conforti, P.; Kennett, R.; Perkins, T.; Hawes, F.; van den Bosch, J. MODTRAN6: A Major Upgrade of the MODTRAN Radiative Transfer Code. In Proceedings of the Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XX; Velez-Reyes, M., Kruse, F.A., Eds.; SPIE, June 13 2014.
111. MODTRAN® Available online: <http://modtran.spectral.com/> (accessed on 15 February 2023).
112. Clough, S.A.; Shephard, M.W.; Mlawer, E.J.; Delamere, J.S.; Iacono, M.J.; Cady-Pereira, K.; Boukabara, S.; Brown, P.D. Atmospheric Radiative Transfer Modeling: A Summary of the AER Codes. *J. Quant. Spectrosc. Radiat. Transf.* 2005, 91, 233–244.
113. Smith, H. J. P. Dube, D. J. Gardner, M. E. Clough, S. A. Kneizys, F. X. *FASCODE - Fast Atmospheric Signature Code (Spectral Transmittance and Radiance)*; VISIDYNE INC BURLINGTON MA, 1978;
114. Matricardi, M.; López-Puertas, M.; Funke, B. Modeling of Nonlocal Thermodynamic Equilibrium Effects in the Classical and Principal Component-Based Version of the RTTOV Fast Radiative Transfer Model. *J. Geophys. Res.* 2018, 123, 5741–5761.
115. Rodgers, C.D. *Inverse Methods For Atmospheric Sounding: Theory And Practice*; World Scientific, 2000; ISBN 9789814498685.
116. Rodgers, C.D. Retrieval of Atmospheric Temperature and Composition from Remote Measurements of Thermal Radiation. *Rev. Geophys.* 1979, doi:10.1029/RG014i004p00609.
117. Steck, T. Methods for Determining Regularization for Atmospheric Retrieval Problems. *Appl. Opt.* 2002, 41, 1788–1797.
118. Mailler, S.; Menut, L.; Khvorostyanov, D.; Valari, M.; Couvidat, F.; Siour, G.; Turquety, S.; Briant, R.; Tuccella, P.; Bessagnet, B.; et al. CHIMERE-2017: From Urban to Hemispheric Chemistry-Transport Modeling. *Geosci. Model Dev.* 2017, 10, 2397–2423.
119. <https://www.lmd.polytechnique.fr/chimere> Available online: <https://www.lmd.polytechnique.fr/chimere> (accessed on 8 August 2022).
120. Ciarelli, G.; Theobald, M.R.; Vivanco, M.G.; Beekmann, M.; Aas, W.; Andersson, C.; Bergström, R.; Manders-Groot, A.; Couvidat, F.; Mircea, M.; et al. Trends of Inorganic and Organic Aerosols and Precursor Gases in Europe: Insights from the EURODELTA Multi-Model Experiment over the 1990--2010 Period, *Geosci. Model Dev.*, 12, 4923--4954 2019.

121. Cholakian; Beekmann; Colette Simulation of Fine Organic Aerosols in the Western Mediterranean Area during the ChArMEx 2013 Summer Campaign. *Atmos. Clim. Sci.* 2018.
122. Deroubaix; Flamant; Menut Interactions of Atmospheric Gases and Aerosols with the Monsoon Dynamics over the Sudano-Guinean Region during AMMA. *Atmos. Clim. Sci.* 2018.
123. Fortems-Cheiney, A.; Dufour, G.; Foret, G.; Siour, G.; Van Damme, M.; Coheur, P.-F.; Clarisse, L.; Clerbaux, C.; Beekmann, M. Understanding the Simulated Ammonia Increasing Trend from 2008 to 2015 over Europe with CHIMERE and Comparison with IASI Observations. *Atmosphere* 2022, 13, 1101.
124. [Http://www.prevoir.org/](http://www.prevoir.org/) Available online: <http://www.prevoir.org/> (accessed on 27 July 2022).
125. Menut, L.; Schmechtig, C.; Marticorena, B. Sensitivity of the Sandblasting Flux Calculations to the Soil Size Distribution Accuracy. *J. Atmos. Ocean. Technol.* 2005, 22, 1875–1884.
126. Hauglustaine, D.A.; Hourdin, F.; Jourdain, L.; Filiberti, M.-A.; Walters, S.; Lamarque, J.-F.; Holland, E.A. Interactive Chemistry in the Laboratoire de Météorologie Dynamique General Circulation Model: Description and Background Tropospheric Chemistry Evaluation. *J. Geophys. Res. D: Atmos.* 2004, 109, doi:10.1029/2003JD003957.
127. Oreggioni, G.D.; Monforti-Ferraio, F.; Crippa, M.; Schaaf, E.; Guizzardi, D.; Muntean, M.; Duerr, M.; Vignati, E. EDGAR v5.0: A Tool to Evaluate the Influence of Technology Incorporation and Regulatory Frameworks on Global Greenhouse Gases and Air Pollutant Emissions 2020.
128. White, B.R. Soil Transport by Winds on Mars. *J. Geophys. Res.* 1979, 84, 4643.
129. Marticorena, B.; Bergametti, G. Modeling the Atmospheric Dust Cycle: 1. Design of a Soil-Derived Dust Emission Scheme. *J. Geophys. Res.* 1995, 100, 16415.
130. van Leer, B. Towards the Ultimate Conservative Difference Scheme. V. A Second-Order Sequel to Godunov's Method. *J. Comput. Phys.* 1979, 32, 101–136.
