Inverse modeling of nitrogen oxides emissions in urbanized regions of the Northern Hemisphere : Application of the CIF-CHIMERE inversion system to OMI and TROPOMI observations

Dilek Savas

Dilek Savas. Inverse modeling of nitrogen oxides emissions in urbanized regions of the Northern Hemisphere : Application of the CIF-CHIMERE inversion system to OMI and TROPOMI observations. Environmental Sciences. Université Paris Cité, 2023. English. ⟨NNT : ⟩. ⟨tel-04349344⟩

HAL Id: tel-04349344
https://hal.science/tel-04349344v1

Submitted on 17 Dec 2023

Université Paris Cité

Ecole doctorale des Sciences de l'Environnement d'Ile de France (ED129)

Laboratoire Interuniversitaire des Systèmes Atmosphériques (LISA UMR7583)

Inverse modeling of nitrogen oxides emissions in urbanized regions of the Northern Hemisphere : Application of the CIF-CHIMERE inversion system to OMI and TROPOMI observations

Par Dilek Savas

Thèse de doctorat en Météorologie et Qualité de l'Air

Dirigée par Gaëlle Dufour

Présentée et soutenue publiquement le 8 Juin 2023

Devant un jury composé de :

Matthias Beekmann, Directeur de recherche, CNRS & Université Paris Cité & Université Paris Est Créteil, Président du jury
Claire Granier, Directrice de recherche, CNRS & Université de Toulouse III, Rapporteure
Vincent Guidard, Directeur de recherche, Météo-France, Rapporteur
Trissevgeni Stavrakou, Directrice de recherche, BIRA-IASB, Examinatrice
Solène Turquety, Professeur, Sorbonne Université, Examinatrice
Gaëlle Dufour, Directrice de recherche, CNRS & Université Paris Cité & Université Paris Est Créteil, Directrice de thèse

Titre : Modélisation inverse des émissions d'oxydes d'azote dans les zones urbanisées de l'hémisphère Nord : Application du système d'inversion CIF-CHIMERE aux observations OMI et TROPOMI

Résumé : Selon l'Organisation mondiale de la santé, l'exposition à la pollution de l'air ambiant serait à l'origine de 4.2 millions de décès prématurés par an dans le monde. Ce problème environnemental majeur est apparu au début du XXe siècle en raison de l'industrialisation et de l'urbanisation intenses des sociétés, obligeant les décideurs et les gouvernements à contrôler les émissions anthropiques de polluants atmosphériques. Depuis 1990, des initiatives importantes ont été prises dans les pays industrialisés, comme l'Europe, pour réduire les émissions de polluants en mettant en œuvre des stratégies d'atténuation. Plus récemment, la mise en œuvre de politiques d'atténuation a commencé à être observée dans des pays en développement comme la Chine.

L'évaluation des politiques de réduction des émissions repose souvent sur des inventaires officiels, établis à partir des déclarations d'activités des émetteurs et de données statistiques. Cette approche dite "bottom up" estime les émissions en extrapolant les mesures des facteurs d'émission qui ne sont disponibles qu'à des résolutions spatiotemporelles réduites. Cette méthode conduit des incertitudes élevées, car elle ne tient pas compte de la grande variabilité spatio-temporelle des flux d'émission. En outre, les incertitudes liées aux inventaires des émissions entraînent des incohérences dans l'évaluation des stratégies les plus appropriées pour prévenir les épisodes de pollution atmosphérique. Par conséquent, il est important de pouvoir suivre avec précision l'évolution des émissions et la tendance des polluants pour évaluer les politiques de réduction. L'un des moyens prometteurs pour cela est l'approche par inversion atmosphérique, qui utilise les observations atmosphériques disponibles pour contraindre les émissions par le biais de la modélisation atmosphérique et de méthodes d'inversion. Cette approche dite "top down" apporte des informations complémentaires aux inventaires "bottom up". Elle estime les émissions tout en exploitant la grande variabilité spatio-temporelle des observations satellitaires et la puissance de calcul.

Les oxydes d'azote (NO_x = $NO+NO_2$) sont parmi les polluants les plus réglementés en tant que précurseurs d'autres polluants atmosphériques, tels que l'ozone et les aérosols secondaires. Dans le cadre de cette thèse, nous avons tout d'abord mis en place et testé la capacité d'inversion des émissions quotidiennes de NO_x du système de modélisation inverse CIF, couplé au modèle de chimie-transport CHIMERE et son adjoint, à une résolution modérée, en utilisant les observations du satellite OMI. Les résultats nous ont permis de déterminer les paramètres et les sensibilités du système CIF pour les inversions de NO_x.

Le système CIF-CHIMERE a d'abord été appliqué pour évaluer l'impact des fortes réglementations sur les émissions de NO_x mises en place par la Chine depuis 2011. Nous avons assimilé les observations de NO_2 des instruments du satellite OMI et estimé les émissions de NO_x pour 2015 et 2019 avec une résolution de 50x50 km^2 sur la Chine. L'année 2010 a été choisie comme base de référence ou inventaire a priori, car elle se situe juste avant la mise en œuvre des réductions d'émissions. Les résultats montrent que la réduction des émissions de NO_x est limitée aux zones urbanisées et industrialisées tout en restant dans les limites des objectifs d'atténuation (10-15 %). Les émissions de NO_x estimées ont également été utilisées pour simuler les concentrations de NO_2 en surface, et permettent d'améliorer légèrement la comparaison avec les mesures au sol.

Le système CIF-CHIMERE a également été appliqué pour évaluer les changements abrupts des émissions de NO_x en Europe causés par la pandémie de COVID-19. En particulier, nous avons exploré le potentiel d'assimilation des observations à haute résolution de TROPOMI NO_2 pendant la période de confinement de 2020.

Mots clefs : Pollution Atmosphérique, Qualité de l'Air, Émission de NOx, Assimilation de Données, Modélisation Inverse, Oxydes d'Azote

Title : Inverse modeling of nitrogen oxides emissions in urbanized areas of the Northern Hemisphere: Application of the CIF-CHIMERE inversion system to OMI and TROPOMI observations

Abstract : According to the World Health Organization, exposure to ambient air pollution is estimated to cause 4.2 premature deaths per year in urban and rural areas worldwide. Poor air quality also leads to damage to infrastructure, soil and water resources. This major environmental problem emerged at the beginning of the 20th century due to the intense industrialization and urbanization of societies, forcing policy makers and governments to control anthropogenic emissions of air pollutants. Since 1990, important initiatives have been taken in industrialized countries, such as Europe, to reduce emissions of pollutants by implementing mitigation strategies. More recently, the implementation of mitigation policies has started to be seen in developing countries such as China.

The assessment of emission reduction policies is often based on official emission inventories derived from emitters' statements of activities and statistical data. This so-called bottom-up approach estimates emissions by extrapolating emission factor measurements that are only available in a sparse spatial and temporal network. This involves high uncertainties, as it does not incorporate the high spatiotemporal variability of emission fluxes. Besides, uncertainties in emission inventories lead to inconsistencies in the assessment of appropriate strategies to prevent air pollution episodes. Therefore, being able to accurately monitor the development of emissions and the trend of pollutants is an important issue for evaluating reduction policies. One of the promising ways to overcome this problem is the atmospheric inversion approach, which uses available atmospheric observations to constrain emissions through atmospheric modeling and the inversion system. This so-called top down approach brings complementary information to bottom-up inventories. It estimates emissions while exploiting the high spatiotemporal variability of the satellite observations and computational power.

Nitrogen oxides ($NO_x = NO+NO_2$) are among the most regulated pollutants as precursors of other air pollutants, such as ozone and secondary aerosols. In the framework of this thesis, first, we set up and tested the daily NO_x emission inversion capability of the state-of-art inverse modeling system CIF, embedded with the CHIMERE Chemistry Transport Model and its adjoint at moderate resolution using OMI satellite observations. The results lead us to determine the settings and sensitivities of the CIF system for the NO_x inversions.

The CIF-CHIMERE system was applied first to evaluate the impact of strong NO_x emission regulations implemented by China since 2011 within its 5-Year Plans. We assimilated NO_2 observations from OMI satellite instruments and estimated NO_x emissions for 2015 and 2019 with a resolution of 50x50 km^2 over Eastern China. The year 2010 was chosen as a baseline or a priori bottom-up inventory, as it was just before the mitigation implementation. The results show that the reduction in NO_x emissions is limited to urbanized and industrialized areas but remains within the mitigation targets (10-15 %). The estimated NO_x emissions were also used to simulate NO_2 surface concentrations, and the inversion approach was found to improve the comparison with ground-based measurements slightly.

The CIF-CHIMERE system was also applied to assess the abrupt changes in NO_x emissions in Europe caused by the COVID-19 pandemic. In particular, we explored the potential of assimilating high-resolution TROPOMI NO_2 observations during the 2020 lock-down period.

Keywords : Atmospheric Pollution, Air Quality, NOx Emission, Data Assimilation, Inverse Modeling, Nitrogen Oxides

Acknowledgements

I would like to express my deep gratitude to everyone who has made meaningful contributions to my academic and personal development during this time. Above all, I sincerely thank my supervisor Gaëlle Dufour for her valuable guidance, profound expertise and feedback throughout the whole process. Her support and encouragement were of critical importance to my success, and I am deeply grateful for her mentorship.

I am also grateful to the committee members; Vincent Guidard, Claire Granier, Trissevgeni Stavrakou, Matthias Beekmann, and Solène Turquety for their insightful feedback and constructive criticism, which have contributed additional perspectives to my research.

I would like to thank the agencies that supported this research; the French National Research Agency (ANR) and Carole Daniel from the French National Centre for Space Studies (CNES) for the doctoral grant. Their financial assistance has made this research possible, and I am deeply grateful for their support.

I also would like to thank Adriana Coman for her continuous support and generous knowledge sharing. It has always been a great pleasure for me to talk to her and learn from her expertise. Without her support, much of this work would not have taken the shape it has. I am honored to have had the privilege of working with her.

I would like to acknowledge the support and encouragement of my colleagues at the LISA, and our collaborators at the LSCE within the ANR-ARGONAUT project for their help and scientific insights.

Additionally, I am grateful to my family and friends who have been a constant source of love, support, and encouragement throughout my academic journey. Their unwavering support, understanding, and patience have been critical in helping me stay focused and motivated especially during difficult times such as COVID-19 pandemic.

I would like to give special thanks to my mom and dad, Vildan-Metin Savas, for their unconditional love and tremendous support over the years to help me pursue my dreams.

Last but not least, I would like to express my sincere gratitude to Atahan Atadil, my dear life partner who has always made me feel his support, even from thousands of kilometers away. I cannot wait to explore the universe on our starship.

Thank you all! This journey would not have been possible without your support and encouragement, and I am deeply grateful for your contributions.

“*On ne découvre pas de terre nouvelle sans consentir à perdre de vue, d'abord et longtemps, tout rivage.”* – André Gide

Dedicated to those brave enough to challenge themselves beyond their comfort zones.

Abbreviations

CH_3CHO	acetaldehyde
AMF	air mass factor
AN	alkyl nitrates
ARGONAUT	pollutAnts and gReenhouse Gases emissiOns monitoring from spAce at high resolUTion
AKs	averaging kernels
BLUE	best linear unbiased estimator
CO	carbon monoxide
CO_2	carbon dioxide
CNES	Centre national d'études spatiales
CCD	charge coupled device
CTMs	chemistry transport models
CNEMC	China National Environmental Monitoring Centre
CAAQS	Chinese ambient air quality standards
UTC	Coordinated Universal Time
CAMS	Copernicus Atmosphere Monitoring Service
CFCs	chlorofluorocarbons
CLAES	Cryogenic Limb Array Etalon Spectrometer
°E	degrees east
°N	degrees north
DOAS	Differential Optical Absorption Spectroscopy
EOS	Earth Observing System
EEA	European Environmental Agency
ENIAC	Electronic Numerical Integrator and Computer
ECCAD	Emissions of atmospheric compounds and compilation of ancillary data
EDGAR	Emissions Database for Global Atmospheric Research
enKF	ensemble Kalman filter
4D-EnVar	Ensemble 4D Variational
EEA	European Environment Agency
ESA	European Space Agency

e.g.	exempli gratia
EPA	Environmental Protection Agency
Eq	equation
EU	European Union
ECMWF	European Centre for Medium-Range Weather Forecasts
EKF	extended Kalman filter
FMI	Finnish Meteorological Institute
4D-Var	four dimensional variational
GLCF	Global Land Cover Facility
GOME	Global Ozone Monitoring Experiment
GHG	greenhouse gas
hPa	hectopascal
HTAP	Hemispheric Transport of Air Pollution
HO_2	hydroperoxyl radical
OH	hydroxyl radical
INERIS	Institut National de l'Environnement Industriel et des Risques
INCA	Interaction with Chemistry and Aerosols
AR4	IPCC Forth Assessment Report
TAR	IPCC Third Assessment Report
C_5H_8	isoprene
KF	Kalman filter
LMD	Laboratoire de Météorologie Dynamique
LMDz	Laboratoire de Météorologie Dynamique Zoom
LISA	Laboratoire Interuniversitaire des Systèmes Atmosphériques
lat	latitude
LT	local time
lon	longitude
CH_4	methane
MEE	Ministry of Ecology and Environment
MEGAN	Model of Emissions of Gases and Aerosols from Nature
MEIC	Multiresolution Emission Inventory for China
NASA	National Aeronautics and Space Administration
NAAQS	national ambient air quality standards

NIR	near infrared
NIVR	Netherlands’ Institute for Air and Space Development
NO_3	nitrate radical
HNO_3	nitric acid
NO	nitric oxide
N_2	nitrogen
NO_2	nitrogen dioxide
NO_x	nitrogen oxides
N_2O	nitrous oxide
NMVOCs	non methane volatile organic compounds
NMB	normalized mean bias
NGST	Northrop Grumman Space Technology
NWP	numerical weather prediction
OFFL	offline
OI	optimal interpolation
RO_2	organic peroxy radical
OECD	Organization for Economic Co-operation and Development
O_2	oxygen
O_3	ozone
OMI	Ozone Monitoring Instrument
PDEs	partial differential equations
$CH_3C(O)OO$	peroxyacetyl radical
PANs	peroxyacetyl nitrates
parts per billion	ppb
PDFs	probability density functions
QA4ECV	Quality Assurance for Essential Climate Variables
REAS	Regional Emission Inventory in Asia
RMSE	root mean square error
SCIAMACHY	Scanning Imaging Absorption Spectrometer for Atmospheric Cartography
SOAs	secondary organic aerosols
S5P	Sentinel-5 Precursor
S5P-PAL	Sentinel-5P Product Algorithm Laboratory

SCD	slant column density
SNR	signal to noise ratio
SD	standard deviation
SNAP	Standardized Nomenclature for Air Pollutants
SO_2	sulfur dioxide
SAOZ	Système d'Analyse par Observation Zénitale
TLH	tangent linear hypothesis
Tg N yr^{-1}	teragram nitrogen per year
3D	three dimension
TOMS	Total Ozone Mapping Spectrometer
TGCC	Très Grand Centre de Calcul
TEMIS	Tropospheric Emission Monitoring Internet Service
TROPOMI	TROPOspheric Monitoring Instrument
USA	United States of America
UARS	Upper Atmosphere Research Satellite
ultraviolet-visible	UV-Vis
Var	variational
v	version
VCD	vertical column density
VOCs	volatile organic carbons
WHO	World Health Organization
WMO	World Meteorological Organization

Contents

Introduction

According to the World Health Organization, exposure to ambient air pollution is estimated to cause 4.2 million premature deaths per year in urban and rural areas worldwide. Poor air quality also leads to damage to infrastructures, soil, and water resources. This major environmental problem emerged at the beginning of the 20th century due to the intense industrialization and urbanization of societies, forcing policymakers and governments to control anthropogenic emissions of air pollutants. Since 1990, important initiatives have been taken in industrialized countries, such as Europe, to reduce emissions of pollutants by implementing mitigation strategies. More recently, the implementation of mitigation policies has started to be seen in developing countries such as China.

Among the various air pollutants, nitrogen oxides ($NO_x = NO + NO_2$) are considered one of the major pollutants. Besides their direct impact on health, such as respiratory infections, asthma, and chronic lung diseases, and on ecosystems and cultures as damaging vegetation, NO_x are precursors of other air pollutants, such as ozone and secondary aerosols, which also have adverse effects on human and animal health, vegetation, and climate by being short-lived climate forcers. Their dissolution in water leads to acids, and their transformation into nitrate particulate also damages vegetation, buildings, and materials and contributes to acidification and eutrophication of aquatic and terrestrial ecosystems. Then, NO_x, and especially NO_2, are among the most regulated pollutants. In Europe, road transport and energy (production, distribution and industrial use) are the main emission sectors followed by tertiary and residential sectors, agriculture and non-road transport (in 2019). According to the European Environmental Agency, European NO_x emissions have been reduced by almost 60 % between 1990 and 2019. In China, following a significant increase in NO_x emissions between the 90s and late 2000s due to rapid economic development, the 12th and 13th 5-Year government plans in 2011 and 2016, respectively, decided to implement NO_x emission reductions and set targets of 10 % reduction in 2015 and 15 % reduction in 2020. In China, electricity generation and industrialization, which depend heavily on coal consumption, are the main contributor and responsible for 63 % of total NO_x emissions, while road vehicles are the second largest contributor to NO_x emissions by 31 %.

In parallel with the mitigation strategies, observational capabilities of pollutants and especially NO_2 concentrations have largely increased during the last three decades with the development of air quality measurement networks in Europe and more recently in China sampling various types of situations (background, rural, mountains, suburban, urban, traffic). In the meantime, satellite observations of NO_2 tropospheric columns have shown unprecedented improvements from the GOME satellite instrument with a spatial resolution of a few hundreds of kilometers in the mid-90s to OMI and GOME-2 (resolution of a few tens of kilometers) in the mid-2000s and TROPOMI (resolution of a few kilometers) since 2018. The satellite observations provide a global description of NO_2 distribution, now at high resolution, all over the globe once a day. These developments brought new insights for pollution and emission monitoring from the regional to the global scales. In addition, atmospheric modeling capabilities have also largely developed following the increase in computational power capabilities. Nowadays, chemical transport models are used routinely and operationally to simulate the tridimensional distributions of pollutant concentrations from local to global with a high spatial and temporal resolution. Operational systems, such as the Copernicus Atmospheric Monitoring Service (CAMS) in Europe or the Prev'Air national platform in France, manage to provide estimates of air pollutants at various scales by exploiting the complementarity of modeling and observations and the development of powerful data assimilation systems.

Once emission mitigations are implemented, one key issue is to evaluate their effectiveness and their efficiency. The assessment of emission reduction policies is often based on official emission inventories derived from emitters’ statements of activities and statistical data. This so-called bottom-up approach estimates emissions by extrapolating emission factor measurements that are only available in a sparse spatial and temporal network. This involves high uncertainties, as it does not incorporate the high spatiotemporal variability of emission fluxes. Besides, uncertainties in emission inventories lead to inconsistencies in the assessment of appropriate strategies to prevent air pollution episodes. Therefore, being able to monitor accurately the development of emissions and the trend of pollutants is an important issue for evaluating reduction policies. One of the promising ways to overcome this problem is the atmospheric inversion approach, which uses available atmospheric observations, such as satellite observations, to constrain emissions through atmospheric modeling and an inversion system (data assimilation). This so-called top-down approach brings complementary information to bottom-up inventories. It estimates emissions while exploiting the high spatiotemporal variability of the satellite observations and computational power with a fast update capability. Inverse modeling approaches have acquired a good level of maturity for the last decade. At the French level, the Community Inversion Framework (CIF), firstly dedicated to the inversion and monitoring of greenhouse gases at the global scale, has been in development during the last few years. Developing and testing its capabilities to perform atmospheric inversions of highly reactive species at the regional scale and a high resolution was one of the objectives of the ARGONAUT ANR project and the framework of this PhD work.

In this context, the first chapter of the thesis describes the current state of knowledge about NO_2 air pollutants, their sources, emission inventories, and the purpose of this thesis. The second chapter explains all the necessary tools and instruments to be used in inversions. It is where we set up and tested the daily NO_x emission inversion capability of the CIF inverse modeling system, embedded with the CHIMERE Regional Chemistry Transport Model and its adjoint at moderate and high-resolution using OMI and TROPOMI satellite observations. In the third chapter, the CIF-CHIMERE system was applied first to evaluate the impact of the NO_x emission regulations implemented by China since 2011 within its 5-Year Plans. We assimilated NO_2 observations from OMI satellite instruments and estimated NO_x emissions for 2015 and 2019 with a resolution of 50×50 km^2 over Eastern China. The year 2010 was chosen as a baseline or a priori bottom-up inventory, as it was just before the mitigation implementation. In the fourth chapter, the CIF-CHIMERE system was applied to assess the abrupt changes in NO_x emissions in Europe caused by the COVID-19 pandemic. In particular, we explored the potential of assimilating high-resolution TROPOMI NO_2 observations during the 2020 lock-down period. The conclusion summarizes the main results of the thesis and discusses the possible perspectives of this work.

First part : Context

Chapter 1 : Nitrogen oxides in the atmosphere

Figure 1 Parisian Sunset at the top of the Montparnasse tower (December 2021).

1.1 Earth’s atmospheric composition and structure

The atmosphere is the air mass surrounding a planet or other celestial body. The name originates from two ancient Greek words, “atmós” and “sphaîra,” meaning vapor and sphere. Earth’s atmosphere (air) is one of the four Earth subsystems, as well as the lithosphere (land), the hydrosphere (water), and the biosphere (living beings). It is invisible, odorless, and essential to the survival of every life form on the Earth. The components of the Earth’s atmosphere (hereinafter only atmosphere) have evolved since the formation of the Earth, and today, nitrogen (78.1 %), oxygen (20.9 %), and a small amount of argon (0.9 %) account for over 99.9 % of the air we breathe. The remaining 0.1 % consists of trace constituents relevant to climate and human health, e.g., water vapor, carbon dioxide, methane, ozone, and nitrogen oxides. Besides gaseous species, solid or liquid particles of different sizes, known as aerosols, are also found in the atmosphere e.g., dust, sea salt, fog, and fume.

The atmosphere is composed of several interconnected layers, which have different characteristics mainly due to vertical variations in temperature and pressure (Figure 2).

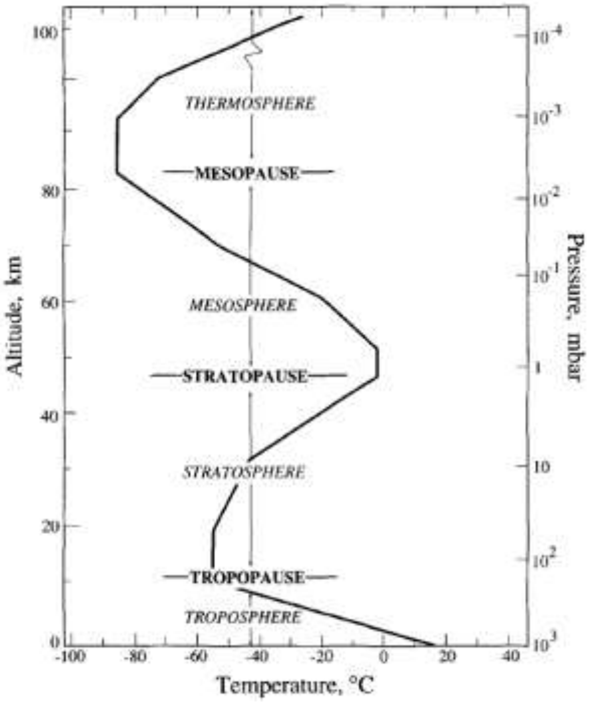


Figure 2 Atmospheric layers (Image taken from Seinfeld and Pandis, 2016)

The troposphere is the lowest layer from the Earth’s surface to the tropopause at an altitude of 10-15 km. The upper-boundary of the troposphere (so-called tropopause) varies with the season (lower in winter, higher in summer) and the latitude (lowest over the poles, highest at the equator). It contains 80 % of the total atmospheric mass and almost all atmospheric water vapor, including clouds. It is, therefore, the layer in which most of the Earth's meteorology occurs. It is also, where air pollutants are emitted, vegetation grows, and interactions with the oceans occur. In the troposphere, the warmest part is the lowest area due to the incoming solar radiation absorbed by the Earth’s surface and transferred into the heat. However, the temperature decreases by 6.5°C per km with an increasing altitude until the tropopause. In addition, this vertical temperature gradient allows for vertical mixing and meteorological disturbances.