131. Couvidat; Bessagnet Development of an Inorganic and Organic Aerosol Model (CHIMERE 2017β v1.0): Seasonal and Spatial Evaluation over Europe. *Geosci. Model Dev.* 2018.
132. Wesely, M.L. Parameterization of Surface Resistances to Gaseous Dry Deposition in Regional-Scale Numerical Models. *Atmos. Environ.* 2007, 41, 52–63.
133. Croft, B.; Lohmann, U.; Martin, R.V.; Stier, P.; Wurzler, S.; Feichter, J.; Hoose, C.; Heikkilä, U.; van Donkelaar, A.; Ferrachat, S. Influences of In-Cloud Aerosol Scavenging Parameterizations on Aerosol Concentrations and Wet Deposition in ECHAM5-HAM. *Atmos. Chem. Phys.* 2010, 10, 1511–1543.

134. Derognat, C. Effect of Biogenic Volatile Organic Compound Emissions on Tropospheric Chemistry during the Atmospheric Pollution Over the Paris Area (ESQUIF) Campaign in the Ile-de-France Region. *Journal of Geophysical Research* 2003, *108*.
135. Couvidat, F.; Debry, É.; Sartelet, K.; Seigneur, C. A Hydrophilic/Hydrophobic Organic (H₂O) Aerosol Model: Development, Evaluation and Sensitivity Analysis. *J. Geophys. Res.* 2012, *117*, doi:10.1029/2011jd017214.
136. Couvidat, F.; Sartelet, K. The Secondary Organic Aerosol Processor (SOAP v1.0) Model: A Unified Model with Different Ranges of Complexity Based on the Molecular Surrogate Approach. *Geoscientific Model Development* 2015, *8*, 1111–1138.
137. Fountoukis; Nenes ISORROPIA II: A Computationally Efficient Thermodynamic Equilibrium Model for K_p ECa 2_p EMg 2_p ENH 1_p ENa 1_p ESO 2_Å ENO 1_Å EC1 1_Å EH₂O Aerosols. *Atmos. Chem. Phys.* 2007.
138. Bian, H.; Prather, M.J. Fast-J2: Accurate Simulation of Stratospheric Photolysis in Global Chemical Models. *J. Atmos. Chem.* 2002, *41*, 281–296.
139. Lemmouchi, F.; Cuesta, J.; Eremenko, M.; Derognat, C.; Siour, G.; Dufour, G.; Sellitto, P.; Turquety, S.; Tran, D.; Liu, X.; et al. Three-Dimensional Distribution of Biomass Burning Aerosols from Australian Wildfires Observed by TROPOMI Satellite Observations. *Remote Sensing* 2022, *14*, 2582.
140. Yu, P.; Davis, S.M.; Toon, O.B.; Portmann, R.W.; Bardeen, C.G.; Barnes, J.E.; Telg, H.; Maloney, C.; Rosenlof, K.H. Persistent Stratospheric Warming Due to 2019–2020 Australian Wildfire Smoke. *Geophys. Res. Lett.* 2021, *48*, doi:10.1029/2021gl092609.
141. Khaykin, S.; Legras, B.; Bucci, S.; Sellitto, P.; Isaksen, L.; Tencé, F.; Bekki, S.; Bourassa, A.; Rieger, L.; Zawada, D.; et al. The 2019/20 Australian Wildfires Generated a Persistent Smoke-Charged Vortex Rising up to 35 Km Altitude. *Communications Earth & Environment* 2020, *1*, 1–12.
142. Chen, L.; Li, Q.; Wu, D.; Sun, H.; Wei, Y.; Ding, X.; Chen, H.; Cheng, T.; Chen, J. Size Distribution and Chemical Composition of Primary Particles Emitted during Open Biomass Burning Processes: Impacts on Cloud Condensation Nuclei Activation. *Sci. Total Environ.* 2019, *674*, 179–188.
143. Li, Y. Cloud Condensation Nuclei Activity and Hygroscopicity of Fresh and Aged Biomass Burning Particles. *Pure and Applied Geophysics* 2019, *176*, 345–356.
144. Kennedy, I.M. The Health Effects of Combustion-Generated Aerosols. *Proc. Combust. Inst.* 2007, *31*, 2757–2770.
145. Freitas, S.R.; Longo, K.M.; Chatfield, R.; Latham, D.; Silva Dias, M.A.; Andreae, M.O.; Prins, E.; Santos, J.C.; Gielow, R.; Carvalho, J.A. Including the Sub-Grid Scale Plume Rise of Vegetation Fires in Low Resolution Atmospheric Transport Models. *Atmospheric Chemistry and Physics* 2007, *7*, 3385–3398.

146. Turquety, S.; Menut, L.; Siour, G.; Mailler, S.; Hadji-Lazaro, J.; George, M.; Clerbaux, C.; Hurtmans, D.; Coheur, P.-F. APIFLAME v2.0 Biomass Burning Emissions Model: Impact of Refined Input Parameters on Atmospheric Concentration in Portugal in Summer 2016. *Geoscientific Model Development* 2020, *13*, 2981–3009.
147. Kahn, R.A.; Chen, Y.; Nelson, D.L.; Leung, F.-Y.; Li, Q.; Diner, D.J.; Logan, J.A. Wildfire Smoke Injection Heights: Two Perspectives from Space. *Geophysical Research Letters* 2008, *35*.
148. Sofiev, M.; Ermakova, T.; Vankevich, R. Evaluation of the Smoke-Injection Height from Wild-Land Fires Using Remote-Sensing Data. *Atmospheric Chemistry and Physics* 2012, *12*, 1995–2006.
149. Sanders, A.F.J.; de Haan, J.F.; Sneep, M.; Apituley, A.; Stammes, P.; Vieitez, M.O.; Tilstra, L.G.; Tuinder, O.N.E.; Koning, C.E.; Veefkind, J.P. Evaluation of the Operational Aerosol Layer Height Retrieval Algorithm for Sentinel-5 Precursor: Application to O₂ A Band Observations from GOME-2A. *Atmospheric Measurement Techniques* 2015, *8*, 4947–4977.
150. Hess, M.; Koepke, P.; Schult, I. Optical Properties of Aerosols and Clouds: The Software Package OPAC. *Bull. Am. Meteorol. Soc.* 1998, *79*, 831–844.
151. Callies, J.; Corpaccioli, E.; Eisinger, M.; Lefebvre, A.; Munro, R.; Perez-Albinana, A.; Ricciarelli, B.; Calamai, L.; Gironi, G.; Veratti, R.; et al. GOME-2 Ozone Instrument Onboard the European METOP Satellites. In Proceedings of the Weather and Environmental Satellites; SPIE, September 15 2004; Vol. 5549, pp. 60–70.
152. Bond, T.C.; Doherty, S.J.; Fahey, D.W.; Forster, P.M.; Berntsen, T.; DeAngelo, B.J.; Flanner, M.G.; Ghan, S.; Kärcher, B.; Koch, D.; et al. Bounding the Role of Black Carbon in the Climate System: A Scientific Assessment. *J. Geophys. Res.* 2013, *118*, 5380–5552.
153. Singh, S.; Fiddler, M.N.; Bililign, S. Measurement of Size-Dependent Single Scattering Albedo of Fresh Biomass Burning Aerosols Using the Extinction-Minus-Scattering Technique with a Combination of Cavity Ring-down Spectroscopy and Nephelometry. *Atmospheric Chemistry and Physics* 2016, *16*, 13491–13507.
154. Konovalov, I.B.; Golovushkin, N.A.; Beekmann, M.; Andreae, M.O. Insights into the Aging of Biomass Burning Aerosol from Satellite Observations and 3D Atmospheric Modeling: Evolution of the Aerosol Optical Properties in Siberian Wildfire Plumes. *Atmospheric Chemistry & Physics* 2021, *21*, 357–392.
155. Hosseini, S.; Li, Q.; Cocker, D.; Weise, D.; Miller, A.; Shrivastava, M.; Miller, J.W.; Mahalingam, S.; Princevac, M.; Jung, H. Particle Size Distributions from Laboratory-Scale Biomass Fires Using Fast Response Instruments. *Atmospheric Chemistry and Physics* 2010, *10*, 8065–8076.
156. Sandvik, O.S.; Friberg, J.; Martinsson, B.G.; van Velthoven, P.F.J.; Hermann, M.; Zahn, A. Intercomparison of In-Situ Aircraft and Satellite Aerosol Measurements in the Stratosphere. *Sci. Rep.* 2019, *9*, 15576.

157. Winker, D.M.; Tackett, J.L.; Getzewich, B.J.; Liu, Z.; Vaughan, M.A.; Rogers, R.R. The Global 3-D Distribution of Tropospheric Aerosols as Characterized by CALIOP. *Atmospheric Chemistry and Physics* 2013, *13*, 3345–3361.
158. Chen, X.; Wang, J.; Xu, X.; Zhou, M.; Zhang, H.; Castro Garcia, L.; Colarco, P.R.; Janz, S.J.; Yorks, J.; McGill, M.; et al. First Retrieval of Absorbing Aerosol Height over Dark Target Using TROPOMI Oxygen B Band: Algorithm Development and Application for Surface Particulate Matter Estimates. *Remote Sensing of Environment* 2021, *265*, 112674.
159. Rao, L.; Xu, J.; Efremenko, D.S.; Loyola, D.G.; Doicu, A. Hyperspectral Satellite Remote Sensing of Aerosol Parameters: Sensitivity Analysis and Application to TROPOMI/S5P. *Front. Environ. Sci. Eng. China* 2022, *9*, doi:10.3389/fenvs.2021.770662.
160. Diner, D.J.; Beckert, J.C.; Reilly, T.H.; Bruegge, C.J.; Conel, J.E.; Kahn, R.A.; Martonchik, J.V.; Ackerman, T.P.; Davies, R.; Gerstl, S.A.W.; et al. Multi-Angle Imaging SpectroRadiometer (MISR) Instrument Description and Experiment Overview. *IEEE Trans. Geosci. Remote Sens.* 1998, *36*, 1072–1087.
161. Limbacher, J.A.; Kahn, R.A. Updated MISR Dark Water Research Aerosol Retrieval Algorithm – Part 1: Coupled 1.1 Km Ocean Surface Chlorophyll a Retrievals with Empirical Calibration Corrections. *Atmos. Meas. Tech.* 2017, *10*, 1539–1555.
162. Garay, M.J.; Witek, M.L.; Kahn, R.A.; Seidel, F.C.; Limbacher, J.A.; Bull, M.A.; Diner, D.J.; Hansen, E.G.; Kalashnikova, O.V.; Lee, H.; et al. Introducing the 4.4 Km Spatial Resolution MISR Aerosol Product.