Above the tropopause, the stratosphere stretches from about 11 km to the stratopause at 50 km. The low temperature of the tropopause acts like a barrier to water being transported into the stratosphere, which makes the stratosphere quite dry compared to the troposphere. The stratosphere contains approximately 10-20 % of the total atmospheric mass. Within the stratosphere, there is the ozone (O_3) layer produced and destroyed naturally (also as a result of human activities) by incoming ultraviolet solar radiation interacting with molecular and atomic oxygen while releasing heat. This phenomenon determines the thermal structure of the stratosphere as the temperature progressively increases from 20 to 50 km until the stratopause. One of the biggest discoveries of the last century and still a hot topic is the degradation of the ozone layer in the Polar Regions due to the emissions of chlorofluorocarbons (CFCs) because of human activities (Farman, Gardiner, and Shanklin 1985; Solomon et al. 1986). Indeed, CFCs, which are entirely human-made, have caused a 'hole in the ozone layer' by breaking the natural cycle of ozone production in the stratosphere. The implementation of agreements such as the Montreal Protocol, signed in 1987, which limits the use of these ozone depletion gases, has resulted with the recovery of the ozone hole, although full recovery is not expected until the end of this century according to the projections (Osani and Haysmith 2023; World Meteorological Organization (WMO) 2018).

Above the stratopause, the mesosphere lies until the mesopause around 80 km, which is the coldest region in the atmosphere with a temperature of less than $-80°C$.

Above the mesopause, the thermosphere lies and is bounded by the thermopause between the altitudes 500 to 1000 km. The temperature increases at the lower levels of the thermosphere because of the absorption of the short wave solar radiation by nitrogen and oxygen at these levels. International Space Station (at 400 km) and low orbit artificial satellites (e.g., NASA Aura at 705 km, Sentinel-5P at 824 km) reside at this layer.

Atop of the thermopause, the exosphere lies, extending into the near vacuum of outer space.

1.2 Atmospheric lifetime

The concentration of the chemical gases and particles in the atmosphere is either directly determined by emissions (through anthropogenic or natural processes), depositions, and atmospheric transport, or indirectly by atmospheric chemical reactions (via production and loss). Atmospheric lifetime, that is the time for a molecule to be converted into another substance by atmospheric chemical reactions (or the time the molecule remains in a steady state), is defined as an indicator of the characteristic of a molecule. Depending on their atmospheric lifetime, molecules show spatial and temporal variabilities. Some of them are chemically inert (or stable) and identified as long-lived species such as methane (CH_4), nitrous oxide (N_2O) and CFCs, whose atmospheric lifetimes are around 10 years for CH_4 and 100 years for N_2O and CFCs (Figure 3). This means that they have enough time to be transported away from their sources over long distances (on a global scale) and at higher altitudes before they chemically interact with other chemical substances and break down (e.g., stratospheric O_3 depletion due to N_2O, (Portmann, Daniel, and Ravishankara 2012).

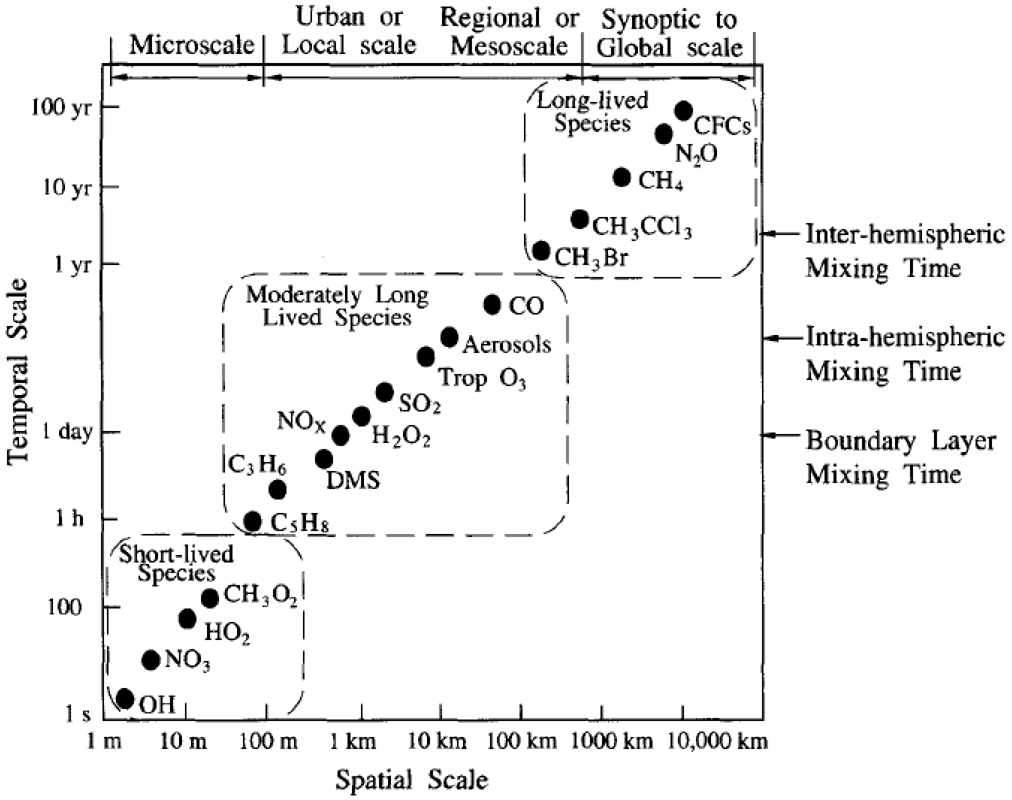


Figure 3 Atmospheric lifetime of different species from short-lived to long-lived species (Image taken from Seinfeld and Pandis, 2016)

Therefore, they are much more mixed in the atmosphere and spread over the entire Earth. On the other hand, substances with a relatively shorter lifetime do not find a chance to be transported far from their emission sources. These short-lived species can be detected close to their emission source compared to the long-lived species before interacting with other substances. Hydroxyl radical (OH) is the freest reactive radical and a prominent short-lived species with a lifetime only of seconds. It reacts with many species and controls the oxidizing capacity of the atmosphere. Other species such as Isoprene (C_5H_8), Nitrogen Oxides (NO_x) and Sulfur Dioxide (SO_2) have lifetimes from hours to days. They are mainly found at the lower levels of the troposphere in urban and regional scales and considered as air pollutants. They are important for the assessment of air quality.

Atmospheric lifetime of the species depends on various chemical and physical factors such that the presence of oxidizers (e.g. O_3, OH), the amount of incoming solar radiation and meteorological conditions (e.g., humidity, temperature). For instance, the amount of incoming solar energy (so-called insolation) differs from tropics to higher altitudes due to the angle of the Sun above the horizon (so-called solar elevation angle). This also yields the differences in the duration of the surface to the solar energy exposure. In conclusion, these phenomena are linked to the seasonality and diurnally. Therefore, atmospheric lifetime of the species, for example in the case of NO_x, shows differences according to the seasons and the time of the day. NO_x has a short lifetime but studies have been shown a longer NO_x lifetime in winter than in summer (e.g., 21h vs 5.9 h in winter and summer, respectively (Shah et al. 2020) and even a longer daytime lifetime (29h) than night-time lifetime (6.3h) for a winter (Kenagy et al. 2018).

1.3 NO_x chemistry (mainly tropospheric)

Nitrogen dioxide (NO_2) and nitric oxide (NO) are together designated the NO_x family. Most of the NO_x is emitted in the form of NO through high-temperature oxidation of the nitrogen present in combustion air. At high temperatures, molecular oxygen (O_2) dissociates into its atomic states (R1) and contributes to a series of reactions with nitrogen (N_2) in order to generate NO (R2-3). This process is shown in the following chemical equilibriums:

$$\mathrm{O_2} \leftrightarrow \mathrm{O} + \mathrm{O} \tag{R1}$$

$$\mathrm{O} + \mathrm{N_2} \leftrightarrow \mathrm{NO} + \mathrm{N} \tag{R2}$$

$$\mathrm{N} + \mathrm{O_2} \leftrightarrow \mathrm{NO} + \mathrm{O} \tag{R3}$$

After the production of NO, the rapid cycling between NO and NO_2, also known as the Leighton cycle occurs by following the reactions:

$$\mathrm{NO} + \mathrm{O_3} \rightarrow \mathrm{NO_2} + \mathrm{O_2} \tag{R4}$$

$$\mathrm{NO_2} + h\nu \rightarrow \mathrm{NO} + \mathrm{O}(^3\mathrm{P}) \tag{R5}$$

$$\mathrm{O}(^3\mathrm{P}) + \mathrm{O_2} + \mathrm{M} \rightarrow \mathrm{O_3} + \mathrm{M} \tag{R6}$$

Noting that M is a generic term stands for a non-reacting molecule but absorbing the energy of the reaction and stabilizing the production molecule(s) of the reaction. During the daytime, NO is oxidized by O_3 (by the titration of ozone), forming NO_2 (R4). Later, NO_2 is photo-dissociated back by radiation with a wavelength between 290 and 420 nanometers (visible and near ultra-violet) to NO, leading to the production of O_3 (R5-6). It is worth noting that this NO-NO_2

diurnal cycle is in a steady photochemical state and does not contribute to the net O_3 production. However, with the existence of different radicals (e.g., hydroperoxyl radical (HO_2) and organic peroxy radical (RO_2)), the NO-NO_2 balance in the Leighton cycle can be broken down as the radicals contribute to the transformation of NO into NO_2 by the sequence of reactions. The sequence is usually triggered by the reaction of VOC and carbon monoxide (CO) with the OH radical, resulting in the production of HO_2 and RO_2 (R7-8).

$$\mathrm{VOC + OH \overset{O_2}{\rightarrow} RO_2 + H_2O} \tag{R7}$$

$$\mathrm{CO + OH \overset{O_2}{\rightarrow} HO_2 + CO_2} \tag{R8}$$

Noting that R in the reaction is the generic chemical formula of any alkyl organic group members e.g., organic peroxy radicals for RO_2 and alkoxy radical for RO. They are closely linked to the formation of secondary organic aerosols (SOAs) (Peng et al. 2022; Molteni et al. 2018; Orlando and Tyndall 2012).

In the presence of NO, this is followed by the conversion of NO to NO_2 through the reactions with RO_2 and HO_2 radicals, which also generates OH (R9-10). Thus, in particular, R10 also increases the oxidizing capacity of the atmosphere.

$$\mathrm{RO_2 + NO \rightarrow NO_2 + RO} \tag{R9}$$

$$\mathrm{NO + HO_2 \rightarrow NO_2 + OH} \tag{R10}$$

Again, the photolysis of NO_2 (from R9 and R10) leads to O_3 production (R5-6). O_3 production, therefore, depends on the amount of NO available and the presence of radicals (RO_2 and HO_2).

NO affects the production of RO (R9) as it reacts with the RO_2 radical formed through VOC oxidation (R7). In other words, the interactions between VOC and NO_x contribute to SOA production. Therefore, NO_x has a role in dispersion and the formation of SOA (Lane et al.,2008). During the daytime, the main sink of NO_x from the atmosphere is via the oxidation of NO_2 with OH to nitric acid (HNO_3) (R11). NO_2 also reacts with O_3 and forms the nitrate radical (NO_3) predominantly through the reaction of (R12).

$$\mathrm{NO_2 + OH + M \rightarrow HNO_3 + M} \tag{R11}$$

$$\mathrm{NO_2 + O_3 \rightarrow NO_3 + O_2} \tag{R12}$$

NO_3 produced during the daytime is rapidly converted back to NO_2 or NO by photolysis or by the reaction with NO.

During the nighttime, there is no photolysis of NO_2 (as in R5) so all NO_x converted to NO_2 due to R4, reacts with O_3 to produce NO_3 which is also considered to be an important night-time oxidant (R12). The NO_3 radical reacts with NO_2 and produces gaseous dinitrogen pentoxide, N_2O_5 (R13, also an equilibrium) which is a nocturnal reservoir of NO_x. It can photo chemically (only during the daytime) or thermally decompose back to NO_2 and NO_3. However, during the nighttime and in the relatively low temperature (enough to prevent thermal decomposition) it can form HNO_3 by hydrolysis (R14). R12-14 are also considered one of the major paths of the removal of NO_x from the atmosphere.

$$\mathrm{NO_2 + NO_3 + M \leftrightarrow N_2O_5 + M} \qquad \text{(R13)}$$

$$\mathrm{N_2O_5 + H_2O \xrightarrow{aerosol} 2HNO_3} \qquad \text{(R14)}$$

HNO_3 is extremely water-soluble and removed by precipitation (wet deposition) causing acid rain or taken up by plants or soil (dry deposition) (R11 and R14). NO_x is then a precursor of atmospheric NO_3 and HNO_3, which contributes to soil acidification, eutrophication and aerosol radiative forcing (Galloway et al. 2004; Liao 2005).

NO_x has a short lifetime (Section 1.2) which varies from a few hours to about several days. However, peroxyacetyl radical ($CH_3C(O)OO$) also reacts with NO_2 to form more stable peroxyacetyl nitrates (PANs) which act as a carrier and a reservoir for NO_x by the following reactions (R15-17).

$$\mathrm{CH_3CHO + OH \rightarrow\ CH_3CO + H_2O} \qquad \text{(R15)}$$

$$\mathrm{CH_3CO + O_2 + M\ \rightarrow CH_3C(O)OO + M} \qquad \text{(R16)}$$

$$\mathrm{CH_3C(O)OO + NO_2 + M\ \rightarrow PAN + M} \qquad \text{(R17)}$$

where CH_3CHO (also called acetaldehyde) is a widespread gas naturally produced by plants and widely used in industries. PAN is not soluble in water and is not removed by deposition. It allows long-range transport of NO_x at low temperatures in the higher altitudes of the troposphere for weeks. Once PAN begins to descent into the warmer lower troposphere, it releases the NO_x due to the thermal decomposition (R18).

$$\mathrm{PAN\ (+heat)\ \rightarrow CH_3C(O)OO\ + NO_2} \qquad \text{(R18)}$$

This means that, even though NO_x has a short lifetime it can be transported far from its sources or from polluted regions to remote regions (Figure 4).

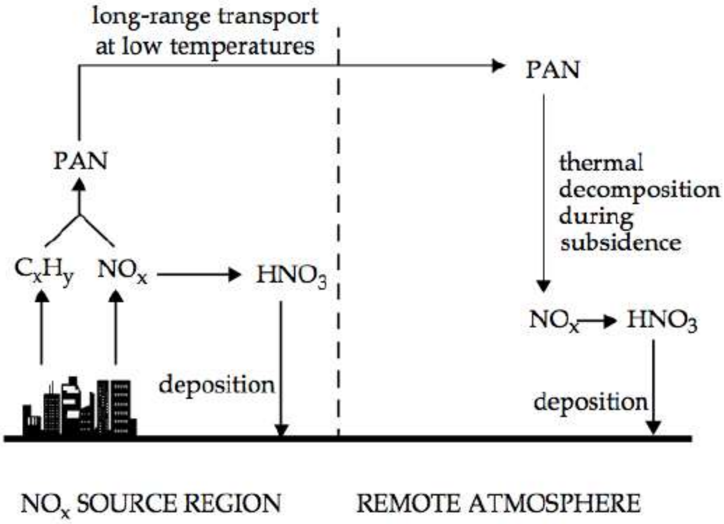


Figure 4 Long-range transportation of NO_x *from PAN. (Taken from Solène Turquety's "Atmospheric chemistry and air quality" course slides, 05/10/2020)*

1.4 NO_x as an air pollutant : discovery and impacts

Air pollution is defined as an atmospheric condition in which the presence of certain substances or air pollutants is higher than normal ambient levels because of anthropogenic or natural activities and has direct or indirect harmful impacts on habitats.

The family of NO_x first introduced as an air pollutant in the early 1950s with the discovery of the chemist Dr.Arie Haagen-Smit in California. During this time, a smog event occurred in the city of Los Angeles, causing reduced visibility, eye irritation, heavy smell, and damage to vegetation. He discovered the smog event, which is caused by the increased number of vehicles in the traffic and the emissions of NO_x species from these vehicles. In fact, the emissions of NO_x and VOCs (e.g., isoprene, ethane) from vehicle exhausts reacted with the incoming solar radiation (or UV) to produce O_3 (which is the key component of the smog) and atmospheric aerosols, and caused so-called brownish-grey photochemical smog. This finding brought attention to the role of NO_x in the photochemical mechanisms related to air pollution in the troposphere (Section 1.3).

Therefore, not only NO_x as a primary pollutant harm the ecosystems and human health, but also secondary pollutants, such as O_3 and HNO_3, and SOA, which are a product of various chemical reactions through NO_x, cause serious environmental and health problems.

In terrestrial ecosystems, exposure to high concentrations of NO_x and O_3 has direct negative impacts on plant growth and the productivity of agricultural crops. Leaves, mainly through their stomata, absorb these pollutants directly, resulting in reduced seed production and crop loss (de Vries 2021). Moreover, when NO_x reservoir gas HNO_3 dissolves in water, acid rain damages plants, aquatic ecosystems, and cultural heritage. In aquatic ecosystems, elevated NO_x and NO_3 concentrations can cause eutrophication and acidification in freshwater, affecting nutrient balances and oxygen levels in the water.

Both short and long-term exposures to high NO_x and O_3 concentration levels affect various human body systems and organs, such as in the respiratory system from respiratory infections, asthma to lung cancer and in the cardiovascular system, from angina to heart diseases (Kampa and Castanas 2008).

NO_x emissions have sophisticated impacts on climate change, both speeding up and slowing down the effects (Lasek and Lajnert 2022). Although NO_x is a reactive trace gas, it is responsible for the formation of tropospheric O_3, which in addition to be a pollutant is also a greenhouse gas (GHG) (Mitchell 1989; Szopa et al. 2021). Ground level O_3 absorbs terrestrial and solar radiation, causing positive radiative forcing which results trapping of the radiation in the troposphere (Stordal et al. 2003). By this way, NO_x indirectly affects climate warming through tropospheric O_3 production. On the other hand, NO_x is also responsible for the formation of OH radicals or cleaning agents. On a global scale, OH reacts with CO to form carbon dioxide (CO_2). Nevertheless, it also reacts with CH_4, which is also one of the most abundant anthropogenic GHG after CO_2, and reduces the CH_4's lifetime.

1.5 NO_2 monitoring tools

It was only in the late 19th century the first monitoring of air pollutants began with O_3 at the Pic du Midi Observatory in southwestern France with the intention of using O_3 as a medicine (Marenco et al. 1994; Palmer 2017).

After the discovery of the NO_x family of being the most significant precursor of photochemical O_3 production and HNO_3, and their possible impacts, monitoring of NO_2, its vertical profiles, and the global trend in NO_2 concentrations aroused interest for use in studies. Vertical profiles can be derived using balloons (Berthet et al. 2017), airborne (Merlaud et al. 2011), and ground-based instruments (e.g., Système d'Analyse par Observation Zénitale (SAOZ), (Kreher et al. 2020)) but their deployment is limited spatially and temporally. On the other hand, in situ surface measurements are usually more accurate and provide measurements continuously over longer period. However, an important concern of the surface in situ instruments is the representativeness of the measurement as they can be influenced easily with the surrounding environment (e.g., instruments situated at the stations nearby the traffic). Even if there are some techniques to evaluate the measurement type (Flemming, Stern, and Yamartino 2005; Lachatre et al. 2020), quantifying the representativeness remains difficult. Moreover, it can be costly to set up a network of ground-based instruments to cover the area of interest and operate them continuously. Therefore, observations from space are complementary to the other measurement techniques for NO_2 monitoring especially thanks to their increased spatial and temporal coverage. They are able to provide global view of NO_2. The analysis of satellite measurements requires precise comprehension of the radiative properties of the atmosphere and sophisticated methods to retrieve information on NO_2 amounts. Satellites do not measure surface concentrations; they measure vertically integrated concentrations as columns. One of the limitations of these observations are the clouds, which prevent observations. Long-term continuous monitoring of NO_2 concentrations using satellites have been possible since the 1990s (Anenberg et al. 2022).

1.6 NO_2 concentration levels in the atmosphere

Official authorities (e.g., governments) or non-profit organizations (e.g., World Health Organization (WHO)) determine thresholds and establish guidelines to keep ambient air pollutant levels below hazardous levels. As NO_2 is an air pollutant, these regulations usually include limits on the amount of NO_2 that can be emitted to keep ambient NO_2 concentrations below certain thresholds. According to the WHO's latest report in 2021, air quality standards for the protection of public health are defined as follows : ambient NO_2 concentrations should not exceed an annual average of 10 $\mu g/m^3$ and a 24-hour average of 25 $\mu g/m^3$ (WHO Global Air Quality Guidelines 2021). However, despite the WHO's relatively restrictive limits, the European Environment Agency's (EEA) “Air quality in Europe 2022” report shows that, NO_2 concentrations are recorded above WHO levels in all Organization for Economic Co-operation and Development (OECD) member countries (at total 38 countries) (European Environment

Agency 2022). The EEA's less restrictive threshold is set at 40 µg/m^3 on an annual average. Similar to the EEA, China's ambient air quality standard limits (CAAQS) are 40 µg/m^3 annual average and 80 µg/m^3 24-hour average for all areas (including urban and industrial areas) nationwide. And according to the National Ambient Air Quality Standards (NAAQS) of the United States Environmental Protection Agency (EPA), ambient NO_2 concentrations should not exceed 53 parts per billion (ppb) (or equivalent to 99.64 µg/m^3 using the WHO's conversion factor) (EPA).

Figure 5 shows the global spatial distribution of tropospheric NO_2 columns from the satellite instrument, Ozone Monitoring Instrument (OMI) averaged over the years from 2005 to 2018 with a spatial resolution of 0.25° × 0.25°. It is quite apparent that the red spots with higher NO_2 concentrations are mostly located in the northern hemisphere, such as Eastern Asia (Eastern China, Japan), Western Europe (e.g., The Netherlands, Belgium, France), and the United States of America (USA) (e.g., New York City, Philadelphia). These places indeed have been observed the highest levels of NO_2 concentrations over the last two decades (Richter et al. 2005; R. J. van der A et al. 2006; Paraschiv et al. 2017; Jamali, Klingmyr, and Tagesson 2020).

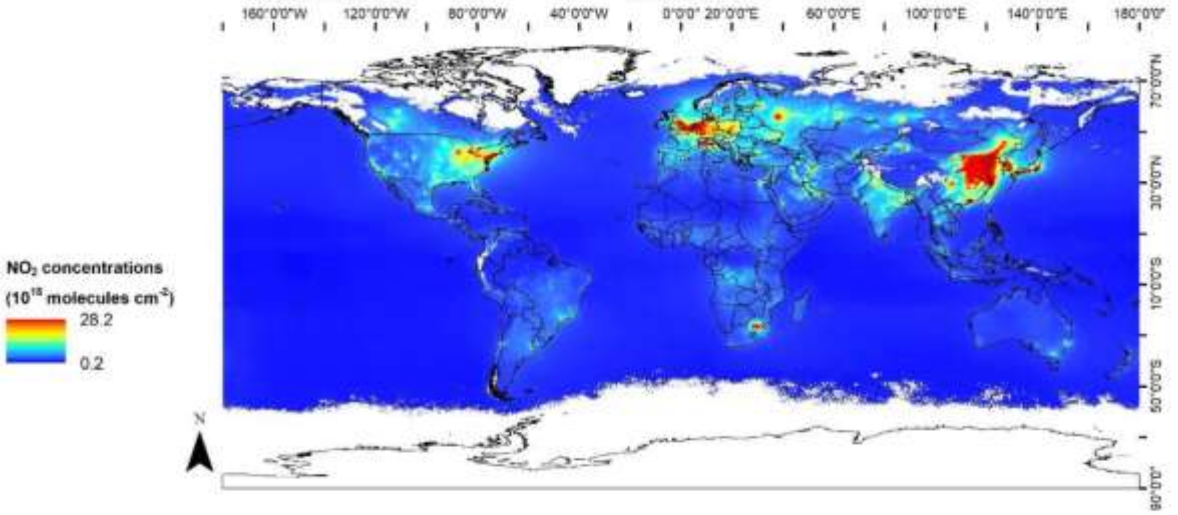


Figure 5 Global map of mean tropospheric NO_2 concentrations from 2005 to 2018 as observed by the OMI satellite instrument. Figure is taken from Jamali, Klingmyr, and Tagesson 2020.