163. Vandenbussche, S.; Kochenova, S.; Vandaele, A.C.; Kumps, N.; De Mazière, M. Retrieval of Desert Dust Aerosol Vertical Profiles from IASI Measurements in the TIR Atmospheric Window. *Atmospheric Measurement Techniques* 2013, *6*, 2577–2591.
164. Cuesta, J.; Flamant, C.; Eremenko, M.; Dufour, G.; Laurent, B.; Bergametti, G.; Aires, F.; Ryder, C. Three-Dimensional Distribution of a Major Saharan Dust Outbreak in June 2011 Derived from IASI Satellite Observations. In Proceedings of the 4th IASI International Conference; 2016.
165. Cuesta, J.; Flamant, C.; Gaetani, M.; Knippertz, P.; Fink, A.H.; Chazette, P.; Eremenko, M.; Dufour, G.; Di Biagio, C.; Formenti, P. Three-dimensional Pathways of Dust over the Sahara during Summer 2011 as Revealed by New Infrared Atmospheric Sounding Interferometer Observations. *Quarterly Journal of the Royal Meteorological Society* 2020, *146*, 2731–2755.
166. Clarisse, L.; Coheur, P.-F.; Prata, F.; Hadji-Lazaro, J.; Hurtmans, D.; Clerbaux, C. A Unified Approach to Infrared Aerosol Remote Sensing and Type Specification. *Atmos. Chem. Phys.* 2013, *13*, 2195–2221.
167. Bernath, P.; Boone, C.; Crouse, J. Wildfire Smoke Destroys Stratospheric Ozone. *Science* 2022, *375*, 1292–1295.

168. Guerrazi, H.; Sellitto, P.; Cuesta, J.; Eremenko, M.; Lachatre, M.; Mailler, S.; Carboni, E.; Salerno, G.; Caltabiano, T.; Menut, L.; et al. Quantitative Retrieval of Volcanic Sulphate Aerosols from IASI Observations. *Remote Sensing* 2021, *13*, 1808.
169. Boer, M.M.; Resco de Dios, V.; Bradstock, R.A. Unprecedented Burn Area of Australian Mega Forest Fires. *Nat. Clim. Chang.* 2020, *10*, 171–172.
170. Filkov, A.I.; Ngo, T.; Matthews, S.; Telfer, S.; Penman, T.D. Impact of Australia's Catastrophic 2019/20 Bushfire Season on Communities and Environment. Retrospective Analysis and Current Trends. *Journal of Safety Science and Resilience* 2020, *1*, 44–56.
171. Liu, X.; Bhartia, P.K.; Chance, K.; Spurr, R.J.D.; Kurosu, T.P. Ozone Profile Retrievals from the Ozone Monitoring Instrument. *Atmospheric Chemistry and Physics* 2010, *10*, 2521–2537.
172. Cuesta, J.; Eremenko, M.; Liu, X.; Dufour, G.; Cai, Z.; Höpfner, M.; von Clarmann, T.; Sellitto, P.; Forêt, G.; Gaubert, B.; et al. Satellite Observation of Lowermost Tropospheric Ozone by Multispectral Synergism of IASI Thermal Infrared and GOME-2 Ultraviolet Measurements over Europe. *Atmos. Chem. Phys.* 2013, *13*, 9675–9693.
173. Wang, P.; Tuinder, O.N.E.; Tilstra, L.G.; de Graaf, M.; Stammes, P. Interpretation of FRESCO Cloud Retrievals in Case of Absorbing Aerosol Events. *Atmospheric Chemistry and Physics* 2012, *12*, 9057–9077.
174. Ludewig, A.; Kleipool, Q.; Bartstra, R.; Landzaat, R.; Leloux, J.; Loots, E.; Meijering, P.; van der Plas, E.; Rozemeijer, N.; Vonk, F.; et al. In-Flight Calibration Results of the TROPOMI Payload on Board the Sentinel-5 Precursor Satellite. *Atmospheric Measurement Techniques* 2020, *13*, 3561–3580.
175. Tropomi Monitoring Portal Available online: <http://mps.tropomi.eu/dashboard> (accessed on 31 March 2022).
176. Torres, O.; Tanskanen, A.; Veihelmann, B.; Ahn, C.; Braak, R.; Bhartia, P.K.; Veefkind, P.; Levelt, P. Aerosols and Surface UV Products from Ozone Monitoring Instrument Observations: An Overview. *J. Geophys. Res.* 2007, *112*, doi:10.1029/2007jd008809.
177. <https://S5phub.Copernicus.Eu> Available online: <https://s5phub.copernicus.eu/> (accessed on 6 October 2021).
178. Platnick, S.; Meyer, K.; Wind, G.; Holz, R.E.; Amarasinghe, N. The NASA MODIS-VIIRS Continuity Cloud Optical Properties Products. *Remote sensing* 2021.
179. Earthdata Search Available online: <https://search.earthdata.nasa.gov/> (accessed on 31 March 2022).
180. Lutz, R.; Loyola, D.; García, S.G.; Romahn, F. OCRA Radiometric Cloud Fractions for GOME-2 on MetOp-A/B. *Atmospheric Measurement Techniques* 2016, *9*, 2357–2379.