Figure 6 represents the spatial distribution of the surface annual means of NO_2 concentrations (in µg/m^3) in 2019 for Europe taken from the surface stations. The measurement data is based on the European air quality database. There are at total 3,480 stations available for this year. Of these, 230 exceed the EEA's limit standards (corresponds to 7 %), while 2820 (81 %) exceed the WHO's limits on an annual base.

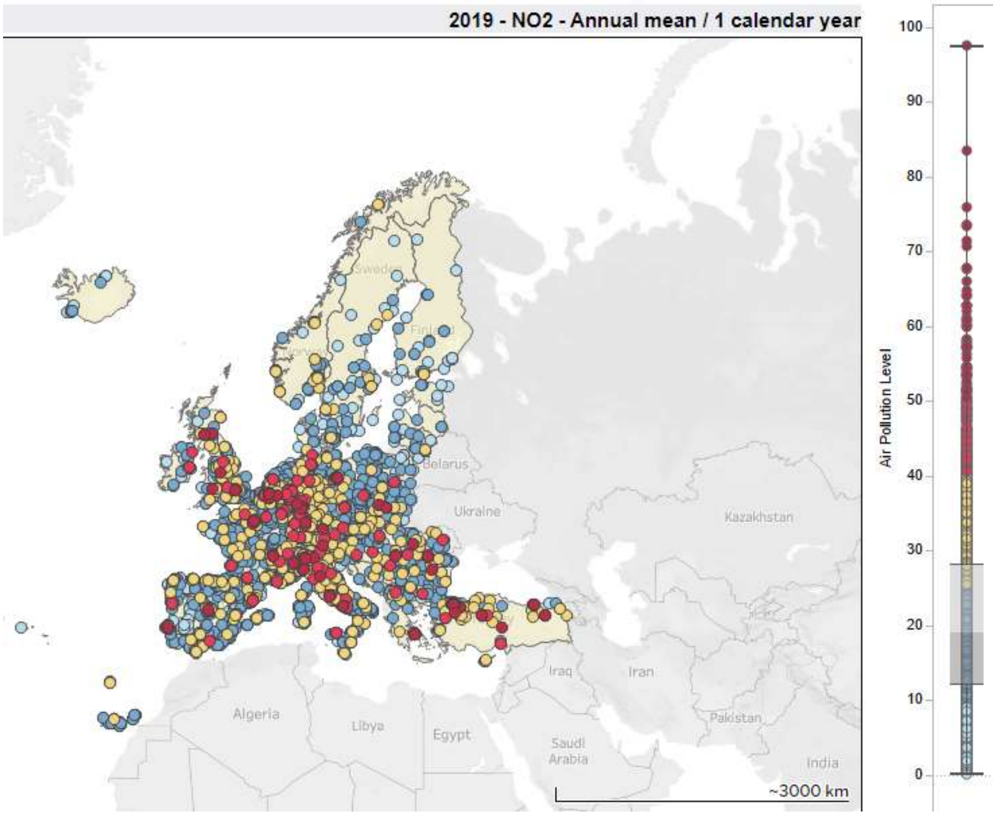


Figure 6 Ambient annual mean of surface NO_2 Concentrations for 2019 (in µg/m³), extracted from https://www.eea.europa.eu/data-and-maps/dashboards/air-quality-statistics

Similar to Figure 6, Figure 7 shows the spatial distribution of surface annual means of NO_2 concentrations (in µg/m³) in 2019 for China. There are 626 stations, of which 43 exceed the CAAQS recommended level (corresponds to 7 %) and 597 exceed the WHO recommended level (corresponds to 95 %).

Accordingly, a significant number of stations do not meet the requirements of WHO levels, while a significant number of stations fall below the CAAQS levels. Ultimately, the aim of policy makers or authorities is to implement mitigation policies to keep pollutant concentrations below these harmful levels. Where measurements cannot be made, for example for forecasting, or to evaluate the best mitigation policies, air quality models are used to simulate NO_2 concentrations in order to predict future ambient levels and impacts of the chosen policies.

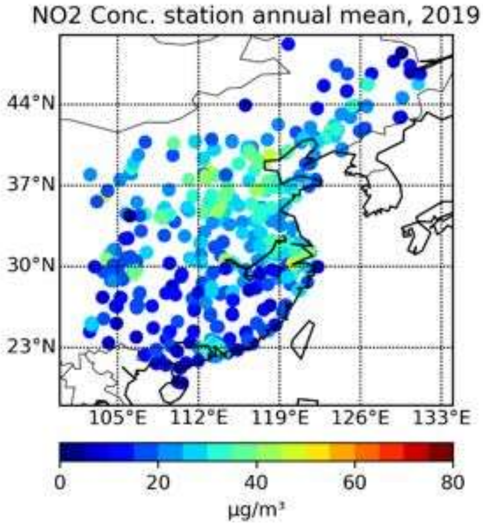


Figure 7 Spatial distribution of annual mean surface NO_2 concentration in 2019 for China (Visualized by using data of the CNEMC network).

It is also important to note that NO_2 concentrations show daily, seasonal and inter-annual trends. For instance, Figure 8 represents the trends of the monthly mean of the tropospheric NO_2 vertical columns over East Central China from 1996 to 2004 (Richter et al. 2005). NO_2 measurements were taken from the early phase of the satellite era, from the Global Ozone Monitoring Experiment (GOME) (Burrows et al. 1999; European Space Agency 1995) satellite instrument between 1996 and 2002, and then from the SCIAMACHY (Bovensmann et al. 1999) instrument for the years 2003 and 2004. The time series shows a slight increase in the NO_2 vertical columns from 1996 to 2004 for East Central China. Especially due to seasonality, higher values are observed in NO_2 measurements in winter months and these values have been increasing year by year (in particular after the 2000s).

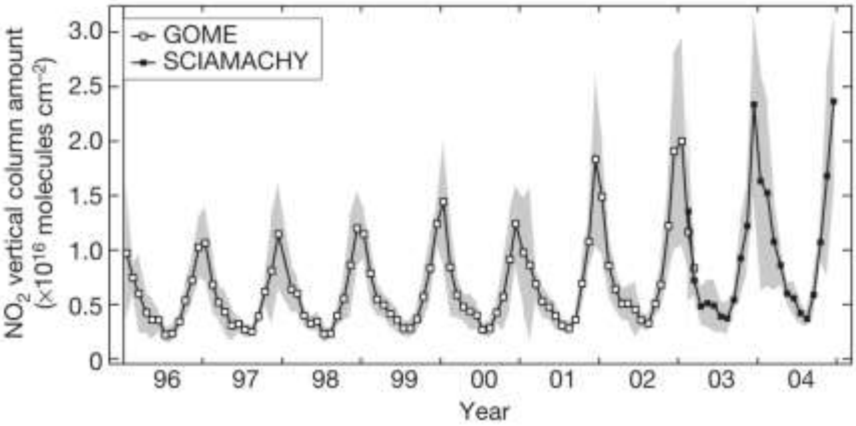


Figure 8 Inter-annual monthly averages of NO_2 tropospheric vertical columns over East Central China measured by GOME and SCIAMACHY satellite instruments. Figure taken from Richter et al., 2005 (Shaded areas show the standard deviation)

Looking at the global trends, NO_2 concentrations show a slight decreasing trend between 2004 and 2008, followed by an increase from 2008 to 2018, with a significant increase in the first 3 years (Figure 9, from Jamali, Klingmyr, and Tagesson 2020). However on a continental scale, this situation differs with a sharp decreasing trend in Europe from 2006 to 2008 and from 2016 to 2018. Still between 2008 and 2016 a slight increase is observed. China has increasing trends of NO_2 concentrations until 2011, followed by a sharp descrease until 2016 and a slight increase (or insignificant change) between 2016 and 2018. These figures show that despite the increasing global trends in the NO_2 concentrations the contribution of different regions to this global trend shows substantial regional differences. This may be a result of the implementation of different mitigation policies in different time periods.

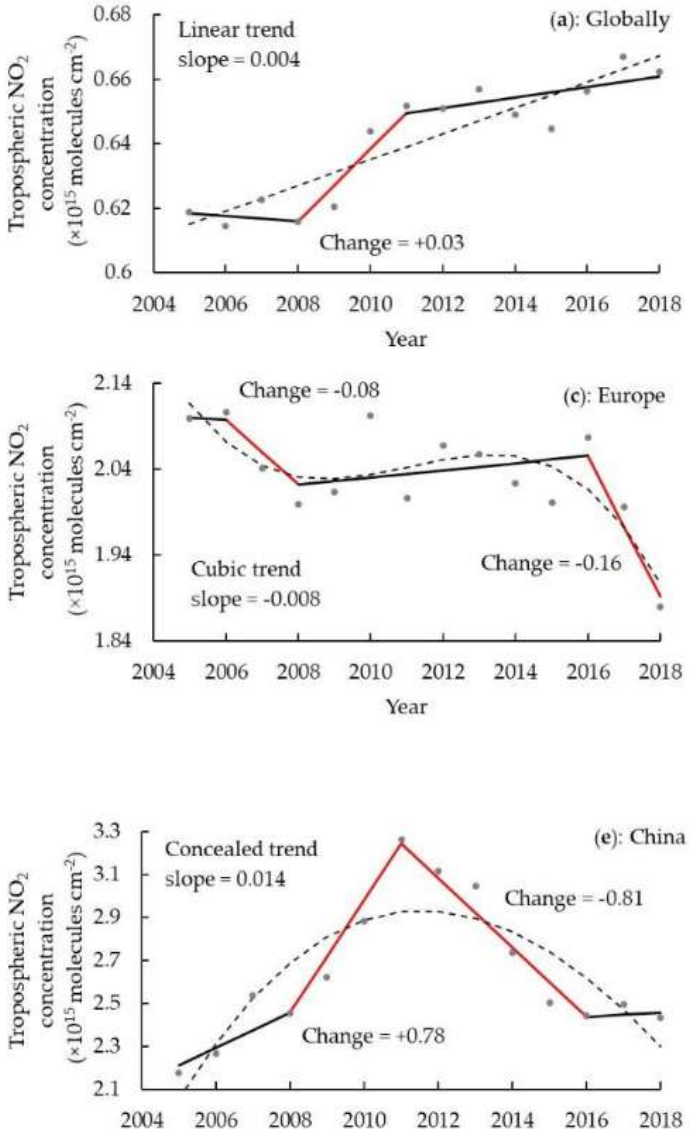


Figure 9 Time series of annual mean of tropospheric NO_2 concentration using DBEST (Jamali et al.,2015) trend analysis method and OMI satellite measurements for the Globe, Europe and China. Figures are taken from Jamali, Klingmyr, and Tagesson 2020

For example, in China, the 12th Five-Year Plan, implemented by the Chinese government in 2011, aims to reduce annual NO_x emissions by 10% within five years (de Foy, Lu, and Streets 2016). This is also consistent with the decrease in NO_2 satellite concentrations starting in 2011 as seen in Figure 9. In European Union (EU), the Gothenburg protocol has been conducted to improve combustion processes, thus reducing NO_x emissions since 2005 (United Nations Economic and Social Council 2015). The task group of the Gothenburg protocol have developed policy scenarios in the framework of Long-range Transboundary Air Pollution, and set long-term goals to reduce NO_x emissions by about 50 to 80 % in the EU, North America, and the Western Balkans, and about 20 % in the Eastern Europe, the Caucasus and Central Asia between 2005 and 2030. In addition, for instance, there are also global conventions such as the Marpol Convention, which aims to reduce global NO_x ship emissions since 2005 (www.imo.org, Last Access: 10/09/2022).

These regulations commonly include limits on the amount of NO_x that can be emitted by vehicles and industrial facilities, as well as requirements for the use of emissions control technologies.

1.7 NO_x emissions

1.7.1 NO_x emission sources

Humankind first began to influence atmospheric composition by emitting pollutants through agricultural activities (e.g., developing land for crops and transportation near water sources or along trade routes, and raising livestock for meat and dairy products). However, with the Industrial Revolution that began in the late 18th century, increasing anthropogenic activities began to increase air pollutant emissions significantly in industrialized regions. These activities became a dominant source of changes in the atmospheric compositions of air pollutants (Abram et al. 2016).

Increase in NO_x emissions due to urbanization, industrial and economic development and consequently augmented energy consumption and transportation has a direct effect on the increase of NO_2 concentrations, although there is a non-linear relationship between NO_2 observations and NO_x emissions (Vestreng et al. 2009; Konovalov et al. 2010; Huszar et al. 2021).

NO_x emissions originate from both biogenic (or natural) and anthropogenic sources. In the nature, NO_x can be produced by lightning with a considerable amount in the upper troposphere (more than 1.2 billion flashes occur annually (NASA/Goddard Space Flight Center 2009), globally at total 5 Tg N yr^{-1} as in Table 1 or estimated between 2 and 8 Tg N yr^{-1} by Schumann and Huntrieser 2007 (not mentioned in the table). Because during lightning, release of very high temperatures can split the molecular oxygen, which can latter react with abundant nitrogen (R1-3). Soil is also another natural source of NO_x (Almaraz et al. 2018; Xiao Lu et al. 2021). NO is the predominant forms of soil NO_x emissions produced by biogenic sources, such as microorganism activities in the soil via nitrification and denitrification processes. These processes can be affected by nitrogen and ammonia availability in the soil, and moisture, temperature, structure, type and pH of the soil. Therefore, human also influence soil NO_x emissions through different agricultural activities such as using nitrogen fertilizer, irrigation, crop harvesting, and land clearance. Biomass burning (e.g., deforestation and agriculture) can also be both biogenic and anthropogenic sources of NO_x emissions by burning of living or dead vegetation. Nevertheless, the significant contribution of NO_x emissions comes from anthropogenic sources, particularly from combustion processes (Table 1). The combustion engines are included in the sectors like; energy sector (e.g., thermal power plants especially using fossil fuels such as coal as a raw material), industry, transportation (e.g., traffic, shipping, aircraft), residential (e.g., home heaters, gas stoves). Figure 10 visualizes the main NO_x emission sectors of road transport, electricity and heat generation, and manufacturing and construction for EU countries in 2019 (European Union Emission Inventory Report 2021).

Table 1 Overview of global NO_x emission budgets and their uncertainties in parenthesis for selected sources and sectors

Source	**Estimated NO_x emissions for different sources and sectors (Tg N yr^{-1})**		
Anthropogenic Sources	**Lee et al. 1997**[a]	**TAR**[b]	**AR4**[c]
Fossil fuel combustion	22 (13-31)	33 (20-24)	25.6 (21-28)
Biomass Burning	7.9 (3-15)	7.1 (2-12)	5.9 (6-12)
Aircraft	0.9	0.7 (0.2-0.9)	
Agriculture		2.3 (0-4)	1.6
Atmospheric Deposition			0.3
Anthropogenic Total	**30.8**	**43.1**	**33.4**
Biogenic Sources			
Soils	7 (4-12)	3.3 (3-8)	7.3 (5-8)
Lightning	5 (2-20)	5 (2-12)	1.1-6.4 (3-7)
Biogenic Total	**12**	**8.3**	**8.4-13.7**
Total	**42.8**	**51.4**	**41.8-47.1**

[a] Best estimate source and uncertainties adapted from Table 3 in the paper.
[b] The IPCC Third Assessment Report (TAR) (Houghton et al. 2001), Table 4.8.
[c] The IPCC Forth Assessment Report (AR4) (Core Writing Team, Pashauri, and Reisinger 2008), Table 7.7.

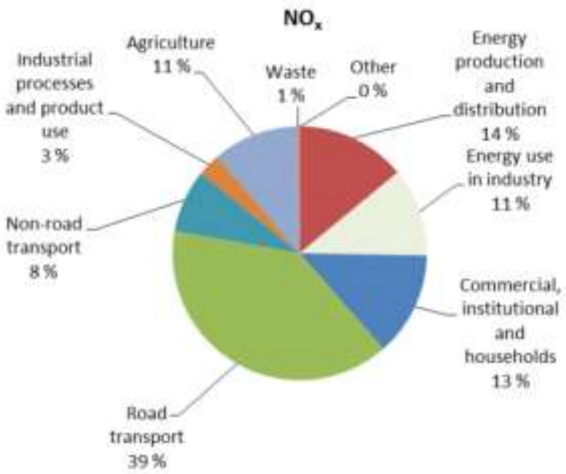


Figure 10 EU Anthropogenic NO$_x$ emission sources by sectors for 2019 (Taken from EU emission inventory report, 2021)

1.7.1 NO$_x$ emission inventories

A series of emission sources or inventories are compiled into a comprehensive report accounting the amount of various pollutants released into the atmosphere from different sources (or emitters) over a period, usually annually. This so-called emission inventories are then used to track changes in pollutant levels over time, assess the effectiveness of pollution reduction policies and regulations, and provide data for scientific research on the impacts of pollution on the environment and human health. They are generally compiled by government agencies, but can also be compiled by private companies, non-governmental organizations, and research institutions.

1.7.1.1 Bottom-up approach

There are different ways to build emission inventories. The so-called bottom-up approach is one of the main and widely used approaches. Inventories using bottom-up method is based on the individual emitters. Using this approach, emissions are calculated, sector by sector, by multiplying the amount of pollutants emitted per unit of activity (such as fuel consumption or electricity generation) by the number of activities that occurred during a given period. This requires detailed information on the characteristics of each source, such as the type of fuel used, emission activity rate, emission factor, the efficiency of the technology, and the operating conditions. Due to the high number of emission calculations, and input data collected from one-to-one sources, it takes time, effort and becomes complicated to build a complete bottom-up inventory of emissions. And even more complicated for a larger region. For instance, estimating the amount of NO$_x$ emissions from an individual car is a challenging task, since it depends on various factors such as, the car's model, the speed during the trip, the length of the route, and the maintenance of the vehicle. This becomes even more complicated when one tries to calculate the total emissions from every car in a country. Then combining this information from each activity and accurately reporting the true total amount of anthropogenic emissions involves high uncertainties. These uncertainties comes from the activity data, the emission factor, spatial allocation and timing of emissions. For example, Zhao et al. 2011 estimated the uncertainty of national, bottom-up inventory for NO_x emissions using the Monte Carlo approach to 37 % for China mainly due to the oversimplified source classifications and roughly estimated emission factors. However, despite the uncertainties, bottom-up inventories provide highly detailed information. Therefore, this bottom-up approach is often used in national and regional emissions inventories.

1.7.1.2 Top-down approach

In contrast, the top-down approach estimates emissions by measuring pollutant concentrations in the atmosphere and using models to inverse the emissions that would be necessary to produce those concentrations. In particular, recent advances in satellite technology have enabled the use of observations from these sources, and the increase in computational power has made it possible to solve the highly complex problems. As a result, top-down approach is conducted using mathematical atmospheric inversion techniques with modeling and observations (widely used satellite observations) to constrain emissions on high performance computer clusters. Despite the fact that, it relies on assumptions in the models and uncertainties come from the model and instruments, the use of satellite observations brings some advantages. For instance, the observations cover a larger area at the same time (spatial) and provide continuous measurements over the area of interest on a daily base (temporal). Also, the increase in the resolution of satellite observations helps to better capture emission sources (higher resolution). All of these advantages helps to improve uncertainties contained in bottom-up inventories. For this reason, establishing top-down emission inventories by exploiting the potential of satellite observations is promising and still developing research area of interest.

1.7.1.3 Examples and comparison of two approaches

National or regional inventories are built using local pieces of information, while global inventories are created using the mosaic method by combining these inventories from different countries or regions into a uniform emissions inventory over a larger domain (Figures 11 and

12). For example, the Multiresolution Emission Inventory for China (MEIC) (Liu et al. 2017; Zheng et al. 2018) v1.0 bottom-up inventory, with a 0.25° × 0.25° spatial resolution, covers China as in Figure 11 for the period 2007 to 2017, and is used together with other local inventories to compile the regional MIX inventory for the whole of Asia (Li, Zhang, et al. 2017a). Furthermore, used in the mosaic of regional and global HTAP v2.2 inventory as in the Figure 12 (Janssens-Maenhout et al. 2015).

Figure 11 Compilation of regional MIX Asia bottom-up emission inventory with local bottom-up inventories (Figure is taken from Li M. et al. 2017)

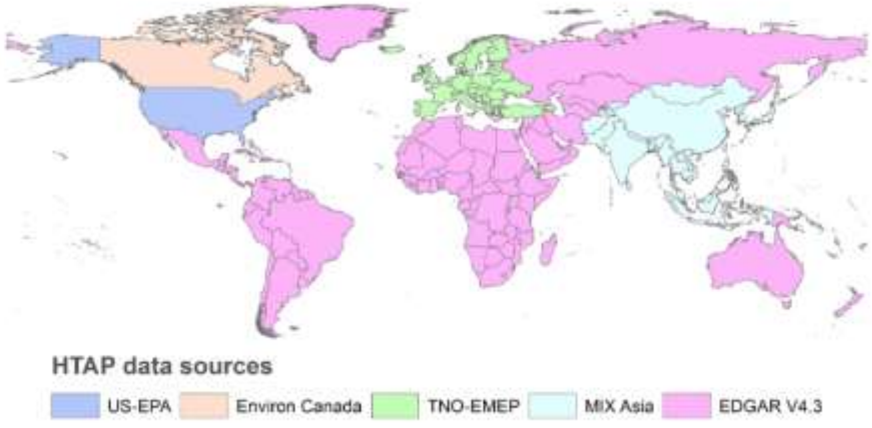


Figure 12 Compilation of several regional bottom-up emission inventories to build a global HTAP v2.2 emission inventory (Figure is taken from Janssens-Maenhout et al., 2015)

The Emissions Database for Global Atmospheric Research (EDGAR) (Janssens-Maenhout et al. 2015) v4.3.1 (Crippa et al. 2016) is another bottom-up inventory with a resolution of 0.1° × 0.1° and covers a relatively large domain for the period from 1970 to 2010 as in Figure 12 compared to other inventories. CAMS (Copernicus Atmosphere Monitoring Service) operates the ECCAD (Emissions of atmospheric Compounds and Compilation of Ancillary Data) database, which contains various emission inventory datasets at the global (1° × 1°) and regional scales (up to 0.05° × 0.1°). It includes national EMEP, TNO bottom-up inventories for Europe as well as, inventories obtained by using top-down approach.

Figure 13 shows an inter-annual comparison between different emission inventories: bottom-up inventories (REAS v1.2, REAS v2.2, MEIC and EDGAR) and top-down inventories (DECSO-OMI, DECSO-GOME2a, EnKF-MIROC and EnKF-CHASER) for the mainland China (taken from Ding et al. 2017).

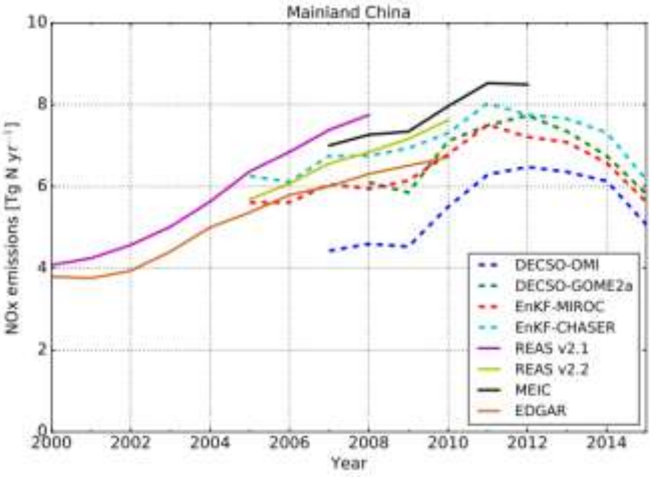


Figure 13 Annual time series of NO_x emissions from different emission inventories over mainland China. (Figure taken from Ding et al., 2017)

According to the figure, top-down inventories (dotted lines) provide recent information on the emissions compared to the bottom-up inventories (plain lines), which is related to the difficulties of the real-time updates of bottom-up inventories. Because time is required for collecting and processing information from each emitter. Moreover, bottom-up inventories such as MEIC and Regional Emission Inventory in Asia (REAS) v2.2 contain relatively short-term information, such as a couple of years. Furthermore, all emission inventories show differences in the amount of annual NO_x emissions compared to each other. The intersection of the values for some years only exists in top-down inventories (e.g., DECSO-GOME2a and EnKF-MIROC in 2011.)

To conclude, bottom-up emission inventories are subject to uncertainties and limitations (e.g., 50.7 % in NO_x emissions for 2012 in EU28 (Crippa et al. 2018). However, accurate and up-to-date emission inventories are crucial for understanding the contribution of various human activities, modeling and projecting the associated changes in the atmospheric composition, and developing the most feasible strategies to mitigate emissions. To this end, benchmarking studies have shown that although top-down inventories also consist of uncertainties, they can provide complementary information to improve bottom-up emission inventories by using their advantages (N. Elguindi et al. 2020).

1.8 Objectives of the thesis

As shown previously, top-down approaches are promising to complement bottom-up inventories providing regional to global information with a fast update capability. These methodologies have reached a good level of maturity in the last decade with the development of several tools (Müller and Stavrakou 2005; Miyazaki et al. 2017; Qu et al. 2017; Fortems-Cheiney et al. 2021) and new satellite observations. At the beginning of this PhD work, the new satellite instrument TROPOMI was starting to provide NO_2 tropospheric columns at an unprecedented spatial resolution of a few kilometers. At the French level, a new inversion system was in development, the Community Inversion Framework (CIF) (Berchet et al. 2021) based on a previous less versatile version of PYVAR (Fortems-Cheiney et al. 2021). This tool was first dedicated to the inversion and monitoring of greenhouse gases (CO_2, CH_4) at the global scale. Capabilities to apply top-down approaches to highly reactive species at the regional scale and the high resolution were limited. In this context, the main objective of the ANR ARGONAUT project was to address the question of the French's capabilities to develop the regional inverse modeling at the high resolution of reactive species, taking advantages of the CIF development, the national regional model CHIMERE, and the newly TROPOMI observations. The PhD work has been developed in this framework and co-funded a grant from CNES. The main objective of my work was to test the capabilities of the CIF tool coupled with the CHIMERE model. To do so, I focused my work mainly on NO_x emissions because NO_2 satellite products are the most mature in terms of product for primary pollutants with reasonable uncertainties and good temporal resolution (daily). Considering atmospheric chemistry, NO_x chemistry is also better known and represented in the model than other species, such as VOCs. NO_x inversion was then a good candidate to start with the evaluation of a new inversion system. The objective was to test the CIF-CHIMERE in various configurations (domains, resolutions, and observations) to evaluate its performances and conditions of work. Three domains have been considered Europe, France, and China, two at medium resolution (50 km for Europe and China) and one at high resolution (10 km for France). The objective was also to evaluate, somehow, the gain of using observations at different resolutions, comparing inversion with OMI and TROPOMI.

Scientifically, I focus my work first on Chinese NO_x emissions. Indeed, the Chinese government developed new mitigation strategies for these emissions during the last decades. I evaluated the performances of the new CIF-CHIMERE system and the NO_2 observations from OMI to follow and quantify the emission reductions for two years (2015 and 2019) at the end of each 5 Year-Plan of the Chinese government. It was not possible to use the new observations from TROPOMI for this study as they started only in 2018. Secondly, I made preliminary tests to apply the CIF-CHIMERE in Europe and France using TROPOMI and OMI to evaluate improvements given by the high resolution of TROPOMI and also to analyze the potential limitations. I focus mainly on the first lockdown period in April 2020 due to the COVID-19 pandemic.

In the second part of the thesis (Chapter 2), the "Theoretical" Tools used in this PhD work are described: the historical development and theoretical bases of inverse modeling and the state-of-the-art CIF inverse modeling system are discussed, as well as the CHIMERE CTM, including the model inputs, and the different satellite and surface observations used. The various tests done to characterize the system and set up the most appropriate configuration are also presented.

Chapter 3 is devoted to the first case study, the inverse modeling of NO_x emissions over China using OMI satellite observations, followed by Chapter 4, which presents the preliminary tests in Europe and France applied to the COVID-19 shutdown period, and discusses expected improvements from TROPOMI compared to OMI, difficulties and possible limitations. A general discussion of the results and findings and suggestions for the extension of this work concludes the manuscript.

Second part : Methodology and Instruments

Chapter 2 : Inverse modeling and its tools

In this thesis, the variational data assimilation technique is applied to inverse NO_x emissions using NO_2 satellite observations. Therefore, this chapter describes the development of the data assimilation theory under the framework of inverse modeling with more emphasis on the variational approach. Moreover, it also includes descriptions of the models and tools used during the inversion and the data sources used to assess the inversion results.

2.1 The concept and theory

2.1.1 What is an inverse problem?

To predict the state or future of an atmospheric system, it is necessary to build physical or mathematical models. A direct or forward problem finds the effects of a given cause (e.g., used as inputs in a model) using these models, while an inverse problem does the reverse and aims to find the cause of a given effect. In 1976, the American mathematician J.B.Keller called two problems inverse if “the formulation of each problem contains all or part of the solution of the other” (J. B. Keller 1976). However, the main difference between these problems is that direct problems are well-posed, while the inverse problems are ill-posed (e.g,, there is no solution of the problem, or no unique solution, or the solution is unstable). This concept was introduced by the French mathematician Hadamard (Hadamard 1923). On the other hand, regularization or penalization methods (e.g., Tikhonov regularization (Tikhonov 1963)) are designed to solve these ill-posed inverse problems and aim to find an existing, unique, and stable (continuous dependence on the parameters) solution of the problem (Hansen 2010). General information about inverse problems and their uses, including other fields of science, can be found briefly in Kabanikhin’s survey paper (Kabanikhin 2008).

2.1.2 Overview of data assimilation: past to present

2.1.2.1 Assimilation as an inverse problem

Inverse modeling or state-estimation theory is often referred to as data assimilation in geosciences (Carrassi et al., 2017), and data assimilation methods are used to solve inverse problems. For example, an inverse modeling that aims to find a set of model parameters (e.g., emissions) that best fits the observed data (e.g., concentrations) requires the assimilation of observations to estimate the system's state. Asch, Bocquet, and Nodet 2016 describes in detail the strong mathematical connections between techniques in inverse problems and data assimilation methods. The mathematical theory behind data assimilation comes from the 18th-century English statistician Thomas Bayes. His idea simply consists of modifying our prior belief with objective information; therefore, the posterior updated belief combines both prior and that new information (Bayes 1763). He described a theorem (after named the Bayes' theorem), which deals with the conditional probability of an event (y) based on the values of known probabilities of prior event (x) or in other words, to deal with the probabilities of the combination of states of information following the *Bayes' rule* (Equation 1):

$$p(x|y) = \frac{p(y|x)\,p(x)}{p(y)} \qquad \text{(Eq.1)}$$

Tarantola has extensively elaborated this probabilistic approach to inverse modeling problems, from Bayes up to the day of his book's publication in 2005 (a rewritten and the latest version of the book from 1987 by the same author) (Tarantola 2005).

Data assimilation aims to combine optimally all the information from model simulations and observations, taking into account their uncertainties, to produce a better estimate of systems representing reality. This estimation process can aim to find the best estimate of the parameters of a model or the state of the system based on available observations. However, during the process, there are issues that make data assimilation challenging and add some difficulties in the calculations of the estimations: (i) systems are complex (e.g., the non-linearity of atmospheric systems or poorly known interactions). (ii) observations are noisy, sparse, and irregularly distributed in space and time (heterogeneous observations). (iii) the problem is ill-posed (regularizations modifies the main problem; thus the result would be the approximation of the true solution).

The invention and development of the least squares method for studying the position of stars by Gauss and Legendre in the early 1800s happened to be a pivotal event for today's data assimilation techniques (Stigler 1981) and also for the probability theory. The idea of the method was based on the minimization of the sum of the squares of the residuals (or errors) that is the departure between an observed value and the fitted value provided by a simulation model. In addition to that, the Gauss-Markov theorem guarantees that this minimization is optimal under some assumptions (e.g., the errors are uncorrelated, have zero mean and have equal variances) in a linear model and the best linear unbiased estimator (BLUE, described in detail later) of the coefficients is the least-squares estimator. While these advances in data assimilation theory were groundbreaking, atmospheric models and applications of atmospheric data assimilation using these models were still in their infancy.

2.1.2.2 First usages of data assimilation in atmospheric science

Data assimilation was first used in numerical weather prediction (NWP) models with the aim of short-term forecasting of meteorological conditions in 1960s, while the history of NWP models goes back much further. At the beginning of the 20th century, Cleveland Abbe described the physical laws of the atmosphere (e.g., principles of thermodynamics and hydrodynamics) in the need of forecasting the meteorological conditions (Abbe 1901). Vilhelm Bjerknes suggested an estimation or prediction of weather could be as a deterministic initial value problem subject to its initial conditions such as temperature, humidity, pressure, density (Bjerknes 1904). He developed a graphical method for solving a system of non-linear partial differential equations. He also found that solving these equations numerically was challenging, and no analytical solutions were possible (Lynch 2008). Then in 1922, Lewis Fry Richardson, an English meteorologist, influenced by Bjerknes, brought a novelty to the weather prediction by applying numerical methods (NWP models) and solving the equations analytically (Richardson 1922). He proposed the first data assimilation technique by interpolating the observations. However, in his analysis, the observations had not been properly assimilated because he observed significant errors due to the inaccuracies of the fields used as the initial conditions. Nevertheless, in early 1950, under the leadership of an American meteorologist Jule Charney, the development of the first generation electronic computer (called the Electronic Numerical Integrator and Computer, ENIAC), the necessary mathematical analysis and design of a numerical algorithm made it possible to have a first successful 24-h NWP forecasts (Lynch 2008), despite the fact that, in their experiment they used Richardson's simplified equation set.

In the late 1950s, George Cressman's successive correction algorithm based on the method of Bergthörsson and DÖÖ brought the idea of weighting the background by observations, and adding linear combinations of the observational increments (observation minus background) to the backgrounds in three dimensions (3D) (Cressman 1959; Bergthörsson and DÖÖs 1955). These weights were determined statistically. They used previous forecasts for the initialization (first guess) of the model to provide complete information (background) over the domain. The algorithm combined the observations with models in a statistical sense via the least square method and with the intention to find a 'best fit', which is called optimal interpolation (OI).

2.1.2.3 Sequential approach: the Kalman filter and its variants

By the 1960s, OI methods began to be globally operational in many meteorological centers to make weather predictions. The optimal interpolation algorithm is a pioneering and simpler version of the Kalman Filter (KF) method (Kalman 1960). KF and its variants are one of the two main data assimilation approaches called sequential (statistical) data assimilation. KF is an optimal linear estimator of the parameters, updating the previous optimal estimates as new observations (noisy) available (therefore, sequential) and providing the uncertainty of the analysis at each step. The detailed descriptions and the analysis of the method can be found in Maybeck and Siouris 1980. However, KF is only optimal in the case of Gaussian statistics and linear systems, so it could not be directly applied to nonlinear systems. As the real life problems involve non-linear functions, the extension of KF that includes Taylor series expansion, to help to get a linear approximation of the non-linear functions was necessary. This is through an approximation of the non-linear function using first derivative of Taylor series called Jacobian matrix. This version of the filter is called the extended Kalman Filter (EKF) and was developed by Jazwinski 1970. However, the numerical cost for large atmospheric systems is very high and the dimensionality of the computational grid and storage cost of the matrix led to the subsequent development of other KF methods, such as the ensemble Kalman filter (EnKF) developed by Evensen 1994. The method generates an ensemble of initializations to be used to ensemble forecast in order to reduce the dimensionality. In the end, this method is a simplification and close approximation to the KF and can provide similar results if the ensemble size is large enough and if the model errors are well represented by the ensemble. Houtekamer and Mitchell implemented EnKF for atmospheric data assimilation for the medium-range ensemble forecasting systems in various of their work (Houtekamer and Mitchell 1998; 2001; Houtekamer et al. 2005).

Assimilating data to predict the chemical state of the troposphere is even much more computationally costly than weather forecasting. This is because atmospheric chemistry involves chemical reactions and the number of parameters taken into account in the calculations is greater than in meteorology (even higher if aerosol chemistry is considered too). The application of the data assimilation techniques to the atmospheric trace gases only began in 1993 with Hartley and Prinn. They investigated the feasibility of using KF method to determine regional surface fluxes for CFCs by comparing observations and models in 3D and found that the higher differences between the model and observations lead higher biases in the estimations (Hartley and Prinn 1993). Haas-Laursen, Hartley, and Prinn 1996 used KF to constrain trace gases and to investigate their temporal source/sink cycle. Napelenok et al. 2008 developed a method using KF, to constrain current NO_x emission inventories by NO_2 column observations from GOME satellite instrument. Mijling and van der A 2012 developed an algorithm (so-called DESCO) using KF method for short-lived species. Miyazaki et al. 2017 estimated global surface NO_x emissions for a long term from 2005 to 2014 using EnKF and various observation sources (e.g., OMI, SCIAMACHY satellite instruments).

2.1.2.4 Variational approach

Meanwhile, in addition to the sequential approaches, in the 80-90s, so-called variational approaches using optimal control theory were introduced into the 3D variational (3D-Var) method and later into the 4D variational (4D-Var) method, which considers the spatial and temporal distributions of observations (Phillips 1982; Dimet and Talagrand 1986; Talagrand and Courtier 1987; Courtier et al. 1998; Rabier et al. 1998). The 3D-Var and 4D-Var approaches aim to combine observations and background information in an optimal way (indeed using optimization techniques e.g., descent algorithm) to produce the best possible estimate of the state of the system (their formulation is explained in the following sub-sections). It should be noted that, under linearity assumptions, Kalman filter method, its variants and the 4D-Var provide the same results at the end of the assimilation period (Lorenc 1986), since they all satisfy BLUE (explained in the section 4D-Var and its pioneering 3D-Var).

First, in 1995 Fisher and Lary, performed 4D-Var approach for the stratosphere to chemically active trace gases using a reduced photochemical box model and satellite measurements using NO_2 observations from the Cryogenic Limb Array Etalon Spectrometer (CLAES) Instrument onboard the Upper Atmosphere Research Satellite (UARS) (Fisher and Lary 1995). They developed the adjoint model of reduced stratospheric gas phase mechanism combined with the box model. Later, Eskes et al. 1999 implemented 4D-Var approach using 2D model and total ozone column data from GOME to estimate global ozone field, in particular over the poles.

For the troposphere, first in 1997, Elbern et al. developed and studied the feasibility of the 4D-Var with the adjoint model of a tropospheric gas phase mechanism by assimilating the O_3 ground-based measurements with the assimilation window of hours only (Hendrik Elbern, Schmidt, and Ebel 1997). They mentioned the necessity of a 3D meteorological modeling and its adjoint to include diffusion and convection operators, and cloud parameters. Few years later, in 1999, Elbern and Schmidt presented the first application of the 4D-Var method for a 6h data assimilation window with a full comprehensive regional 3D Eulerian CTM and its adjoint, including chemical reactions, transport, and diffusion by using surface observations such as NO_2 and O_3 to obtain a chemical state of the troposphere (Hendrik Elbern and Schmidt 1999). They found the 4D-Var method to be feasible and useful but emphasized the importance of having appropriate prior knowledge on the concentration levels to be used as a background. So far, the parameters to be optimized were initial values, such as concentration levels. However, H. Elbern et al. 2000 optimized the emissions within an assimilation window of 6h. They investigated the performance of the 4D-Var to estimate NO_x and VOCs' (O_3 precursors) emission rates by considering SO_2 and O_3 surface observations. Hendrik Elbern and Schmidt 2001 focused on the optimal analysis of chemical evolution of the ozone episode occurred over central Europe. This was the first 4D-Var application on a real case event by assimilating the surface observations of NO_x and O_3 as well as O_3 radiosonde records with an extended data assimilation window to 9h. However, a longer assimilation window, such as 24 hours representing a full day, can be a strong constraint as it can reflect the diurnal profile of emissions. Thus, H. Elbern et al. 2007 performed 4D-Var within 24h assimilation window to estimate emission rates of all gaseous species emitted at each surface grid point using surface in-situ observations and a full comprehensive regional CTM over Europe.

In addition to the ground-based networks used in the assimilations, dedicated flight missions and various satellite instruments also used as a source of independent information. For example, Talbot 2003 performed the first tropospheric airborne assimilation using 4D-Var to determine reactive nitrogen species over the Western Pacific. Müller and Stavrakou 2005 performed the first co-assimilation of annual global emissions (at continental scales) of different ozone precursor chemical species (CO and NO_x) within 6 hours of assimilation window, using observations from the GOME satellite instrument. The simultaneous assimilation of CO and NO_x can account better the chemical response of tropospheric ozone production cycle as they are both taking a role in the cycle. However, the reliability of the optimized emissions are not only depend on the independent observations but also chemical feedback between other species.

Kurokawa et al. 2009 developed an assimilation system with a parameterized NO_x chemistry scheme in order to optimize NO_x emissions over eastern China using GOME instrument. They run an assimilation system over a daily assimilation window containing observations from different times. Since then, there have been many studies that have widely used the 4D-Var assimilation approach in air quality assessments using different satellite observations and CTMs at a spatial resolution of a few kilometers to several hundred square kilometers (from a point source to a global) according to the interests (Pison, Menut, and Bergametti 2007; Boisgontier et al. 2008; Chai et al. 2009; Fortems-Cheiney et al. 2021; Qu et al. 2022).

In the last decade, hybrid approaches have received increasing attention and have become a popular research area (e.g., Ensemble 4D Variational (4D-EnVar)). These approaches (called hybrid) combine the two main methods, variational and statistical, by aiming to exploit the advantages of both methods, and they are still evolving (Bannister 2017). Moreover, there are studies in which these approaches are supported by deep learning models with the aim of providing more robust results (C. A. Keller and Evans 2019; Petetin et al. 2020; He et al. 2022).

2.1.3 General notations of data assimilation

This section defines standardized notations and describes the main elements of data assimilation that form the basis of the model theory used in this study. Ide et al. 1997 proposed these standardized notations to help the development of more advanced data assimilation methods.

2.1.3.1 The state vector, control variable and observations

2.1.3.1.1 The state vector

Let us consider a discrete dynamical model (e.g. CTMs) for a physical system (e.g., atmospheric, oceanic) that evolves over time. The state of this model is represented by a collection of numbers equivalent to a column matrix called the **state vector** $\boldsymbol{x}$ of dimension *n* in the state space $\boldsymbol{X}$. If we define $\boldsymbol{t_i}$ as a time index at the *i*-th time, $\boldsymbol{x(t_i)}$ and $\boldsymbol{x(t_{i+1})}$ become two state vectors at the consecutive times $\boldsymbol{t_i}$ and $\boldsymbol{t_{i+1}}$. This system from time $\boldsymbol{t_i}$ to time $\boldsymbol{t_{i+1}}$ can be described as follows:

$$\boldsymbol{x(t_{i+1}) = \mathcal{M}_{t_i \to t_{i+1}} x(t_i)} \qquad \text{(Eq.2)}$$

where $\boldsymbol{\mathcal{M}}$ is the dynamics operator of a computer model. It can be non-linear and consists of a discretization of non-linear partial differential equations (PDEs). It is assumed that the model is deterministic and perfect, so more or less it represents the reality to avoid considering the model error. However, a model is limited by its resolution, so it cannot simulate the real or **true state** ($\boldsymbol{x^t}$) of the system. Therefore, the state vectors represent an approximation of the state of the atmosphere. Before the assimilation experiment, the true state can be estimated by the so called first guess of the true state or **background state** ($\boldsymbol{x^b}$). This is often called initial or prior state and mostly derived from a CTM model, a previous prediction or a climatology before the **analysis** ($\boldsymbol{x^a}$) is carried out.

2.1.3.1.2 Control variable

In the complex systems, it is not easy to solve the analysis problem for all components of the state vector. It is because either the interactions between components are not completely well known (e.g., chemical interactions in an atmospheric model) or we are limited with the computational power. For this reason, the analysis problem corresponds to the search for a correction or **increment** ($\boldsymbol{\delta x}$) of the background which belongs to a subspace called **control variable** space as follows:

$$\boldsymbol{x^a = x^b + \delta x} \qquad \text{(Eq.3)}$$

where $\boldsymbol{x^a}$ is as close as possible to $\boldsymbol{x^t}$. By this transition, essentially, our problem is to search for $\boldsymbol{x^a - x^b}$ instead of searching for $\boldsymbol{x^a}$.

2.1.3.1.3 Observations

The collection of all available observations (e.g., from spectrometers, radars, in-situ measurements, satellite instruments) are defined by a variable **observation vector** $\boldsymbol{y}$ in the observation space $\boldsymbol{Y}$. All the parameters of the control space may not be observable and are not necessarily observed. So, there are fewer observations than parameters in the model. Moreover, to use the information contained in the observation vector, they need to be comparable with the state vector. However, observations are not regularly disposed and they do not correspond to a point of the grid of the model. Therefore, in order to compare the observations to the model, a function of $\boldsymbol{\mathcal{H}}$ or an **observation operator** is defined. It maps the state space onto observation space. $\boldsymbol{\mathcal{H}}$ is supposed to be known, most of the time it is assumed to be linear. If it is linear, then a matrix $\mathbf{H}$, called tangent linear observation operator, exists. However, it can also be complex and non-linear.

Noting that, if the dimension of the observation vector $\boldsymbol{y}$ at *i*-th time is $\boldsymbol{p_i}$, usually $\boldsymbol{p_i} \ll \boldsymbol{n}$, (*n* being the dimension of the state space) because all the parameters of the state space may not be observed. There are not enough observations to estimate the state of space accurately. This fact is causing the inverse modeling being an ill-posed problem. Nevertheless, it is important to note that for some applications the dimensions of both state vector and observation vectors are quite large (≈ e.g., from 10^6 to 10^8).

2.1.3.2 Analysis, innovation and residuals

Analysis ($\boldsymbol{x^a}$) is an estimation process of the true state of a system for a given time using a combination of $\boldsymbol{\mathcal{M}}$ computational model, $\boldsymbol{y}$ independent observations from instruments and $\boldsymbol{x^b}$ background information as well as initial and boundary conditions. It is the result of the so-called **data assimilation** process. During 4D-Var assimilation process, the **assimilation window** defines the time window of all observations considered at once. Data assimilation uses all this information to estimate the best possible representation of the true state. An important element of an analysis is the use of discrepancies between the observations and the state vector mapped by the operator as follows:

$$\boldsymbol{y} - \boldsymbol{\mathcal{H}}(\boldsymbol{x}) \tag{Eq.4}$$

When it is calculated using $\boldsymbol{x^b}$, the difference is called **innovations** and when $\boldsymbol{x^a}$ is used the term is called **analysis residuals**.

2.1.3.3 Errors

One of the crucial elements in data assimilation is the representation of the unknown errors associated with the uncertainties found in the background, observations and analysis. Even if it is difficult to determine the exact value of the errors, it is possible to characterize them using probability density functions (PDFs) and using statistics associated to the PDFs. Therefore, it is not expected to know what the exact error is but statistics are calculated for the estimation of this error. The statistical formulas are recalled in Annex A: Statistical formulas.

Background errors represent the difference between the background state and true state of the system as follows:

$$\boldsymbol{\varepsilon^b} = \boldsymbol{x^b} - \boldsymbol{x^t} \tag{Eq.5}$$

Its covariance **B** (**background error covariance matrix**) is represented by

$$\mathbf{B} = \mathbb{E}\left(\boldsymbol{\varepsilon^b}\boldsymbol{\varepsilon^b}^T\right) = \overline{\left(\boldsymbol{\varepsilon^b} - \overline{\boldsymbol{\varepsilon^b}}\right)\left(\boldsymbol{\varepsilon^b} - \overline{\boldsymbol{\varepsilon^b}}\right)^{\mathrm{T}}} \tag{Eq.6}$$

where $\mathbb{E}$ represents the mathematical expectation, T represents the transposition and the average is $\overline{\boldsymbol{\varepsilon^b}}$. The diagonal terms of the **B** matrix are the variances and give the confidence that we have in the background information. The off-diagonal terms of the matrix give how this information propagates spatially from the background knowledge locations. Therefore, they critically influence the outcome of the inversion. Background error consists of misrepresentation of the physical processes of the model or systematic and random errors. However, constructing a **B** matrix is a challenging task. Sometimes there are experts who determine and make available these values using different methods (e.g. Monte Carlo simulations) to determine the effects of uncertainties in the background information.