181. Loyola, D.G.; Gimeno García, S.; Lutz, R.; Argyrouli, A.; Romahn, F.; Spurr, R.J.D.; Pedernana, M.; Doicu, A.; Molina García, V.; Schüssler, O. The Operational Cloud Retrieval Algorithms from TROPOMI on Board Sentinel-5 Precursor. *Atmospheric Measurement Techniques* 2018, *11*, 409–427.
182. Hersbach, H.; Bell, B.; Berrisford, P.; Hirahara, S.; Horányi, A.; Muñoz-Sabater, J.; Nicolas, J.; Peubey, C.; Radu, R.; Schepers, D.; et al. The ERA5 Global Reanalysis. *Quart. J. Roy. Meteor. Soc.* 2020, *146*, 1999–2049.
183. Rothman, L.S.; Gordon, I.E.; Babikov, Y.; Barbe, A.; Chris Benner, D.; Bernath, P.F.; Birk, M.; Bizzocchi, L.; Boudon, V.; Brown, L.R.; et al. The HITRAN2012 Molecular Spectroscopic Database. *J. Quant. Spectrosc. Radiat. Transf.* 2013, *130*, 4–50.
184. Chance, K.; Kurucz, R.L. An Improved High-Resolution Solar Reference Spectrum for Earth’s Atmosphere Measurements in the Ultraviolet, Visible, and near Infrared. *Journal of Quantitative Spectroscopy and Radiative Transfer* 2010, *111*, 1289–1295.
185. Sentinel-5p+ Innovation Available online: https://www.grasp-sas.com/projects/aod-brdf_sentinel-5p-innovation/ (accessed on 31 March 2022).
186. European Space Agency (ESA) *TROPOMI ATBD of the Directionally Dependent Surface Lambertian-Equivalent Reflectivity*;
187. Sarpong, E.; Smith, D.; Pokhrel, R.; Fiddler, M.N.; Bililign, S. Refractive Indices of Biomass Burning Aerosols Obtained from African Biomass Fuels Using RDG Approximation. *Atmosphere* 2020, *11*, 62.
188. Tran, H.; Boulet, C.; Hartmann, J.-M. Line Mixing and Collision-Induced Absorption by Oxygen in the A Band: Laboratory Measurements, Model, and Tools for Atmospheric Spectra Computations. *J. Geophys. Res.* 2006, *111*, doi:10.1029/2005jd006869.
189. ISRF Dataset Available online: <http://www.tropomi.eu/data-products/isrf-dataset> (accessed on 31 March 2022).
190. Spurr, R.; Christi, M. On the Generation of Atmospheric Property Jacobians from the (V)LIDORT Linearized Radiative Transfer Models. *J. Quant. Spectrosc. Radiat. Transf.* 2014, *142*, 109–115.
191. Peterson, D.A.; Fromm, M.D.; McRae, R.H.D.; Campbell, J.R.; Hyer, E.J.; Taha, G.; Camacho, C.P.; Kablick, G.P.; Schmidt, C.C.; DeLand, M.T. Australia’s Black Summer Pyrocumulonimbus Super Outbreak Reveals Potential for Increasingly Extreme Stratospheric Smoke Events. *npj Climate and Atmospheric Science* 2021, *4*.
192. Zoogman, P.; Liu, X.; Chance, K.; Sun, Q.; Schaaf, C.; Mahr, T.; Wagner, T. A Climatology of Visible Surface Reflectance Spectra. *J. Quant. Spectrosc. Radiat. Transf.* 2016, *180*, 39–46.
193. Clark, R.N.; Swayze, G.A.; Wise, R.A.; Livo, K.E.; Hoefen, T.M.; Kokaly, R.F.; Sutley, S.J. *USGS Digital Spectral Library Splib06a*; US Geological Survey, 2007;

194. Tikhonov, A.N. On the Solution of Ill-Posed Problems and the Method of Regularization. In Proceedings of the Doklady Akademii Nauk; Russian Academy of Sciences, 1963; Vol. 151, pp. 501–504.
195. Eremenko, M.; Dufour, G.; Foret, G.; Keim, C.; Orphal, J.; Beekmann, M.; Bergametti, G.; Flaud, J.-M. Tropospheric Ozone Distributions over Europe during the Heat Wave in July 2007 Observed from Infrared Nadir Spectra Recorded by IASI. *Geophys. Res. Lett.* 2008, 35, doi:10.1029/2008gl034803.
196. Bilal, M.; Qiu, Z.; Campbell, J.R.; Spak, S.N.; Shen, X.; Nazeer, M. A New MODIS C6 Dark Target and Deep Blue Merged Aerosol Product on a 3 Km Spatial Grid. *Remote Sensing* 2018, 10, 463.
197. Fernald, F.G. Analysis of Atmospheric Lidar Observations: Some Comments. *Appl. Opt.* 1984, 23, 652.
198. Fernald, F.G.; Herman, B.M.; Reagan, J.A. Determination of Aerosol Height Distributions by Lidar. *J. Appl. Meteorol. Climatol.* 1972, 11, 482–489.
199. Landgraf, J.; de Brugh, J.A.; Scheepmaker, R.; Borsdorff, T.; Hu, H.; Houweling, S.; Butz, A.; Aben, I.; Hasekamp, O. Carbon Monoxide Total Column Retrievals from TROPOMI Shortwave Infrared Measurements. *Atmospheric Measurement Techniques* 2016, 9, 4955–4975.