Observation errors represent the difference between the observations and the corresponding state in the reality as follows:

$$\boldsymbol{\varepsilon^o} = \boldsymbol{y} - \boldsymbol{\mathcal{H}}(\boldsymbol{x^t}) \tag{Eq.7}$$

with the average of $\overline{\boldsymbol{\varepsilon^o}}$. Its covariance **R** (**observation error covariance matrix**) is represented by

$$\mathbf{R} = \mathbb{E}\left(\boldsymbol{\varepsilon^{o}}\boldsymbol{\varepsilon^{o}}^{T}\right) = \overline{(\boldsymbol{\varepsilon^{o}} - \overline{\boldsymbol{\varepsilon^{o}}})(\boldsymbol{\varepsilon^{o}} - \overline{\boldsymbol{\varepsilon^{o}}})^{\mathrm{T}}} \qquad \text{(Eq.8)}$$

R is assumed to be diagonal because observations are considered mutually uncorrelated from each other. It contains instrumental errors (measurement error and errors raising during the conversion of satellite measurement into concentrations or columns), the modeling errors comes from the definition of the observation operator ($\boldsymbol{\mathcal{H}}$ **or H**), and representativeness errors due to the discretization of PDEs.

Analysis errors represent the difference between the analysis and the true state of the system as follows:

$$\boldsymbol{\varepsilon^{a}} = \boldsymbol{x^{a}} - \boldsymbol{x^{t}} \qquad \text{(Eq.9)}$$

with the average of $\overline{\boldsymbol{\varepsilon^{a}}}$. A measure of amplitude $\overline{\|\boldsymbol{\varepsilon}^{\mathrm{a}} - \overline{\boldsymbol{\varepsilon}^{\mathrm{a}}}\|}$ is provided by the trace of the **analysis error covariance matrix A**,

$$\mathrm{Tr}(\mathbf{A}) = \overline{\|\boldsymbol{\varepsilon}^{\mathrm{a}} - \overline{\boldsymbol{\varepsilon}^{\mathrm{a}}}\|^{2}} \qquad \text{(Eq.10)}$$

The optimization results with in an analysis error smaller than the background error. The averages calculated over the errors are called **biases**.

2.1.3.4 Hypotheses

For simplification, the concept and the techniques of data assimilation rely on some assumptions as follows (Lorenc 1986; Daley 1991; Ghil 1989):

- Tangent linear hypothesis (TLH) allows the observation operator $\boldsymbol{\mathcal{H}}$ to be linearized (denoted by **H**) , e.g., the variables ($\boldsymbol{x}$) of the observation operator in the neighborhood of the background ($\boldsymbol{x^{b}}$) are assumed to be linear:

$$\boldsymbol{\mathcal{H}}(\boldsymbol{x}) - \boldsymbol{\mathcal{H}}\left(\boldsymbol{x^{b}}\right) = \mathbf{H}(\boldsymbol{x} - \boldsymbol{x^{b}}) \qquad \text{(Eq.11)}$$

- Errors $\boldsymbol{\varepsilon^{b}}$ and $\boldsymbol{\varepsilon^{o}}$ are non-trivial: The covariance matrices **B** and **R** are positive definite.
- Errors $\boldsymbol{\varepsilon^{b}}$ and $\boldsymbol{\varepsilon^{o}}$ are unbiased: $\mathbb{E}\left(\boldsymbol{\varepsilon^{b}}\right) = \mathbb{E}(\boldsymbol{\varepsilon^{o}}) = \overline{\boldsymbol{x^{b}} - \boldsymbol{x^{t}}} = \overline{\boldsymbol{y} - \boldsymbol{\mathcal{H}}(\boldsymbol{x^{t}})} = \mathbf{0}$
- Errors $\boldsymbol{\varepsilon^{b}}$ and $\boldsymbol{\varepsilon^{o}}$ are uncorrelated: $\overline{(\boldsymbol{x^{b}} - \boldsymbol{x^{t}})\left(\boldsymbol{y} - \boldsymbol{\mathcal{H}}(\boldsymbol{x^{t}})\right)^{\boldsymbol{T}}} = \mathbf{0}$

- The analysis $\boldsymbol{x}^{\boldsymbol{a}}$ is linear: The analysis, defined by corrections to the background state, depends linearly on the differences between the background and the observations.
- The analysis $\boldsymbol{x}^{\boldsymbol{a}}$ is optimal: The analyzed state is as close as possible to the true state in accordance with the root mean square, e.g., an estimate of the minimum variance.

2.1.4 Assimilation techniques

All data assimilation techniques rely on the classical Bayesian inversion framework with Gaussian assumptions (in other words, PDF of error is fully described by their mean and covariance). Having a priori information about the state of the atmosphere, the search for analysis is the search for the most probable state of the atmosphere knowing the observations. Therefore, the Gaussian Bayesian formulation consists in computing the following PDF:

$$p(\boldsymbol{x}|\boldsymbol{y}^{\boldsymbol{o}}, \boldsymbol{x}^{\boldsymbol{b}}) \sim N(\boldsymbol{x}^{\boldsymbol{a}}, \mathbf{A})$$

where the solution of an inverse problem $\boldsymbol{x}^{\boldsymbol{a}}$ is a random vector with a Gaussian or normal distribution assigned with $\boldsymbol{N}$.

2.1.4.1 Statistical data assimilation

The aim is to produce an estimate $\boldsymbol{x}^{\boldsymbol{a}}$ of the true input parameters $\boldsymbol{x}^{\boldsymbol{t}}$ of the system. To this end, statistical interpolation (also called OI, which is a simpler version of KF) used to combine two independent sources of information. This method allows writing them as a linear combination between the background ($\boldsymbol{x}^{\boldsymbol{b}}$) vector and observations ($\boldsymbol{y}$) vector as follows:

$$\boldsymbol{x}^{\boldsymbol{a}} = \boldsymbol{x}^{\boldsymbol{b}} + \mathbf{K}\left(\boldsymbol{y} - \boldsymbol{\mathcal{H}}(\boldsymbol{x}^{\boldsymbol{b}})\right) \qquad \text{(Eq.12)}$$

in which an innovation $\boldsymbol{y} - \boldsymbol{\mathcal{H}}(\boldsymbol{x}^{\boldsymbol{b}})$ is weighted by the linear operator $\mathbf{K} = \mathbf{BH^T(HBH^T + R)^{-1}}$ called gain or weight matrix. Noting that, $\mathbf{H}$ is the tangent linear version of the operator $\boldsymbol{\mathcal{H}}$**.** Under the assumptions above, the best estimate of $\boldsymbol{x}^{\boldsymbol{a}}$ in Eq.12 is called **BLUE-Best Linear Unbiased Estimator** via least squares analysis. In other words, BLUE predicts an unbiased, optimal weighting for a linear correction of the state of the model from the observations. Thus, the problem seeks to find the best unbiased estimation of $\boldsymbol{x}$ that minimizes the variance of the analysis errors (Eq.9) (optimality criterion). The covariance matrix of the analysis error $\mathbf{A}$ is calculated using the following error equivalent in Eq.13 (This equation is obtained using Eq.5, Eq.7, and Eq.9 in Eq.12):

$$\boldsymbol{\varepsilon^a} = \boldsymbol{\varepsilon^b} + \mathbf{K}\left(\boldsymbol{\varepsilon^o} - \mathbf{H}(\boldsymbol{\varepsilon^b})\right) \qquad \text{(Eq.13)}$$

Therefore, for all **K** (not necessarily to be an optimal),

$$\begin{aligned}\mathbf{A} = \mathbb{E}(\boldsymbol{\varepsilon^a}\boldsymbol{\varepsilon^{a^T}}) &= \mathbb{E}\left[\left(\boldsymbol{\varepsilon^b} + \mathbf{K}\left(\boldsymbol{\varepsilon^o} - \mathbf{H}(\boldsymbol{\varepsilon^b})\right)\right)\left(\boldsymbol{\varepsilon^b} + \mathbf{K}\left(\boldsymbol{\varepsilon^o} - \mathbf{H}(\boldsymbol{\varepsilon^b})\right)\right)^{\mathrm{T}}\right] \\ &= \mathbb{E}\left[\left((\mathbf{I} - \mathbf{KH})\boldsymbol{\varepsilon^b} + \mathbf{K}\boldsymbol{\varepsilon^o}\right)\left((\mathbf{I} - \mathbf{KH})\boldsymbol{\varepsilon^b} + \mathbf{K}\boldsymbol{\varepsilon^o}\right)^{\mathrm{T}}\right] \\ &= \mathbb{E}\left[\left((\mathbf{I} - \mathbf{KH})\boldsymbol{\varepsilon^b}\left(\boldsymbol{\varepsilon^b}\right)^{\mathrm{T}}(\mathbf{I} - \mathbf{KH})^{\mathrm{T}}\right)\right] + \mathbb{E}[\mathbf{K}\boldsymbol{\varepsilon^o}(\boldsymbol{\varepsilon^o})^{\mathrm{T}}\mathbf{K}^{\mathrm{T}}] \\ &= (\mathbf{I} - \mathbf{KH})\mathbf{B}(\mathbf{I} - \mathbf{KH})^{\mathrm{T}} + \mathbf{KRK}^{\mathrm{T}} \qquad \text{(Eq.14)}\end{aligned}$$

where **I** is the identity matrix. If **K** is the optimum gain (so it minimizes the scalar error Eq.10 or the matrix entries along the principal diagonal of **A**), then the analysis error covariance matrix becomes $\mathbf{A} = (\mathbf{I} - \mathbf{KH})\mathbf{B}$. Full details of the computation of **K** and **A** can be found in Burgers, Jan van Leeuwen, and Evensen 1997.

2.1.4.2 Variational data assimilation

As seen in the previous section, BLUE requires the direct computation of an optimal gain matrix **K**. However, in the complex systems, this matrix is very large and therefore expensive in computing time. Nevertheless, BLUE can be obtained without computation of the gain **K** using variational calculus, e.g., through the minimization of a functional. Therefore, variational methods, based on the optimal control theory, are promising in data assimilation. They consist of minimizing the least squares functional. The proof of their equivalence to the optimal interpolation problems can be found in (Lorenc 1986; Talagrand 1997).

2.1.4.2.1 Formulation of variational data assimilation

Let a state vector $\boldsymbol{x}(\boldsymbol{t}) \in \boldsymbol{\mathcal{X}}$ and time $\boldsymbol{t} \in \mathbf{T}$, which contains any physical variable(s) of the state of the system (e.g., chemical compositions for air quality modeling). In reality, the evolution of the state is continuous and described by a set of nonlinear differential equations in a set of $\boldsymbol{\Omega}$ as follows:

$$\frac{d\boldsymbol{x}}{dt} = \boldsymbol{\mathcal{M}}(\boldsymbol{x}) \quad \text{in } \Omega \times [0, \mathrm{T}] \qquad \text{(Eq.15)}$$
$$\boldsymbol{x}\,(t = 0) = \boldsymbol{x}_0$$

where the true initial condition is not known. Supposing that $\boldsymbol{y(t)} \in \boldsymbol{\mathcal{O}}$ are observations of the true state and $\boldsymbol{\mathcal{H}}$ is observation operator, in the continuous context, **the objective $\boldsymbol{J}$ function** (also called **cost function**) defines the distance of the state of the atmosphere with respect to the background and the observations as follows:

$$J(\boldsymbol{x}) = \frac{1}{2}\int_0^{\mathrm{T}} \left\| \boldsymbol{y}(t) - \boldsymbol{\mathcal{H}}\big(\boldsymbol{x}(\boldsymbol{x}_0, t)\big) \right\|_{\boldsymbol{\mathcal{O}}}^2 dt + \frac{1}{2}\left\| \boldsymbol{x} - \boldsymbol{x^b} \right\|_{\boldsymbol{\mathcal{X}}}^2 \qquad \text{(Eq.16)}$$

While in the first term, integral calculates the distance between observations and modelled state under the observation operator in the infinite and continuous context, the second term comes from a regularization, as inverse problems are known to be ill posed and quantify the misfit to the background state. Adding the second term makes the problem well posed. The problem would then be to search for $\boldsymbol{x^a}$ that minimizes the cost function J:

$$\boldsymbol{x^a} = argmin\, J(\boldsymbol{x}) \qquad \text{(Eq.17)}$$

This minimization requires numerical optimization techniques. Quasi-Newtonian descending algorithms are used to solve this minimization problem iteratively with high dimensionality, starting with an initial state vector $\boldsymbol{x}_0$. For the existence of a (local) minimum of Eq.17, the gradient of the cost function is necessarily $\nabla J(\boldsymbol{x^a}) = 0$. However, the calculation of the derivatives of the cost function at each step with high dimensions can be computationally expensive. For this reason, these gradients are obtained by running an adjoint model simultaneously in parallel to the inverse model. This is why variational data assimilation is also known as ‘adjoint’ method.

2.1.4.2.2 4D-Var and its pioneering 3D-Var

This subsection introduces 4D-Var data assimilation formulation and its adjoint, as well as its pioneering 3D-Var method in a finite-dimensional and discrete context. In the finite-dimensional case, the two norms under the integral (Eq.16) are represented by the error covariance matrices **R** and **B**, respectively. The observation operator $\boldsymbol{\mathcal{H}}$ and the computational model $\boldsymbol{\mathcal{M}}$ are assumed to be linear (thus the cost function is quadratic) according to the hypotheses made earlier and defined as their Jacobian matrix **H** and **M**, respectively. Their adjoint $\boldsymbol{\mathcal{H}}^{*}, \boldsymbol{\mathcal{M}}^{*}$ can be identified by the transpose for the Jacobian $\mathbf{H}^{\mathrm{T}}$ and $\mathbf{M}^{\mathrm{T}}$. Under these assumptions, the direct model $\mathbf{M}$ is defined by,

$$\begin{aligned} \boldsymbol{x}_i &= \mathbf{M}_i(\boldsymbol{x}_{i-1}) \ , \qquad i = 1, n \\ \boldsymbol{x}_0 &= \boldsymbol{x}^{\boldsymbol{b}} \end{aligned} \qquad \text{(Eq.18)}$$

where the state trajectory is $\boldsymbol{x} = (\boldsymbol{x}_0, \boldsymbol{x}_1, \dots, \boldsymbol{x}_{n-1}, \boldsymbol{x}_n,)^{\mathrm{T}}$, n is the number of time steps per assimilation window. Then, for a fixed time, the Eq.16 would define 3D-Var cost function as follows,

$$J(\boldsymbol{x}) = \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}^{\boldsymbol{b}})^{\mathrm{T}}\mathbf{B}^{-1}(\boldsymbol{x} - \boldsymbol{x}^{\boldsymbol{b}}) + \frac{1}{2}(\boldsymbol{y} - \mathbf{H}(\boldsymbol{x}))^{\mathrm{T}}\mathbf{R}^{-1}(\boldsymbol{y} - \mathbf{H}(\boldsymbol{x})) \qquad \text{(Eq.19)}$$

And the gradient of the J function is given by:

$$\nabla J(\boldsymbol{x}) = \mathbf{B}^{-1}(\boldsymbol{x} - \boldsymbol{x}^{\boldsymbol{b}}) - \mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}(\boldsymbol{y} - \mathbf{H}(\boldsymbol{x})) \qquad \text{(Eq.20)}$$

where **B** and **R** are stated as the background and observation error covariance matrices respectively. Calculations using these matrices are not computationally storable and manipulable due to the high dimensionality of the problem. Therefore, the main idea of the variational method is to minimize the cost function J in order not to deal with these difficulties. This minimization is achieved by an adjoint model involving the calculation of the gradient of the cost function and a suitable descent algorithm in the gradient direction. This minimization can be stopped by a stopping criterion that sets a decrease in the norm of the gradient of J ($\|\nabla J(\boldsymbol{x})\|$) small enough or artificially limits the maximum number of iterations. When there is no time dependence, like in the case of 3D, the adjoint model is not necessary. Because the adjoint simply includes the transpose of **H** matrix (Eq.20). However, the 3D-Var method is a particular case of the 4D-Var method. While 3D-Var does not take into account temporal dimension in minimization process, 4D-Var method carries the minimization over a time interval. The process involves the observations, which are obtained at different times, and considers all the observations at once in the assimilation window (e.g., in the Figure 14 the assimilation window is from $\boldsymbol{t}_0$ to $\boldsymbol{t}_n$ (Lahoz and Schneider 2014)). Therefore, the adjoint model is required for the transpose of both the model and operator dynamics.

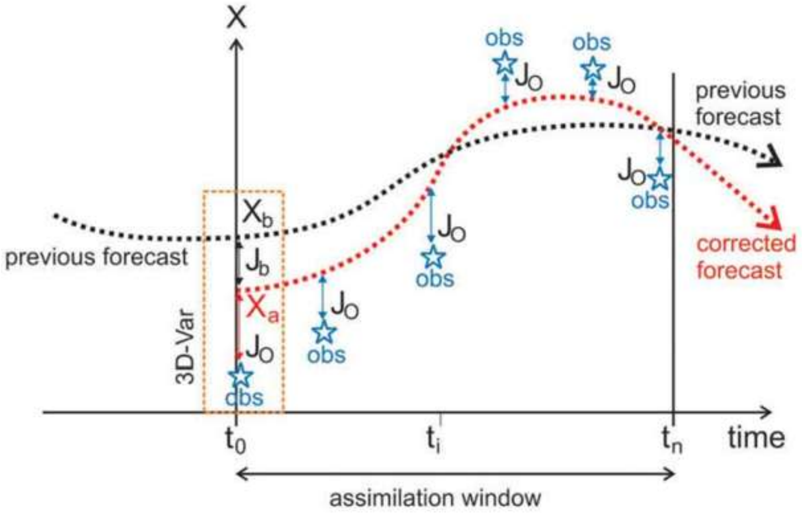


Figure 14 Illustration of 3D-Var & 4D-Var (Taken from Lahoz and Schneider, 2014)

As a result, 4D-Var data assimilation scheme is defined by the minimization of a cost function $J(\boldsymbol{x}) = J^{\boldsymbol{b}}(\boldsymbol{x}) + J^{\boldsymbol{o}}(\boldsymbol{x})$ from time $\boldsymbol{i} = \boldsymbol{0}$ to $\boldsymbol{i} = \boldsymbol{n}$ as follows:

$$J(\boldsymbol{x}) = \underbrace{\frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}^{\boldsymbol{b}})^{\mathrm{T}}\mathbf{B}^{-1}(\boldsymbol{x} - \boldsymbol{x}^{\boldsymbol{b}})}_{J^{b}(x)} + \underbrace{\frac{1}{2}\sum_{\mathrm{i}=0}^{\mathrm{n}}(\boldsymbol{y}_i - \mathbf{H}_{\mathrm{i}}(\boldsymbol{x}_i))^{\mathrm{T}}\mathbf{R}_{\mathrm{i}}^{-1}(\boldsymbol{y}_i - \mathbf{H}_{\mathrm{i}}(\boldsymbol{x}_i))}_{J^{o}(x)} \qquad \text{(Eq.21)}$$

where the state $\boldsymbol{x}$ at time i is obtained by an iterative composition of the model $\mathbf{M}$, e.g., $\boldsymbol{x}_i = \mathbf{M}_i\mathbf{M}_{i-1} \ldots \mathbf{M}_2\mathbf{M}_1\boldsymbol{x}$. Then the cost function J is as follows,

$$J(\boldsymbol{x}) = \frac{1}{2}\left(\boldsymbol{x} - \boldsymbol{x}^{\boldsymbol{b}}\right)^{\mathrm{T}}\mathbf{B}^{-1}\left(\boldsymbol{x} - \boldsymbol{x}^{\boldsymbol{b}}\right) + \frac{1}{2}\sum_{\mathrm{i}=0}^{\mathrm{n}}\mathbf{d}_i^{\mathrm{T}}\mathbf{R}_i^{-1}\mathbf{d}_i \qquad \text{(Eq.22)}$$

where the innovation vector $\mathbf{d}_i = \boldsymbol{y}_i^{\mathbf{o}} - \mathbf{H}_i\mathbf{M}_i\mathbf{M}_{i-1} \ldots \mathbf{M}_1(\boldsymbol{x})$. Then the gradient of the $\nabla J(\boldsymbol{x})$,

$$\nabla J(\boldsymbol{x}) = \mathbf{B}^{-1}\left(\boldsymbol{x} - \boldsymbol{x}^{\boldsymbol{b}}\right) - \sum_{\mathrm{i}=0}^{\mathrm{n}}\mathbf{M}_{\mathrm{i}}^{\mathrm{T}}\mathbf{H}_i^{\mathrm{T}}\mathbf{R}_i^{-1}\mathbf{d}_i \qquad \text{(Eq.23)}$$

By this way, the 4D-Var algorithm calculates first J followed by ∇J one direct model integration and one adjoint model integration. One can find the evolution of Eq.22 and Eq. 23 from Eq.21 in Annex B: 4D-Var Cost and the Gradient Evolution.

2.1.4.2.3 4D-Var algorithm

- *Initialization*: $i = 0$, $\boldsymbol{x^b} = \boldsymbol{x}_0$
- While $\|\nabla J\| > \boldsymbol{\delta}$ or $i \leq i_{\max} = n$:
- Compute J using the direct model $\mathbf{M}$ and the observation operator $\mathbf{H}$ (forward mode)
- Compute ∇J using the backwards integration of the adjoint model $\mathbf{M}^{\mathrm{T}}$ and the adjoint of the observation operator $\mathbf{H}^{\mathrm{T}}$ (inverse mode)
- Gradient descent minimization and update of $\boldsymbol{x}$
- $i = i + 1$

At the initialization step, trajectory of the model is determined by the state of the starting point $\boldsymbol{x}_0$ and set it to the background $\boldsymbol{x^b}$. This is a first guess. At Step 1, Model runs forward in time to the end of the assimilation window to obtain the first-guess trajectory using Eq.18 and Eq.22. Using the first guess trajectory from Step 1, Step 2 integrates the adjoint model backwards in time from t_i to t_0 using Eq.23. In Step 3, the optimal state of the cost function is calculated using the outputs from Step 1 and Step 2 via Gradient Descent Minimization (Figure 15, Bouttier and Courtier 1999) and set to the state variable. This procedure iteratively continues via Step 4 until it reaches the maximum number of steps (i_{max}) or the gradient of the cost function is sufficiently smaller than a delta ($\boldsymbol{\delta}$) value determined by the user.

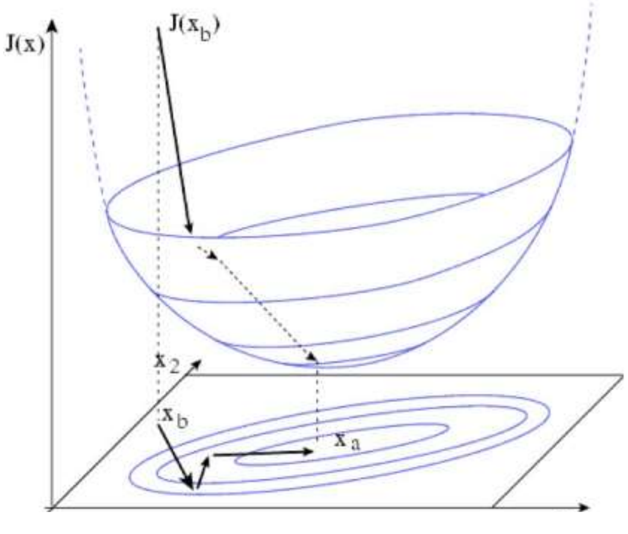


Figure 15 Visualization of gradient descent minimization for the state $\mathbf{x}$ *in 2D.* J *is a quadratic convex cost function. Its minimum is obtained* $\mathbf{x}^a$. *The image taken from, Bouttier and Courtier, ECMWF,1999*

2.1.4.2.4 Limitations of 4D-Var

Although the 4D-Var technique is powerful and robust, it is not easy to implement because it is highly dependent on the problem (linearity, dimensionality) and makes the calculations computationally expensive because the algorithm relies on the availability of derivative information (e.g., coding of the adjoint operators). The adjoint model is a mathematical trick used to calculate sensitivities between model parameters outside the forward model. However, developing an adjoint model is computationally costly and consists of more complex equations than forward model. At the same time, each adjoint model is specific to the forward model, so any changes in the forward model also require recoding in the adjoint model.

Also, the development of background $\mathbf{B}$ and the observation error covariance matrix $\mathbf{R}$ are challenging and require attention. Because they give perception about how much we trust the given information (e.g., how much we trust the observations), any misled specification of these matrices may strongly influence the results of the inversion, especially at high resolutions.