200. Borsdorff, T.; Hu, H.; Hasekamp, O.; Sussmann, R.; Rettinger, M.; Hase, F.; Gross, J.; Schneider, M.; Garcia, O.; Stremme, W.; et al. Mapping Carbon Monoxide Pollution from Space down to City Scales with Daily Global Coverage. *Atmospheric Measurement Techniques* 2018, 11, 5507–5518.
201. Giglio, L.; Schroeder, W.; Justice, C.O. The Collection 6 MODIS Active Fire Detection Algorithm and Fire Products. *Remote Sens. Environ.* 2016, 178, 31–41.
202. Nanda, S.; de Graaf, M.; Veefkind, J.P.; Sneep, M.; ter Linden, M.; Sun, J.; Levelt, P.F. A First Comparison of TROPOMI Aerosol Layer Height (ALH) to CALIOP Data. *Atmospheric Measurement Techniques* 2020, 13, 3043–3059.
203. Kloss, C.; Sellitto, P.; Von Hobe, M.; Berthet, G.; Smale, D.; Krysztofiak, G.; Xue, C.; Qiu, C.; Jégou, F.; Ouerghemmi, I.; et al. Australian Fires 2019–2020: Tropospheric and Stratospheric Pollution Throughout the Whole Fire Season, *Front. Environ. Sci.* 2021, 9, 10–3389.
204. Beringer, J.; Hutley, L.B.; Abramson, D.; Arndt, S.K.; Briggs, P.; Bristow, M.; Canadell, J.G.; Cernusak, L.A.; Eamus, D.; Edwards, A.C.; et al. Fire in Australian Savannas: From Leaf to Landscape. *Glob. Chang. Biol.* 2015, 21, 62–81.
205. Williams, R.J.; Gill, A.M.; Moore, P.H.R. Seasonal Changes in Fire Behaviour in a Tropical Savanna in Northern Australia. *Int. J. Wildland Fire* 1998, 8, 227–239.
206. Chaivaranont, W.; Evans, J.P.; Liu, Y.Y.; Sharples, J.J. Estimating Grassland Curing with Remotely Sensed Data. *Nat. Hazards Earth Syst. Sci.* 2018, 18, 1535–1554.

207. Remer, L.A.; Kaufman, Y.J.; Tanré, D.; Mattoo, S.; Chu, D.A.; Martins, J.V.; Li, R.-R.; Ichoku, C.; Levy, R.C.; Kleidman, R.G.; et al. The MODIS Aerosol Algorithm, Products, and Validation. *J. Atmos. Sci.* 2005, *62*, 947–973.
208. Menut, L.; Bessagnet, B.; Briant, R.; Cholakian, A.; Couvidat, F.; Mailler, S.; Pennel, R.; Siour, G.; Tuccella, P.; Turquety, S.; et al. The CHIMERE V2020r1 Online Chemistry-Transport Model. *Geoscientific Model Development* 2021, *14*, 6781–6811.
209. Bessagnet, B.; Hodzic, A.; Vautard, R.; Beekmann, M.; Cheinet, S.; Honoré, C.; Liousse, C.; Rouil, L. Aerosol Modeling with CHIMERE—Preliminary Evaluation at the Continental Scale. *Atmospheric Environment* 2004, *38*, 2803–2817.
210. Escribano, J.; Boucher, O.; Chevallier, F.; Huneus, N. Impact of the Choice of the Satellite Aerosol Optical Depth Product in a Sub-Regional Dust Emission Inversion. *Atmospheric Chemistry and Physics* 2017, *17*, 7111–7126.
211. Escribano, J.; Boucher, O.; Chevallier, F.; Huneus, N. Subregional Inversion of North African Dust Sources. *J. Geophys. Res. D: Atmos.* 2016, *121*, 8549–8566.
212. Garrigues; Remy; Chimot; Ades Monitoring Multiple Satellite Aerosol Optical Depth (AOD) Products within the Copernicus Atmosphere Monitoring Service (CAMS) Data Assimilation System. *Atmos. Clim. Sci.* 2022.
213. Bocquet, M.; Elbern, H.; Eskes, H.; Hirtl, M.; Žabkar, R.; Carmichael, G.R.; Flemming, J.; Inness, A.; Pagowski, M.; Pérez Camaño, J.L.; et al. Data Assimilation in Atmospheric Chemistry Models: Current Status and Future Prospects for Coupled Chemistry Meteorology Models. *Atmospheric Chemistry and Physics* 2015, *15*, 5325–5358.
214. Sayeed, A.; Eslami, E.; Lops, Y.; Choi, Y. CMAQ-CNN: A New-Generation of Post-Processing Techniques for Chemical Transport Models Using Deep Neural Networks. *Atmospheric Environment* 2022, *273*, 118961.
215. Xu, M.; Jin, J.; Wang, G.; Segers, A.; Deng, T.; Lin, H.X. Machine Learning Based Bias Correction for Numerical Chemical Transport Models. *Atmos. Environ.* 2021, *248*, 118022.
216. Jin; Lin; Segers; Xie Machine Learning for Observation Bias Correction with Application to Dust Storm Data Assimilation. *Chem. Phys. Lipids* 2019.
217. Rasp, S.; Lerch, S. Neural Networks for Postprocessing Ensemble Weather Forecasts. *Monthly Weather Review* 2018, *146*, 3885–3900.