Noting that the model operator $\mathbf{M}$ and observation operator $\mathbf{H}$ are assumed to be linear here. It is known that, under these linearity assumptions, 4D-Var and Kalman filter methods provide the same results at the end of the assimilation period (Lorenc 1986). However, in the real world problems, they are non-linear, and non-linearity of the model $\boldsymbol{\mathcal{M}}$ and observation operator $\boldsymbol{\mathcal{H}}$ making the objective function non-convex, and its gradient complex. In this case, the BLUE formalism is not valid anymore. Thanks to the tangent linear hypothesis (TLH) and adjoint models (Lewis and Derber 1985), non-linear operators are replaced by their tangent linear operators (Annex C: Non-linear operators), resulting in the problem being an approximation of 4D-Var. Courtier, Thépaut, and Hollingsworth 1994 proposed the use of the TL model ($\boldsymbol{\delta}$) using the increment vector in the classical formulation of 4D-Var (called incremental 4D-Var) and solving the problem in the neighborhood of the background state. In this case, the 4D-Var algorithm includes systematic and random errors raised by the linearization.

In the framework of this thesis, the variational approach of the CIF inverse modeling system is used. The next section is therefore devoted to an introduction about CIF and the development of CIF for our inversion problem.

2.2 CIF inverse modeling system

This thesis adapts the Community Inversion Framework (CIF), a state-of-the-art inverse modeling system. This system is developed to unify atmospheric inversion problems and bring together researchers interested in atmospheric inversions (Berchet et al. 2021). It is a user-friendly, easily accessible, transparent, and open-source Python-based tool on http://community-inversion.eu/ (Last access: 08/04/2023). It is used to estimate the atmospheric states, more specifically, fluxes of various greenhouse gases and reactive species by having the flexibility to run CTMs (Lagrangian or Eulerian) at different spatial (regional or global) and temporal scales using different types of observations (e.g., in-situ, satellite) in different data assimilation techniques (e.g., analytical, statistical, variational).

The next section explains how the CIF is structured in this thesis.

2.2.1 Development of CIF in this study

In this thesis, to inverse NO_x emissions using satellite observations, we used variational approach of the CIF (more specifically 4D-Var). Although the variational approach offers a numerical approximation to the solution of the inverse problem, the formulation of the problem does not require complex matrix products. Therefore, the high dimensionality of the matrices is not an obstacle during the calculation process (Berchet et al. 2021). Furthermore, inversion within the assimilation window makes it possible to consider non-linear links between emissions and concentrations along trajectories on time scales equal to the chemical lifetime of the studied species. For this reason, due to the high diurnal characteristics of NO_2 concentrations and their non-linearity with NO_x emissions, the 4D-Var data assimilation technique with 24h assimilation windows (that is daily temporal resolution) is used. Next, it is described the main elements of the 4D-Var used within CIF.

2.2.1.1 The control vector ($\boldsymbol{x}$)

The control vector contains all the parameters related to the species to be optimized in the inversion problem (e.g., fluxes and/or initial and boundary concentration conditions). In this thesis as we aim to estimate NO_x emissions, the control vector contains corrections coefficients (or multiplicative factors) to be applied to the relevant NO_x emissions information (including anthropogenic and biogenic NO_x emissions) at a temporal resolution of 1-day. The dimensions of the control vector is as follows; hours in a day × level of altitude × number of latitudinal grid cells × number of longitudinal grid cellsThe daily corrections obtained from the inversions are applied to the background NO_x emissions ($\boldsymbol{x^b}$) as multiplicative factors. The relationship between the multiplicative factors of the control vector and background NO_x emissions (prior) is then as follows:

$$Estimated\ NOx\ emissions\ (posteriori) = \boldsymbol{x}\ .\ Prior\ NOx\ emissions$$

2.2.1.2 The background vector ($\boldsymbol{x^b}$)

The background vector is an initial state (Eq.18) of the inversion problem. It contains *a priori* information to be considered by the inversion. It is linked to the background NO_x daily emission fluxes (or prior NO_x emissions). The background vector is defined with multiplication factors of 1 as these are reference emissions from bottom-up emission inventories, which will be described in detail in Section 2.3. Similar to the control vector, the background vector has the dimensions as follows; hours in a day × level of altitude × number of latitudinal grid cells × number of longitudinal grid cells.

2.2.1.3 The background error covariance matrix ($\mathbf{B}$)

It is important to note that, the uncertainties in the prior NO_x emissions are assumed to be unbiased, have a Gaussian distribution and are characterized by their **B** covariance matrix. Its diagonal terms are the weights to the components of the vector differences between the control vector and prior NO_x emissions. On the other hand, non-diagonal terms of the covariance matrix give uncertainties about the spatial correlations of the NO_x emission sources. In our configuration, they are calculated using exponentially decaying functions with the correlation length scale derived from the model resolution. As it is difficult to construct **B** matrix standard estimates of prior NO_x emissions, uncertainties from the literature are used as a starting point in this thesis instead of complicated ones.

2.2.1.4 The observation operator ($\mathbf{H}(\boldsymbol{x})$)

The observation operator includes a computation of CTM, which is used to link NO_x surface fluxes to NO_2 atmospheric concentration measurements and projection of simulated NO_2 concentrations to the NO_2 tropospheric column measurements by the satellite instrument. By this way, the background emissions become comparable with the satellite instrument measurements. The CTM model used in this thesis is explained in Section 2.3.

2.2.1.5 The observation vector ($\boldsymbol{y}$)

The observation vector brings independent information to the inversion system using atmospheric observations. These observations can come from ground-based measurements, such as trace gas mixing ratios from fixed or mobile platforms, and remote sensing observations, such as satellite observations. In this thesis, satellite observations of the NO_2 tropospheric column are used because of their high spatial coverage in the area of interest. Section 2.4 explains the types of satellite instruments of which observations are used in this thesis.

2.2.1.6 The observation error covariance matrix ($\mathbf{R}$)

As with the **B** matrix, the observation errors are assumed to be unbiased and Gaussian distributed. In this thesis, errors from the model are neglected and all the diagonal terms of the **R** matrix are taken as the measurement errors reported in the satellite data sets. Additionally, spatial error correlations between NO_2 satellite observations are neglected. Therefore, all non-diagonal terms of the **R** matrix are taken as zero.

2.2.1.7 The minimizer M1QN3

The minimization subroutine M1QN3, which uses the cost function and its adjoint as input, is used to minimize the inversion problem. The minimizer written in Fortran was described by (Gilbert and Lemaréchal 1989; Gilbert, Lemaréchal, and Simulator 2009). It is distributed under the GNU General Public License. It is used to solve large-scale unconstrained minimization problems using a quasi-Newton algorithm. Quasi-Newton methods are designed to use on linear problems. It iteratively moves the cost function along the direction of its steepest descents, which is, calculated an approximation of the Hessian matrices and by the adjoint. However, non-linearity of the chemistry of NO_x reactive species makes our inversion problem non-linear and the cost function non-convex. For this reason, the quasi-Newton algorithm can reach the local minima without reaching a global minimum. To solve this issue, this method needs an initial regularization of the control vector that we want to minimize. In other words, instead of minimizing $\boldsymbol{x}$ we minimize $\boldsymbol{\chi} = \mathbf{B}^{-\frac{1}{2}}(\boldsymbol{x} - \boldsymbol{x}^{\boldsymbol{b}})$. A simpler formulation of the adjoint of the cost function J to be applied regulation is as follows:

$$\nabla J(\boldsymbol{x}) = \mathbf{B}^{-1}(\boldsymbol{x} - \boldsymbol{x}^{\boldsymbol{b}}) - \boldsymbol{\mathcal{H}}^{*}\left(\mathbf{R}^{-1}(\boldsymbol{\mathcal{H}}(\boldsymbol{x}) - \boldsymbol{y})\right)$$

The formulation of the adjoint of the cost function J becomes:

$$\nabla J(\boldsymbol{\chi}) = \boldsymbol{\chi} - \mathbf{B}^{\frac{1}{2}}\boldsymbol{\mathcal{H}}^{*}\left(\mathbf{R}^{-1}\left(\boldsymbol{\mathcal{H}}\left(\mathbf{B}^{\frac{1}{2}}\boldsymbol{\chi} + \boldsymbol{x}^{\boldsymbol{b}}\right) - \boldsymbol{y}\right)\right)$$

The minimizer algorithm is required to meet the convergence criterion before reaching the stopping criterion. In cases where this criterion is n ot met, the algorithm is terminated by the stopping criterion and the optimization is considered non-convergent. In other words, this means that a suitable state vector that minimizes the cost function has not been found. The convergence criterion we adopt in this thesis is as follows; consider the cost function to have converged if, after minimization, the norm of its gradient has decreased by 95 % compared to the initial one. If such a vector that minimizes the cost function cannot be found, stop running the algorithm with at most 20 iterations.

2.2.1.8 Analysis ($\boldsymbol{x}^{\boldsymbol{a}}$)

The optimal estimate of the NO_X emissions is found by iteratively minimizing the cost function J and is so-called the analysis, or posterior NO_X emissions. Figure 16 presents the iterative structure of the CIF used for our inversion problem.

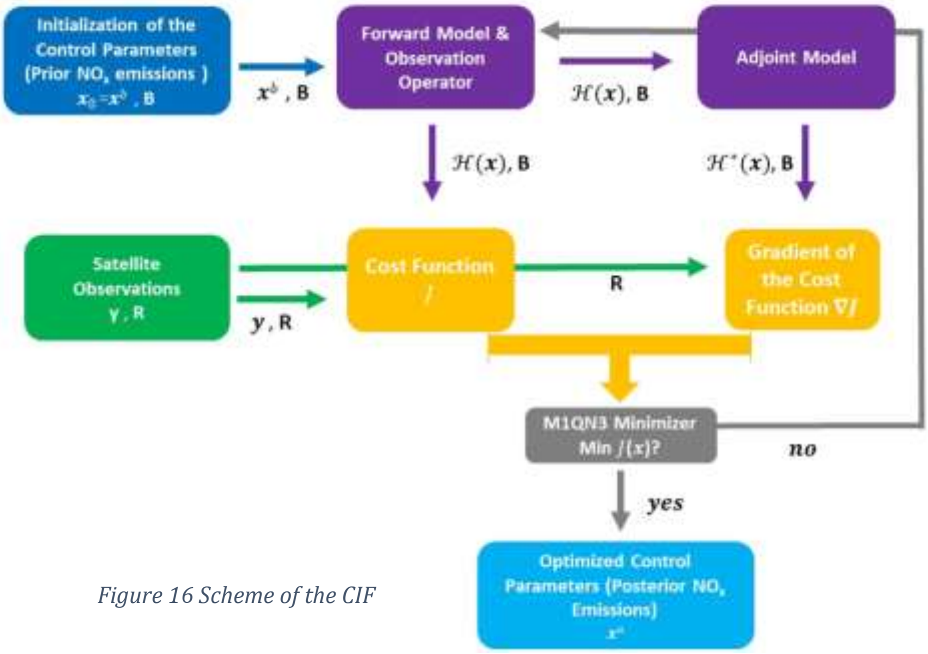


Figure 16 Scheme of the CIF

2.3 Atmospheric Modeling

Atmospheric modeling consists of a set of mathematical equations representing the physical and chemical interactions of the atmosphere. It can simulate the evolution of atmospheric concentrations, taking into account all atmospheric physical and chemical phenomena within and at the boundaries of the study area (domain), but also taking into account emissions and meteorological parameters used as forcing data (e.g., precipitation, wind, temperature). The simulation of atmospheric concentrations is performed using CTMs based on two main approaches: (i) The Lagrangian approach and (ii) The Eulerian approach.

The Lagrangian approach considers a pollutant on the flow and simulates the trajectory of the air parcel with respect to its direction due to the wind. Therefore, the simulation is performed at different locations and different times. The Eulerian approach, on the other hand, considers the domain over which the pollutant moves as fixed and discretized (e.g., by boxes or cells), and simulations are carried individually within these discrete cells (so-called ‘grid cells’). Chemical composition calculations are performed for each grid cell. The size of the grid cells determines the resolution of the CTM. A 3D CTM model is a model that performs simulations at the desired vertical and horizontal resolutions, taking into account the interactions between each grid cell (both vertically and horizontally), and determines the concentrations over the entire domain. It is necessary to take into account the transport between the grid cells and the physical/chemical phenomena in the grid cells in order to calculate the concentrations in each grid cell. For this reason, simulations at each grid cell are performed based on the following mass balance equation to determine the concentration c of the species s at time t.

$$\frac{\mathrm{dc}_s}{\mathrm{d}t} = \left(\frac{\mathrm{dc}_s}{\mathrm{d}t}\right)_{\text{chemistry}} + \left(\frac{\mathrm{dc}_s}{\mathrm{d}t}\right)_{\text{advection}} + \left(\frac{\mathrm{dc}_s}{\mathrm{d}t}\right)_{\text{emission}} + \left(\frac{\mathrm{dc}_s}{\mathrm{d}t}\right)_{\text{dilution}} + \left(\frac{\mathrm{dc}_s}{\mathrm{d}t}\right)_{\text{deposition}} \tag{Eq.24}$$

Let us represent the chemistry balance term $\left(\frac{\mathrm{dc}_s}{\mathrm{d}t}\right)_{\text{chemistry}}$ by R_s which is the net chemical production rate of the species s at time t. The advection term can be expressed as follows,

$$\left(\frac{\mathrm{dc}_s}{\mathrm{d}t}\right)_{\text{advection}} = \frac{u}{\Delta \mathrm{x}}(\mathrm{c}_s^0 - \mathrm{c}_s)$$

where u is the wind speed, Δx is the length of the grid cell, c_s^0 is the background concentration of the species s. So the proportion of the wind speed to the length of the grid cell determines the residence time of the air parcel over the domain, which reduces some of the background concentration of the species s. Sometimes the source of the pollutant is inside of the grid cell so the emission term can be expressed as follows,

$$\left(\frac{\mathrm{d}c_s}{\mathrm{d}t}\right)_{\mathrm{emission}} = \frac{q_s}{h}$$

which represents the emission rate per unit area ($\mathrm{kgm^{-2}h^{-1}}$), where q_s is the quantity of the emission of the species s and h is the height of the grid cell. The dilution balance term $\left(\frac{\mathrm{d}c_s}{\mathrm{d}t}\right)_{\mathrm{dilution}}$ can be neglected when the grid cells have constant height because there will not be any transfer entraining or detraining of the concentrations into or out of the grid cell. The deposition term can be expressed using removal rate as follows,

$$\left(\frac{\mathrm{d}c_s}{\mathrm{d}t}\right)_{\mathrm{deposition}} = -\frac{v_{d_s} c_s}{h}$$

where v_{d_s} is dry or wet deposition velocity of the species s.

Overall, the balance of the species s can therefore be expressed with the simplified mass balance equation as follows:

$$\frac{\mathrm{d}c_s}{\mathrm{d}t} = R_s + \frac{u}{\Delta x}(c_s^0 - c_s) + \frac{q_s}{h} - \frac{v_{d_s} c_s}{h} \qquad \text{(Eq.25)}$$

This equation includes all source and sink terms of chemical constituents in the CTM; therefore, solving this equation gives the concentration of a given species s for a given time t (so-called output). Figure 17 visualizes the mass balance conception of the concentration of a species s at one grid cell and construction of the 3D CTM.

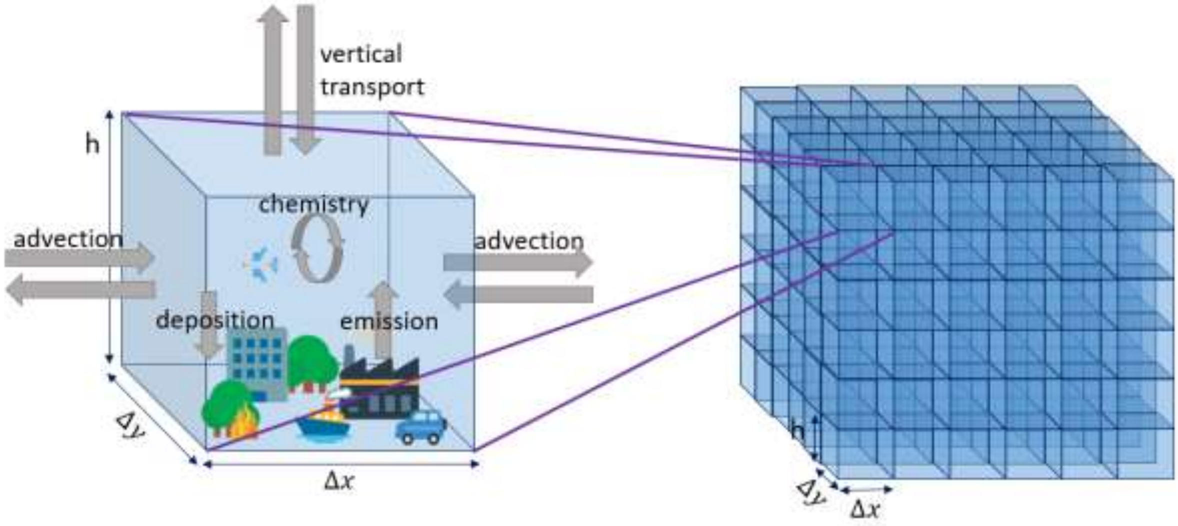


Figure 17 Illustration of simplified mass balance processes represented in each grid cell of a 3D CTM

For each grid cell, solving the mass balance equation numerically consists of the calculation of the sources (e.g., emissions) and sinks (e.g., deposition) while considering the interaction between neighboring grid cells. To do this, some information (so-called input) must be provided to the CTMs before the simulations launch. For example, anthropogenic and biogenic emissions from emission inventories or other CTMs need to be provided as input. In order to determine atmospheric conditions and provide information about the atmospheric situation, meteorological conditions (e.g., temperature, wind, cloud, humidity) need to be defined through offline or online CTM modeling environments and meteorology models. While in an offline setting, a CTM is used to run after the meteorology model determines meteorological conditions in advance, in an online setting, there is back and forth relationship between the CTM and the meteorology model. Moreover, a chemical scheme has to be provided for chemical reaction types and rates of species including aerosol chemistry. Assuming that a CTM simulation covers only part of the globe, meaning that simulations run over a smaller domain, in this case, boundary conditions of the domain need to be presented to take into account transport from neighborhoods. Generally, these conditions are determined running another simulation over the larger domain (usually with lower resolutions compared to the smaller domain). These kinds of simulations are called nested-simulations. Moreover, initial conditions are presented to provide information about the existing atmospheric chemical conditions of the domain, which are fed to the model in the same manner as boundary conditions. Noting that boundary and initial conditions are necessary only for the first time-step. Finally, land use information needs to be presented.

CTMs have advantages that turn them into complementary tools for the measures observed. They provide continuous and potentially very complete information relative to physical sensing instruments. They make it possible to quantify interactions between different processes. On the other hand, like other man made tools, they face some limitations. They can only process the state of the information we have already defined. For example, they cannot take into account information from an as-yet-unknown chemical reaction. As such, they are only a representation of reality, and need to be validated to ensure that the representation is faithful enough for the use of simulations. Furthermore, the amount of information that needs to be used as input is detailed and these inputs also contain some limitations and uncertainties. For example, uncertainties in emissions used in CTMs. However, CTMs are still the only and necessary tool to reconstruct inaccessible information, to predict concentrations anywhere at any time in the past or the future.

In the framework of this thesis, in order to perform an inversion of NO_x emissions on a regional scale, it is necessary to run a CTM to link NO_x emission fluxes and their ambient concentrations. To do that, the regional CTM CHIMERE is used, since it is one of the CTMs whose adjoint has been coded (Laurent Menut 2003) and coupled with the CIF inversion system.

2.3.1 CHIMERE Regional CTM

CHIMERE regional chemistry and transport model has been co-developed since 1997 by Laboratoire de Météorologie Dynamique (LMD), Institut national de l'environnement industriel et des risques (INERIS) and Laboratoire Interuniversitaire des Systèmes Atmosphériques (LISA). It is a 3D offline (and online for the recent versions) Eulerian model designed to run long-term simulations for emission control scenarios and forecasting air pollution events within the atmospheric boundary layer. It uses external forcing (e.g., meteorological fields, pollutant emissions, boundary conditions) to simulate hourly concentrations of gaseous and aerosol species on a regular Cartesian grid from the surface up to the level of interest with resolutions from 1-2 km up to 100-200 km. CHIMERE simulated concentrations were compared and evaluated in different resolutions with various types of measurements from surface stations (Menut et al. 2000; Blanco-Ward et al. 2021; Menut et al. 2020), airborne or various satellite instruments for tropospheric columns (Huijnen et al. 2010; Konovalov et al. 2009).

This thesis uses the CHIMERE v2013b version, which is integrated into the CIF without activating the aerosol module (Menut et al. 2013). This is because the adjoint model of

CHIMERE (Laurent Menut 2003), which helps to calculate the sensitivity of simulated atmospheric concentrations to corrections in fluxes, was developed without including the aerosol module due to its high computational cost. Therefore, the CIF-CHIMERE only considers full gaseous chemistry, not aerosols. Considering aerosols in the model can lead to different simulated NO_2 concentrations as they significantly affect the photochemical process. As an example, we checked the impact of aerosols in the simulation only in the Paris region (within 15 km), where the consideration of aerosol leads to a reduction of the simulated NO_2 concentrations by 5 to 15%.

CHIMERE uses hybrid sigma-pressure $(\sigma - p)$ coordinates which allows user to define any number of vertical layers. The formula to calculate the pressure (in hPa) at the top of each layer k, p_k is as follows: $p_k = a_k p_0 + b_k p_{surf}$ where p_{surf} is the surface pressure, p_0 is the reference state pressure, a_k and b_k are the hybrid coefficients determined by either the user or the model. More detailed information about CHIMERE v2013b and its adjoint can be found in (Menut et al. 2013; Laurent Menut 2003; Fortems-Cheiney et al. 2021).

2.3.1.1 Setup of the CHIMERE Simulations

In the framework of this thesis, the study domains include two principal urbanized regions of the Northern Hemisphere: Eastern China and Europe (studied in the Chapters 3 and 4 respectively); hence, CHIMERE’s configurations vary depending on these regions. Therefore, the setup of the CHIMERE simulations and their inputs are explained separately according to these domains.

2.3.1.1.1 Domain : Eastern China

CHIMERE setup for the NO_x emission inversion to simulate NO_2 concentrations over Eastern Asia covers the regular grid with horizontal resolution of 0.5° × 0.5° and 17 vertical levels from the surface up to 200 hectopascal (hPa) with 8 layers within the first three kilometers. It includes 65 longitude (lon) × 61 latitude (lat) grid cells that correspond to 18°N – 50°N (degrees north) and 102°E – 132°E (degrees east) and covers eastern and southern China, the South China Sea and Korean Peninsula domain as in the Figure 18.

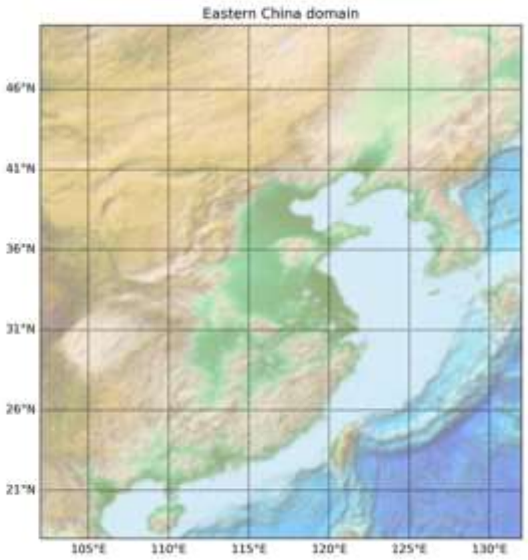


Figure 18 Eastern China domain visualization

To be able to perform simulations over Eastern China domain, the necessary ingredients used as inputs for CHIMERE CTM is summarized in the Figure 19.

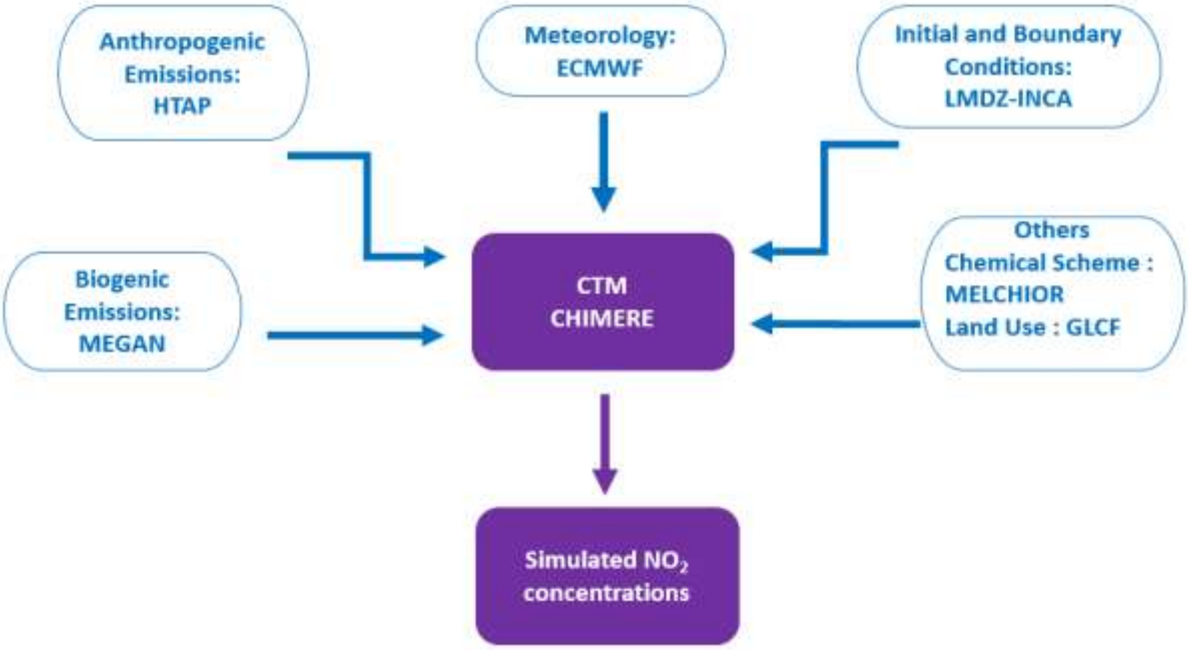


Figure 19 Main inputs and their sources for CTM CHIMERE simulations over Eastern China

Boundary and initial conditions of the domain are given by a climatology from the global Laboratoire de Météorologie Dynamique Zoom (LMDz) model that is coupled with the model Interaction with Chemistry and Aerosols (INCA) (Hauglustaine et al. 2004). Meteorological conditions are coming from the global meteorological forecasts of European Centre for Medium-Range Weather Forecasts (ECMWF) and hourly interpolated to fit the hourly simulations (Owens and Hewson 2018). MELCHIOR-2 gaseous chemical scheme which is the reduced MELCHIOR chemical mechanism (Lattuati 1997; Derognat 2003) comprising 44 species and 120 reactions is used to evaluate tropospheric chemical evolution of gaseous species in CHIMERE (Menut et al. 2013; Derognat 2003). Land use information is coming from the database Global Land Cover Facility (GLCF) of University of Maryland (CHIMERE User's Manual). Biogenic emissions are derived from the Model of Emissions of Gases and Aerosols from Nature (MEGAN) v2.1 model (Guenther et al. 2006). And finally, anthropogenic emissions are those from Hemispheric Transport of Air Pollution (HTAP) v2.2 bottom-up inventory (Janssens-Maenhout et al. 2015). It is important to note that, HTAP v2.2 anthropogenic emissions are taken as a reference (or prior) bottom-up emission inventory in the inversion to estimate NO_x emissions for Eastern China; therefore, the following part is dedicated to HTAP v2.2 prior bottom-up inventory. It is worth noting that, in this study, NO_x emissions are expressed as nitrogen oxides (NO_2 and NO) as budgets with the unit of kilotons (hereinafter kt NO_x or Tg NO_x referred to as kt or Tg).

2.3.1.1.1.1 Anthropogenic Bottom-up Emission Inventory: HTAP

HTAP are used as the prior inventory in our inversions (Henze, Hakami, and Seinfeld 2007). For the region of China, HTAP v2.2 is constructed using regional emission inventories in collaboration with the MICS Asia Scientific Community (MICS-Asia) and REAS experts with a horizontal resolution of 0.25° × 0.25° converted to 0.1° × 0.1°. Emissions are estimated from seven main sectors of anthropogenic activities: air, ships, energy, industry, transport, residential, and agriculture.

In this study, The EDGAR-HTAP V2.2 emissions inventory (Janssens-Maenhout et al. 2015) based on the year 2010 was used to generate anthropogenic emissions of gaseous species. The emissions are initially grouped by sector of activity (SNAP, Standardized Nomenclature for Air Pollutants); they are then broken down to correspond to the actual pollutants, and then aggregated to match the species represented in CHIMERE (a construction of emissions of the type (Middleton, Stockwell, and Carter 1990). Emissions are spatially distributed and daily hourly profiles emissions are allocated (in Coordinated Universal Time (UTC)) for each day of the week.

2.3.1.1.1.2 Anthropogenic Bottom-up Emission Inventory: MEIC

MEIC (Multi-resolution Emission Inventory for China) is a bottom-up emission inventory developed and operated at Tsinghua University since 1990 (http://meicmodel.org.cn/, Last access: 08/04/2023). It is designed to track anthropogenic emissions over continental China at different spatial resolutions. NO_x emissions are mainly presented in five sectors: energy, transportation, industry, agriculture, and residential. For more details about MEIC inventory construction and sectors, see (Li, Liu, et al. 2017; Zheng et al. 2018).

In this thesis work, MEIC version 1.3 from 2015 (Zheng et al. 2018; 2021) is used as an independent inventory to evaluate the estimated NO_x emissions for the same year. Note that this version of MEIC includes the impacts of the 12th Five-Year Plan mentioned in Section 1.6.

2.3.1.1.2 Domain : Europe

CHIMERE setup for the NO_x emission inversion to simulate European NO_2 concentrations covers the regular grid with horizontal resolution of 0.5° × 0.5° and 20 vertical levels from the surface up to 200 hPa with 8 layers within the first three kilometers. It includes 101 lon × 85 lat grid cells that correspond to -15.25°N – 34.75°N and 31.75°E – 73.75°E and covers Europe (Figure 20).

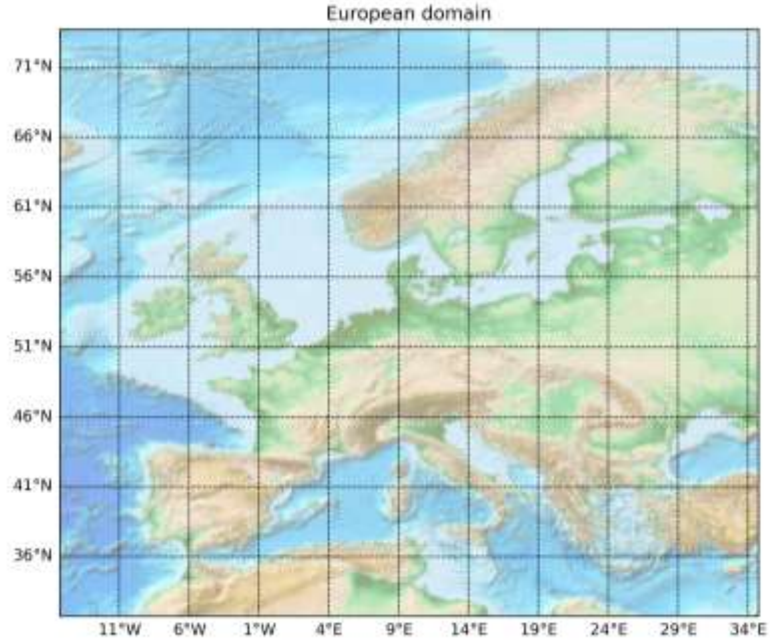


Figure 20 European domain visualization

To be able to perform the simulations, the inputs variables of CHIMERE are as follows (Figure 21): similar to the Eastern China domain (Figure 19), meteorological forecasts are obtained from ECMWF, the chemical scheme MELCHIOR-2 is used. Initial, lateral and boundary conditions are given by a climatology from LMZD-INCA global model. The biogenic emissions are those from MEGAN model as well. Moreover, anthropogenic emissions are provided by Copernicus Atmosphere Monitoring Service (CAMS). In this study, CAMS anthropogenic NOx emissions from 2016 are taken as a reference (or prior) bottom-up emission inventory in the inversion to estimate NO_x emissions for Europe; therefore, the following part is dedicated to CAMS prior bottom-up inventory.

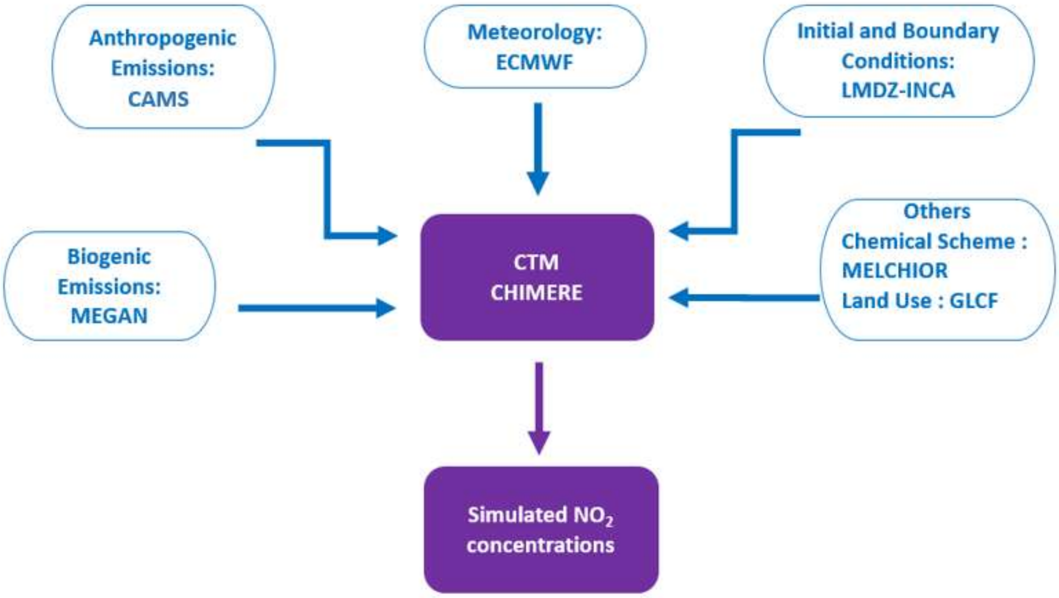


Figure 21 Main inputs and their sources for CTM CHIMERE simulations over Europe

2.3.1.1.2.1 Anthropogenic Bottom-up Emission Inventory: CAMS

CAMS, conducted by ECMWF on behalf European Union, consists of global and regional European anthropogenic emissions for various sectors. The regional CAMS (CAMS-REG) inventory has a spatial resolution of 0.1° × 0.05°. It covers GHG emissions (referred to as CAMS-REG-GHG) and air pollutants (referred to as CAMS-REG-AP). It consists of emissions based on national reporting and a combination of other inventories (e.g., GAINS, EDGAR, TNO inventory estimations) (J. J. P. Kuenen et al. 2014; J. Kuenen et al. 2022) and aggregated across different sectors and fuel combinations (Table 2.2, Granier et al. 2019). Annual NO_x emissions are presented in the sectors such as; energy, residential, transportation, industry, and agriculture. For the documentation describing CAMS inventories, construction and sectors, see Granier et al. 2019.

In this study, the CAMS-REG-AP inventory v3.1 for 2016 published in 2019 was used as anthropogenic emissions.

2.4 Atmospheric observations

In the framework of this thesis, NO_x emissions are constrained using NO_2 tropospheric column observations from the OMI and TROPOMI satellite instruments. The satellite observations subsection is devoted to explaining the nature of these observations and the criteria by which the observation data were selected for the inversions.

After performing the inversions, the obtained NO_x emissions estimations were evaluated by considering independent surface observations. The surface observations subsection is devoted to the description of the ground-based networks at different domains (Eastern China and Europe) stated in this thesis.

2.4.1 Satellite observations

2.4.1.1 OMI

The Ozone Monitoring Instrument (OMI) is an ultraviolet-visible (UV-Vis) nadir-view solar backscatter spectrometer carried by the National Aeronautics and Space Administration's (NASA) Earth Observing System (EOS) Aura satellite launched in July 2004 (Figure 22). The instrument is a contribution of the Netherlands' Institute for Air and Space Development (NIVR), in collaboration with the Finnish Meteorological Institute (FMI).

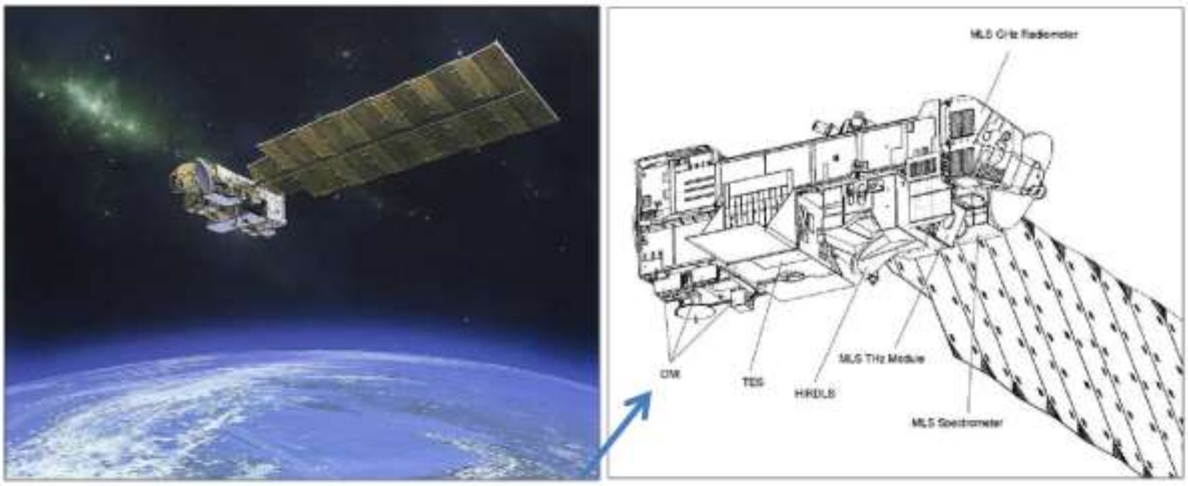


Figure 22 Visualization and illustration of the Aura Spacecraft (images credit: Northrop Grumman Space Technology NGST)

OMI has a wide field of view of 2600 km (Figure 23). This provides near global coverage with 14 orbits per day. The ground pixel size at nadir have a spatial resolution of (nadir) 13km × 24km (along × across-track) and a local overpass time of 13:40 local time (LT) (Levelt et al. 2006).

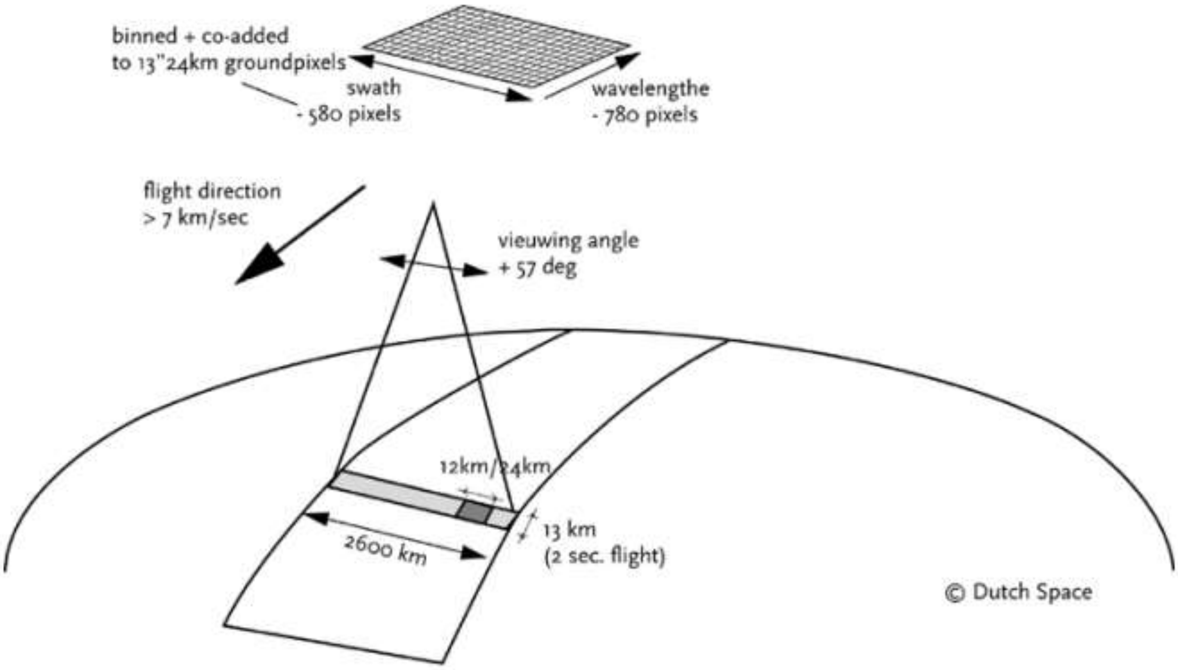


Figure 23 OMI measurement geometry (Taken from Levelt et al., 2006)

It is a successor instrument to the European Global Ozone Monitoring Experiment (GOME) (with a spatial resolution of 320 km × 40 km) (Burrows et al. 1999), the Scanning Imaging Absorption Spectrometer for Atmospheric Cartography (SCIAMACHY) (with a spatial resolution of 30 km × 60 km) (Bovensmann et al. 1999) and NASA's Total Ozone Mapping Spectrometer (TOMS) (with a spatial resolution of 50 km × 50 km) (Wellemeyer et al. 1997) instruments. OMI is the most sophisticated instrument of its time in terms of high resolution and the ability to measure the entire UV-Vis-Near Infrared (NIR) spectrum. Figure 24 shows an example of the spectrum over the Netherlands in a clear sky.

Thanks to its two 2D charge coupled device (CDD) detectors, it covers UV1 (270–310 nm), UV2 (310–365 nm) and VIS-channel (365-500 nm) with a spectral sampling of 0.21 nm and a spectral resolution of 0.63 nm (Boersma et al. 2007). Signal-to-noise ratio (SNR) is estimated to be 400 for UV2 and 500 for VIS for 1 pixel (Zara et al. 2018, Table 1). OMI was initially designed to continue the TOMS total ozone record, so it includes TOMS-type of retrievals as well as other types of retrievals (e.g., Differential Optical Absorption Spectroscopy (DOAS) type, Raman scattering). It also provides mapping of air pollution products such as NO_2, SO_2 and aerosols from an urban to regional scale (Boersma et al. 2007; 2011). This study uses OMI

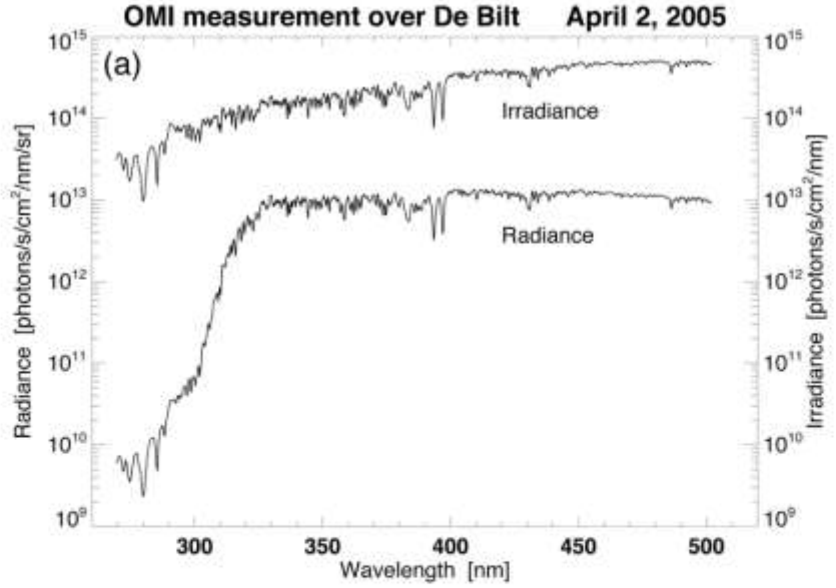


Figure 24 Solar irradiance and the Earth's radiance spectrum measured by OMI over the Netherlands. The image taken from Levelt et al., 2006

NO_2 tropospheric columns v1.1 retrievals from the European Quality Assurance for Essential Climate Variables (QA4ECV) project (Boersma et al. 2018). This version of the OMI observations (so-called QA4ECV NO2 data product) consists of an improved spectral fitting method for the retrieval of NO_2 slant columns (a photon path from the Sun to the atmosphere and satellite using DOAS spectral fitting procedure (Platt and Stutz 2008). Air mass factor (AMF) is another key element in the retrieval process. It is used to correct and convert the slant columns density (SCD) to vertical columns (VCD) density (AMF = SCD / VCD). However, it is calculated by using radiative transfer model and dependent on the model input parameters such as, albedo and clouds. It also requires a priori estimation of the vertical profiles of the species of interest. So, AMF calculations are subject to uncertainties. OMI QA4ECV retrieval includes data assimilation using the TM5-MP CTM model in $1° \times 1°$ degree horizontal resolution to estimate stratospheric background NO_2 to strengthen the AMF and reduce the uncertainty associated with the AMF (Zara et al. 2018; Lorente et al. 2017) and thus improves the retrieval. Averaging kernels (AKs) correlate the derived quantities and the true profile during the retrieval (Rodgers 2000). They give information about the vertical sensitivity of the satellite retrievals, vary from one pixel to another and are used to minimize the discrepancies of model/satellite comparisons. AMFs used to calculate the tropospheric averaging kernels for below the tropopause as follows:

$$A^{trop} = \frac{M}{M^{trop}} A \qquad \text{(Eq.26)}$$

where AK and tropospheric AK is denoted by the symbols A and A^{trop}; AMF and tropospheric AMF is denoted by M, and M^{trop} , respectively.

As in this thesis, we are interested in NO_2 tropospheric columns, it should be noted that OMI tropospheric NO_2 column data are defined as an integrated quantity from the surface to the tropopause for a specified surface and often expressed in molecule/cm^2. Validations studies show that OMI QA4ECV NO_2 column data is negatively biased or even gives good agreement (e.g., bias = -2 % over the polluted Tai'an (China) region in June 2006 (Boersma et al. 2018) compared to the surface observations from stations (Compernolle et al. 2020). For documentation of the dataset, see Boersma et al. 2017, user's guide. In addition, the dataset is freely accessible via the Tropospheric Emission Monitoring Internet Service (TEMIS) (https://www.temis.nl/airpollution/no2.php, Last access: 08/04/2023).

2.4.1.1.1 Data selection and processing

In this thesis, we follow the OMI user's manual data quality recommendations to use the QA4ECV NO_2 data. The selection of OMI observation data is based on the following criteria: (i) pixels with a processing error flag equal to 0 are selected for successful processing; (ii) pixels with the solar zenith angle larger than 80 are selected to avoid observations from polar nocturnal darkness; (iii) pixels selected with the snow ice flag lower than 10 (ice coverage with max 10%) or equal to 255 (that is a flag for the ice-free ocean); (iv) pixels with the ratio of tropospheric AMF over geometric AMF is higher than 0.2 to avoid situations in which the retrieval is based on very low (relative) tropospheric AMFs; and (v) pixels with the cloud fraction is lower or equal than 0.5 meaning that most of the radiance originates from the clear-sky part of the pixel. These limit the contribution of a priori profile made for the calculation of AMF; (vi) In addition to these recommendations, we consider an additional criterion concerning the error. Data with an uncertainty of more than 100% are filtered out.

After the selection and filtering procedure, observation data needs to be processed to become comparable with the model (J^o observation cost function in Eq. 21). Since the satellite data are not on a regular grid (as in the model), and the spatial resolution of the OMI data is finer than the CHIMERE model grid cells, we sampled the observation data into super-observations. To do that, we chose the median of the OMI NO_2 tropospheric columns within the $0.5^\circ \times 0.5^\circ$ grid cell and its corresponding AKs.

As we are dealing with the discrepancies between simulated and observed NO_2 concentrations, to compare with OMI super-observations, hourly-simulated NO_2 concentrations in the same grid cells are selected and are interpolated to OMI's pressure levels. The AKs are then applied to obtain a model estimate with the vertical sensitivity of the tropospheric NO_2 columns (Eq.27).

$$\mathrm{y} = \sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{y}_i^{\mathrm{s}} \mathrm{A}_i \qquad \text{(Eq.27)}$$

where y is the model equivalent of the satellite data, n is the number of levels of the satellite observations, $\mathrm{y}_i^{\mathrm{s}}$ is the simulated concentration interpolated at the i-th level and AK is denoted by the following symbol A_i at the i-th level. Note that, these operations take place in the **H** observation operator, which includes the CHIMERE CTM to relate emissions to concentrations and transformations to calculate the simulated equivalent of observations as explained in the Section 2.2.