218. Taillardat, M.; Mestre, O.; Zamo, M.; Naveau, P. Calibrated Ensemble Forecasts Using Quantile Regression Forests and Ensemble Model Output Statistics. *Mon. Weather Rev.* 2016, *144*, 2375–2393.
219. Nabavi, S.O.; Haimberger, L.; Abbasi, R.; Samimi, C. Prediction of Aerosol Optical Depth in West Asia Using Deterministic Models and Machine Learning Algorithms. *Aeolian Res.* 2018, *35*, 69–84.

220. Sayer, A.M.; Munchak, L.A.; Hsu, N.C.; Levy, R.C.; Bettenhausen, C.; Jeong, M.-J. MODIS Collection 6 Aerosol Products: Comparison between Aqua's e-Deep Blue, Dark Target, and "Merged" Data Sets, and Usage Recommendations. *J. Geophys. Res.* 2014, *119*, 13,965-13,989.
221. Wei, J.; Li, Z.; Peng, Y.; Sun, L. MODIS Collection 6.1 Aerosol Optical Depth Products over Land and Ocean: Validation and Comparison. *Atmos. Environ.* 2019, *201*, 428–440.
222. Gupta; Levy; Mattoo; Remer A Surface Reflectance Scheme for Retrieving Aerosol Optical Depth over Urban Surfaces in MODIS Dark Target Retrieval Algorithm. *Atmos. Clim. Sci.* 2016.
223. Hsu, N.C.; Jeong, M.-J.; Bettenhausen, C.; Sayer, A.M.; Hansell, R.; Seftor, C.S.; Huang, J.; Tsay, S.-C. Enhanced Deep Blue Aerosol Retrieval Algorithm: The Second Generation. *J. Geophys. Res.* 2013, *118*, 9296–9315.
224. Holben, B.N.; Eck, T.F.; Slutsker, I.; Tanré, D.; Buis, J.P.; Setzer, A.; Vermote, E.; Reagan, J.A.; Kaufman, Y.J.; Nakajima, T.; et al. AERONET—A Federated Instrument Network and Data Archive for Aerosol Characterization. *Remote Sens. Environ.* 1998, *66*, 1–16.
225. Menut, L.; Bessagnet, B.; Khvorostyanov, D.; Beekmann, M.; Blond, N.; Colette, A.; Coll, I.; Curci, G.; Foret, G.; Hodzic, A.; et al. CHIMERE 2013: A Model for Regional Atmospheric Composition Modelling. *Geoscientific Model Development* 2013, *6*, 981–1028.
226. Péré, J.C.; Mallet, M.; Pont, V.; Bessagnet, B. Evaluation of an Aerosol Optical Scheme in the Chemistry-Transport Model CHIMERE. *Atmospheric Environment* 2010, *44*, 3688–3699.
227. Skamarock, C.; Klemp, B.; Dudhia, J.; Gill, O.; Liu, Z.; Berner, J.; Wang, W.; Powers, G.; Duda, G.; Barker, D.; et al. A Description of the Advanced Research WRF Model Version 4.1. 2019, doi:10.5065/1dfh-6p97.
228. Alfaro, S.C.; Gomes, L. Modeling Mineral Aerosol Production by Wind Erosion: Emission Intensities and Aerosol Size Distributions in Source Areas. *Journal of Geophysical Research: Atmospheres* 2001, *106*, 18075–18084.
229. Gama, C.; Ribeiro, I.; Lange, A.C.; Vogel, A.; Ascenso, A.; Seixas, V.; Elbern, H.; Borrego, C.; Friese, E.; Monteiro, A. Performance Assessment of CHIMERE and EURAD-IM' Dust Modules. *Atmospheric Pollution Research* 2019, *10*, 1336–1346.
230. Menut, L.; Chiapello, I.; Moulin, C. Previsibility of Saharan Dust Events Using the CHIMERE-DUST Transport Model. *IOP Conference Series: Earth and Environmental Science* 2009, *7*, 012009.
231. Abdoul Aziz Saidou Chaibou, Xiaoyan Ma, Kanike Raghavendra Kumar, Hailing Jia, Yaoguo Tang, Tong Sha Evaluation of Dust Extinction and Vertical Profiles Simulated by WRF-Chem with CALIPSO and AERONET over North Africa. *J. Atmos. Sol. Terr. Phys.* 2020, *199*, 105213.

232. Washington, R.; Bouet, C.; Cautenet, G.; Mackenzie, E.; Ashpole, I.; Engelstaedter, S.; Lizcano, G.; Henderson, G.M.; Schepanski, K.; Tegen, I. Dust as a Tipping Element: The Bodélé Depression, Chad. *Proceedings of the National Academy of Sciences* 2009, *106*, 20564–20571.
233. Bittner, L. R. Bellman, Adaptive Control Processes. A Guided Tour. XVI 255 S. Princeton, N. J., 1961. Princeton University Press. Preis Geb. \$ 6.50. *ZAMM - Zeitschrift für Angewandte Mathematik und Mechanik* 1962, *42*, 364–365.
234. <https://www.python.org/> Available online: <https://www.python.org/> (accessed on 22 July 2022).
235. <https://jupyterbook.org> (accessed on 22 July 2022).
236. Pedregosa; Varoquaux; Gramfort Scikit-Learn: Machine Learning in Python. *of machine Learning ...* 2011.
237. Bengio, Y. Learning Deep Architectures for AI. *Foundations and Trends® in Machine Learning* 2009, *2*, 1–127.
238. Rumelhart, D.E.; Hinton, G.E.; Williams, R.J. Learning Representations by Back-Propagating Errors. *Nature* 1986, *323*, 533–536.
239. MacKay, D.J.C. A Practical Bayesian Framework for Backpropagation Networks. *Neural Comput.* 1992, *4*, 448–472.
240. Dubey, S.R.; Singh, S.K.; Chaudhuri, B.B. Activation Functions in Deep Learning: A Comprehensive Survey and Benchmark. *Neurocomputing* 2022, *503*, 92–108.
241. Dauphin; Pascanu; Gulcehre Identifying and Attacking the Saddle Point Problem in High-Dimensional Non-Convex Optimization. *Adv. Neural Inf. Process. Syst.*
242. Goodfellow, I.J.; Vinyals, O.; Saxe, A.M. Qualitatively Characterizing Neural Network Optimization Problems. *arXiv [cs.NE]* 2014.
243. Zhou, Y.; Yang, J.; Zhang, H.; Liang, Y.; Tarokh, V. SGD Converges to Global Minimum in Deep Learning via Star-Convex Path. *arXiv [cs.LG]* 2019.
244. Du, S.; Lee, J.; Li, H.; Wang, L.; Zhai, X. Gradient Descent Finds Global Minima of Deep Neural Networks. In Proceedings of the Proceedings of the 36th International Conference on Machine Learning; Chaudhuri, K., Salakhutdinov, R., Eds.; PMLR, 09--15 Jun 2019; Vol. 97, pp. 1675–1685.
245. Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G.S.; Davis, A.; Dean, J.; Devin, M.; et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. *arXiv [cs.DC]* 2016.
246. O'Malley, T.; Bursztein, E.; Long, J.; Chollet, F.; Jin, H.; Invernizzi, L.; Others Keras Tuner. Retrieved May 2019, *21*, 2020.

247. Ioffe, S.; Szegedy, C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. In Proceedings of the Proceedings of the 32nd International Conference on Machine Learning; Bach, F., Blei, D., Eds.; PMLR: Lille, France, 07--09 Jul 2015; Vol. 37, pp. 448–456.
248. Fukushima, K. Cognitron: A Self-Organizing Multilayered Neural Network. *Biol. Cybern.* 1975, 20, 121–136.
249. Hinton, G.E.; Srivastava, N.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R.R. Improving Neural Networks by Preventing Co-Adaptation of Feature Detectors. *arXiv [cs.NE]* 2012.
250. Srivastava; Hinton; Krizhevsky Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *The journal of machine.*
251. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv [cs.LG]* 2014.
252. Breiman, L. *Machine Learning* 2000, 40, 229–242.
253. Gordon, A.D.; Breiman, L.; Friedman, J.H.; Olshen, R.A.; Stone, C.J. Classification and Regression Trees. *Biometrics* 1984, 40, 874.
254. Beauchamp, J. <https://commons.wikimedia.org> Available online: https://commons.wikimedia.org/wiki/File:Decision_Tree_vs._Random_Forest.png (accessed on 23 October 2022).
255. Friedman, J.H. Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics* 2001, 29.
256. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; Association for Computing Machinery: New York, NY, USA, August 13 2016; pp. 785–794.
257. Olson, R.S.; Moore, J.H. TPOT: A Tree-Based Pipeline Optimization Tool for Automating Machine Learning. In Proceedings of the Proceedings of the Workshop on Automatic Machine Learning; Hutter, F., Kotthoff, L., Vanschoren, J., Eds.; PMLR: New York, New York, USA, June 24 2016; Vol. 64, pp. 66–74.
258. <https://Xgboost.Readthedocs.io/En/Stable/Parameter.Html> Available online: <https://xgboost.readthedocs.io/en/stable/parameter.html> (accessed on 2 August 2022).
259. Worldview: Explore Your Dynamic Planet Available online: <https://worldview.earthdata.nasa.gov/> (accessed on 31 January 2023).
260. Xu, K.; Zhang, M.; Li, J.; Du, S.S.; Kawarabayashi, K.-I.; Jegelka, S. How Neural Networks Extrapolate: From Feedforward to Graph Neural Networks. *arXiv [cs.LG]* 2020.

261. Lundberg, S.; Lee, S.-I. A Unified Approach to Interpreting Model Predictions. *arXiv [cs.AI]* 2017.
262. Drake, N.; Bristow, C. Shorelines in the Sahara: Geomorphological Evidence for an Enhanced Monsoon from Palaeolake Megachad. *Holocene* 2006, *16*, 901–911.
263. Chudnovsky, A.; Kostinski, A.; Herrmann, L.; Koren, I.; Nutesku, G.; Ben-Dor, E. Hyperspectral Spaceborne Imaging of Dust-Laden Flows: Anatomy of Saharan Dust Storm from the Bodélé Depression. *Remote Sens. Environ.* 2011, *115*, 1013–1024.
264. Hsu, N.C.; Tsay, S.-C.; King, M.D.; Herman, J.R. Aerosol Properties over Bright-Reflecting Source Regions. *IEEE Transactions on Geoscience and Remote Sensing* 2004, *42*, 557–569.
265. Rocha-Lima; Martins; Remer A Detailed Characterization of the Saharan Dust Collected during the Fennec Campaign in 2011: In Situ Ground-Based and Laboratory Measurements. *Atmos. Clim. Sci.* 2018.