2.4.1.2 TROPOMI

TROPOspheric Monitoring Instrument (TROPOMI) is a nadir view solar backscatter instrument on board European Space Agency's (ESA) Copernicus Sentinel-5 Precursor (S5P) satellite, launched into a Sun-synchronous orbit in October 2017. Similar to OMI, TROPOMI has a wide field of view of 2600 km (Figure 25) and provides near global coverage with 14 orbits per day.

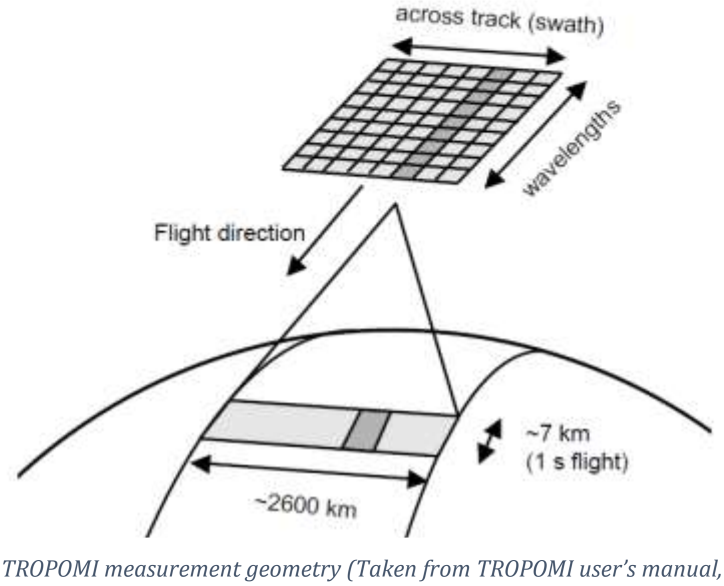


Figure 25 TROPOMI measurement geometry (Taken from TROPOMI user's manual, 2019)

It is a nadir imaging spectrometer with a high spatial resolution of 3.5 km × 5.5 km (initially 3.5 km × 7 km changed on 6 August 2019) making it the most advanced instrument available today including, its predecessor, the OMI (Figures 26 and 27).

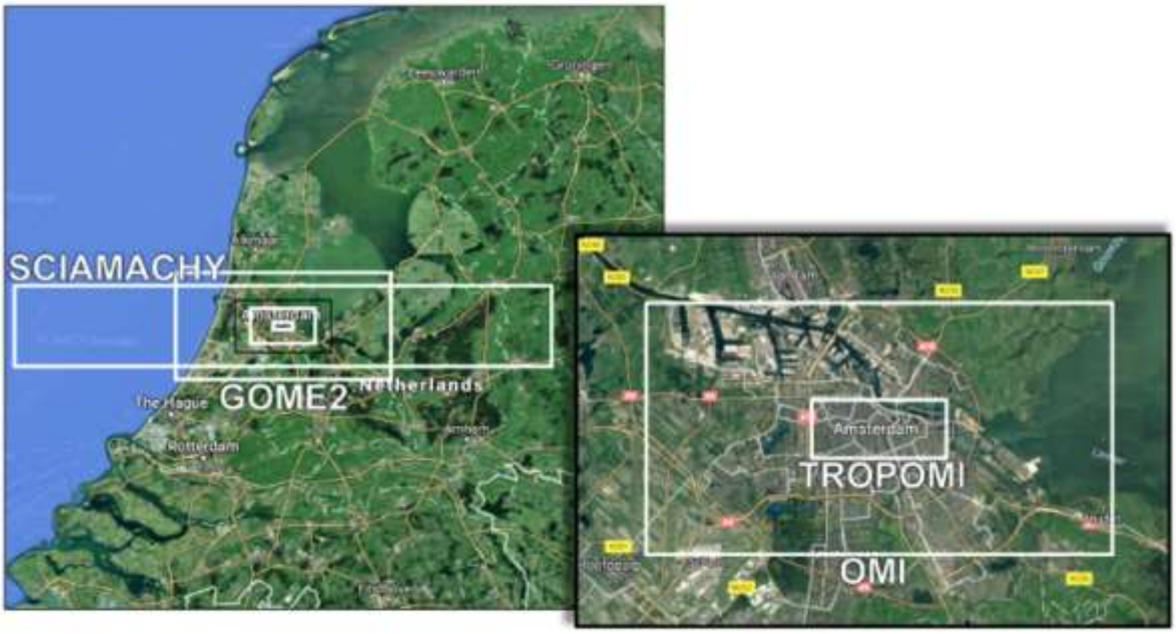


Figure 26 Spatial resolution comparison of satellite instruments. Note that TROPOMI pixel size further decreased now (Taken from BIRA-IASB, www.aeronomie.be/en/news/2021/three-years-tropomi-measurements)

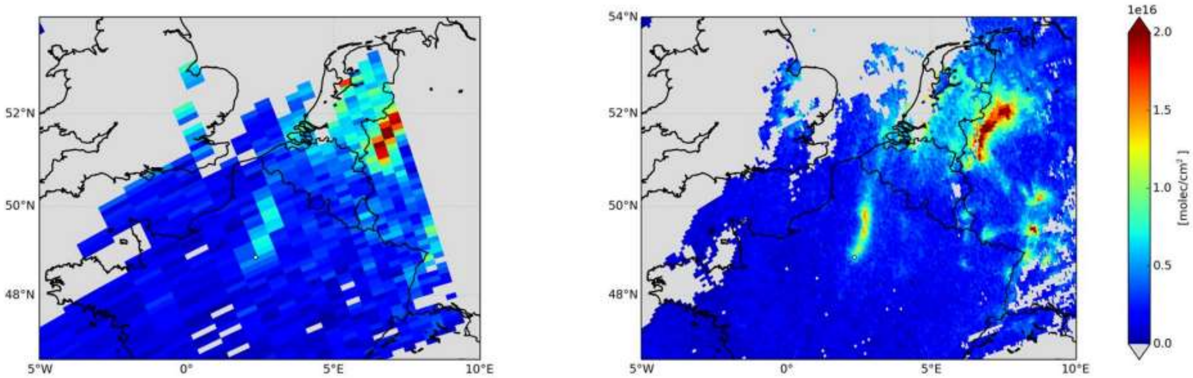


Figure 27 OMI (left) and TROPOMI (right) NO_2 tropospheric column comparison for 17/04/2018 (Taken from Lorente et al., 2019)

The local overpass time is 13:30 LT, almost identical to OMI's. It covers the wavelength bands 270-495 nm and 710-775 nm with spectral resolution 0.55 nm and 2305-2385 nm with spectral resolution 0.25 (from ultraviolet to shortwave infrared) and it has a capacity to measure the atmospheric columns of trace gases such as NO_2, HCHO, SO_2, O3, CO, CH_4 and aerosol and cloud properties (Figure 28). In this thesis, we are interested in NO_2 tropospheric columns.

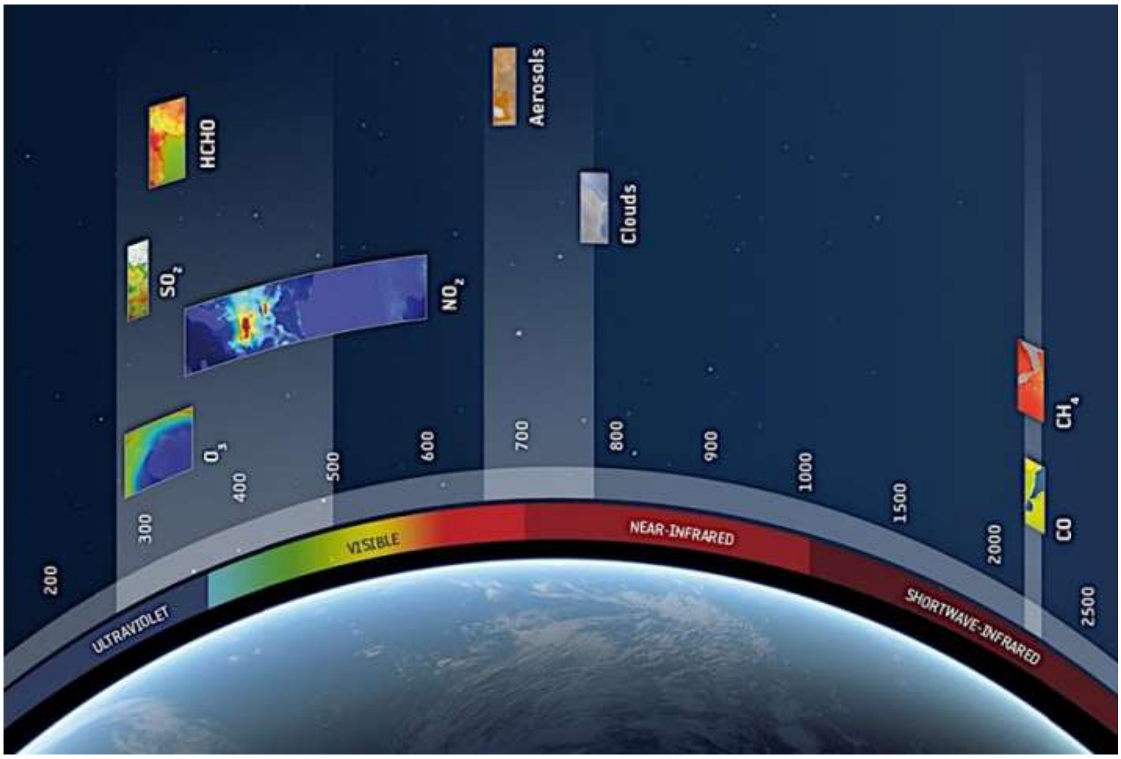


Figure 28 Spectral Coverage of TROPOMI (Image credits: ESA)

Compared to OMI, TROPOMI has 1-5 times higher SNR per-ground pixel (TROPOMI NO_2 data products user manual Van Geffen et al. 2019; Eskes et al. 2019). As with its predecessor OMI, the spectra are analyzed, using the DOAS technique to retrieve SCDs. Figure 29 gives an example of a NO_2 measurement spectral analysis (Borger et al. 2020).

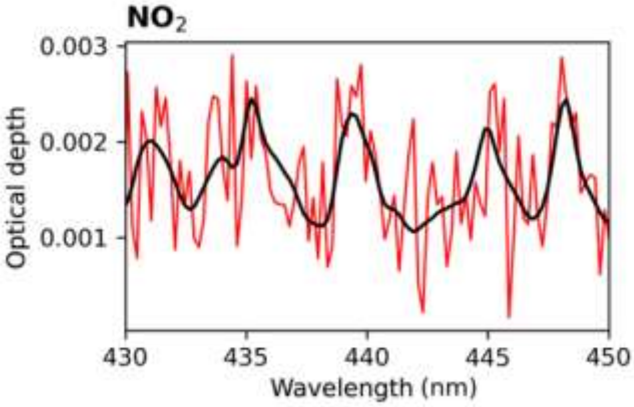


Figure 29 Example of a spectral analysis of a TROPOMI NO_2 measurement spectrum (black line is the best fit while red line is noise) (taken from Borger et al.,2020)

Similar to the OMI, TM5-MP CTM model is used to determine stratospheric and thus tropospheric SCDs. Then, the tropospheric SCDs are converted into a vertical column density using AMF, which is calculated using altitude-dependent AMF look-up tables, radiative transfer model, vertical distribution of NO_2 from an assimilation, cloud fraction, and pressure information from the FRESCO-S algorithm (unlike to OMI) (Lorente et al. 2017; van Geffen et al. 2022). At the beginning of this thesis, the TROPOMI L2 NO2 OFFL (offline) v1.2 and later v1.3 data products were operationally available at the S5P Pre-Operations Data Hub. Since TROPOMI is a new satellite instrument, comparison of TROPOMI L2 NO2 OFFL v1.2 and v1.3 products with ground-based observations shows that underestimation of up to 51% and 37% in extremely and highly polluted areas, respectively (Verhoelst et al. 2021, Table 4). This underestimation may be due to the FRESCO-S cloud pressure algorithm or the use of low-resolution a priori NO_2 profiles from TM5-MP compared to the high resolution of TROPOMI. This problem necessitated an improvement of the data product and since the beginning of 2022; a new product has been made publicly available through the ESA's Sentinel-5P Product Algorithm Laboratory (S5P-PAL). This new PAL v2.2 product shows on average 10-40% higher tropospheric VCDs compared to v1.2 and v1.3 data, depending on pollution level and season (van Geffen et al. 2022). The latest version of the data was reprocessed between 1 May 2018 to 14 November 2021 and is labeled as PAL v2.3 which covers our study period in this thesis (Eskes et al. 2021). In this thesis, in subsection 4.1 and 4.2 we use v1.2 (at the time of the study PAL v2.3 products were not released) and in the rest of Chapter 4 we use PAL v2.3 products of TROPOMI.

2.4.1.2.1 Data selection and processing

The processing of TROPOMI observations is the same as for OMI observations (i) to (iv). To ensure the quality of the retrieval results, the data is distributed with different quality flags corresponding to each ground-pixel and ranging from 0 to 1. In this thesis, we have chosen to use a quality flag value (qa_value) of 0.75, which is a more restricted and also recommended filter. Thus we use pixels with a qa_value greater than 0.75. In addition to that, unlike OMI, data with more than 60% uncertainty are filtered out. This is due to the large amount of TROPOMI observations compared to OMI. We chose to be more restrictive when using TROPOMI data in the inversion to use less informative data. And last, the steps for building super observations are the same as in OMI.

2.4.2 Surface observations

Since true NO_x emissions are not known (and cannot be) in the atmosphere, we used other observational tools currently available to assess our estimated NO_x emissions, constrained by satellite NO_2 observations within the CIF-CHIMERE inversion system. To do that, in this thesis, first we perform CHIMERE simulations using the prior and estimated NO_x emissions (from the inversions) to obtain simulated NO_2 concentrations. Then, the simulated surface NO_2 concentrations are compared to surface observations to evaluate the inversion results in case these new estimate improve the comparisons. NO_2 surface observations are from different in-situ networks according to our study domain.

2.4.2.1 Chinese in-situ network

In this thesis, in-situ observations provided by the China National Environmental Monitoring Centre (CNEMC) network directly affiliated to the Chinese Ministry of Ecology and Environment (MEE) were used to evaluate our estimated NO_x emissions for China (Chapter 3). The CNEMC network established in 1980, incorporates more than 2000 air quality-monitoring stations to evaluate the status of the national air quality by hourly measuring national NO_2, O_3, CO, SO_2, PM2.5 and PM10 since 2014 (Figure 30).

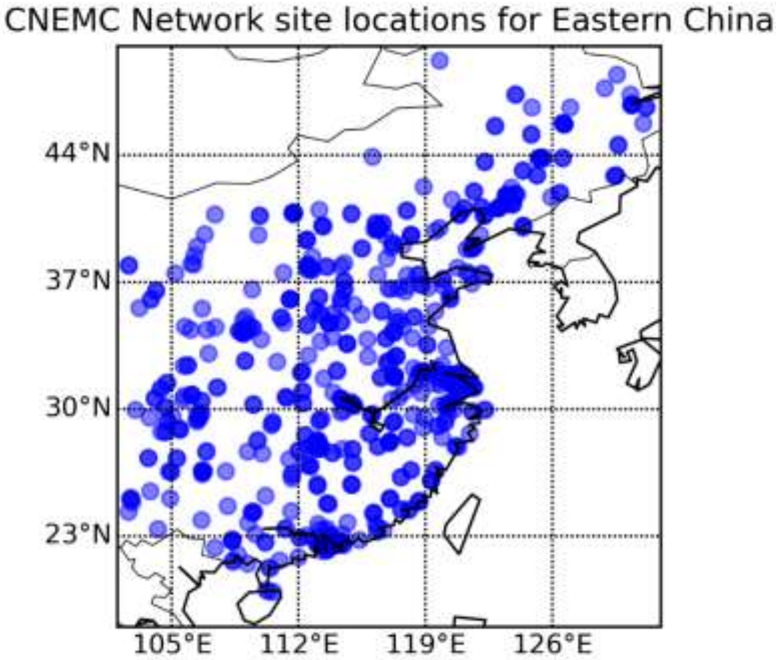


Figure 30 Spatial distribution of the CNEMC Network sites for Eastern China

2.4.2.1.1 Data selection and processing

Networks include different classifications of the stations, depending on their environment types. In this work, the types and environment of the stations follow the classification of Lachatre et al. 2020 based on the Flemming, Stern, and Yamartino 2005 method. Flemming et al. describes a new classification by using hourly concentrations of O_3 and their diurnal variability to evaluate whether stations represent rural, suburban, urban, or traffic environments. In our analyses, we select in-situ data from stations in rural, suburban and urban environments and exclude stations in traffic environments. Because they are close to pollutant sources and cannot be compared to the model with its coarse model resolution. This study considers surface NO_2 concentration measurements from 648 stations in 2015 and 626 stations in 2019, located in Eastern China (also stated in the Figure 30).

These measurements are performed using a commercial chemiluminescence analyzer equipped with molybdenum converter. During the performance of this measurement analyzer catalyzing NO_2 to NO, other reactive oxidized nitrogen compounds (NOz) such as PAN and HNO_3 can also be converted to NO, resulting in an overestimation of ambient NO_2 concentrations (Lamsal et al. 2008). To compensate that, Lamsal et al. suggested using the following correction factor (CF) formula to alleviate and avoid this overestimation of NO_2 (Eq.28).

$$\mathrm{CF} = \frac{\mathrm{NO_2}}{\mathrm{NO_2} + \sum \mathrm{AN} + 0.95\mathrm{PAN} + 0.35\mathrm{HNO_3}} \qquad \text{(Eq.28)}$$

where the sum of all alkyl nitrates (AN), PAN, and HNO_3 at a resolution of $0.5^{\circ} \times 0.5^{\circ}$ are calculated running CHIMEREv2013 simulations with the settings explained in Section 2.3.1.

The values of correction factors are between 0 and 1. In summer, the value is smaller, which means that larger corrections occur, while in winter, it is found to be closer to 1 (Lamsal et al. 2008) and Figure 31. This is because NO_x has a longer lifetime in winter. These correction factors are then bilinearly interpolated by station locations and applied (by multiplication) to the hourly measured surface NO_2 concentrations at these stations.

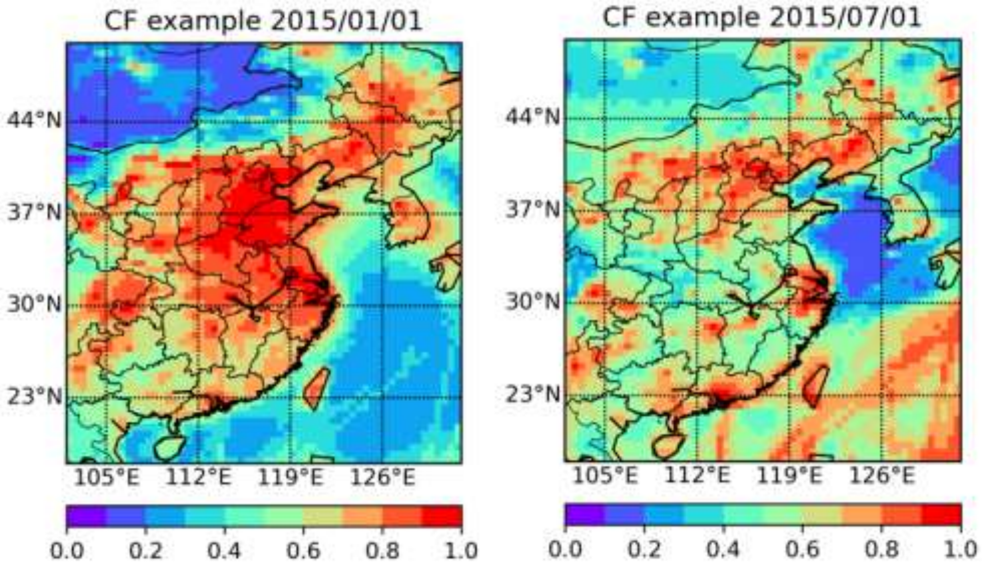


Figure 31 Correction factor examples winter (left) vs summer (right) for 1 day

2.4.2.2 European in-situ network

In-situ NO_2 concentrations from national monitoring networks were retrieved from the European Environmental Agency (EEA, https://www.eea.europe.eu, Last access: 30/03/2023) under the direction of EU. The network provides hourly monitoring data for various pollutants such as NO_2 and O_3 for a large number of stations (more than 3500 stations) since 2013, not only for EU member states but also for non-EU countries located on the European continent (Figure 32). The data provider has predetermined the classification of the stations depending on the predominant emission sources, and distribution of the building around the station. According to these, urban, suburban and rural classification represents the surrounding of the station according to the density of the building from higher to less dense. Traffic stations are located close to a main road (or roads), industrial stations are located close to industrial sources and background stations are located where the general population or vegetation is exposed to average pollution levels. The same correction factor as used for Chinese surface measurements is applied.

Figure 32 Spatial distribution of the EEA sites for Europe (the image is produced in https://www.eea.europa.eu/data-and-maps/explore-interactive-maps/up-to-date-air-quality-data, accessed on 30/03/2023)

2.5 Sensitivity analysis of the CIF inverse modeling system

Since CIF is a new inverse modeling system, we were interested in assessing the sensitivity of some of its parameters, which are used to determine the optimization and convergence of the system and are critical for variational inversion. This allows us to identify which parameters have a higher influence on the outputs of the inversions and to guide our selection parameters more consistently when setting up the inversion system, as in the applications in Chapters 3 and 4. To this end, more than 40 inversion runs for the same day were done over Europe using OMI satellite observations. At the time of the study, the stabilized version of the TROPOMI product (PAL version v2.3.1) was not available. We chose a day in winter where NO_2 are usually higher. The date of 08/02/2020 was chosen because the number of OMI observations was reasonable (amongst the largest for this period) (Figure 33) and the observed NO_2 columns cover a representative range of values. In this section, an overview of these tests is presented. Either a set of parameters, related to the optimization or variational processes, which can influence the inversion results are selected to be tested. A reference test is performed with parameters fixed at values that we estimate to be suitable for a first try based on literature or expert knowledge. The other tests covering a range of values for the different parameters are compared with this reference test by changing one parameter at a time. In the following, these parameters are described for inversion, firstly as optimization parameters (less effective on the solutions), then as **B** error parameters (more effective) and how they are selected is explained.

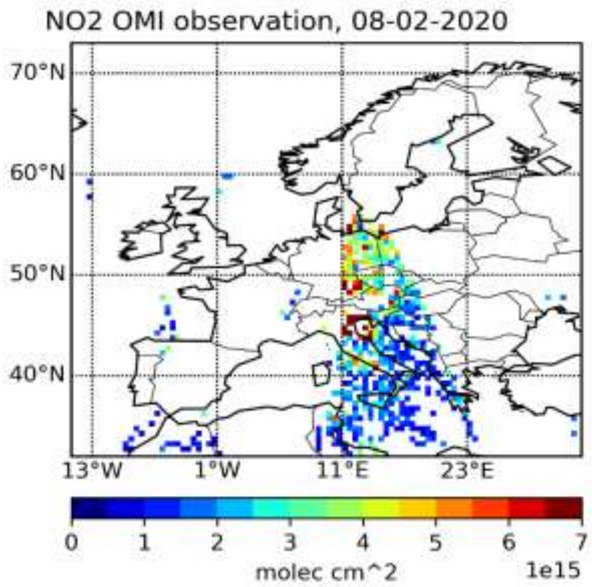


Figure 33 OMI tropospheric column for the reference day (08/02/2020)

2.5.1 Parameters definition

2.5.1.1 Optimization parameters

CIF-CHIMERE uses the M1QN3 limited-memory quasi-Newton minimization algorithm to find the best estimate of NO_x emissions that minimizes the cost function J (Gilbert and Lemaréchal 1989). The parameters related to the performance of the minimization algorithm can be classified under technical parameters. These parameters are as follows:

(i) stopping criteria (so-called 'epsg'). It sets a threshold set on the ratio between the final and initial norms of the cost function gradient. For example, if less than 0.05 is taken as epsg, this equates to a requirement to reduce the final norm by at least 95%.

(ii) maximum number of iterations (so-called maxiter). It sets a limit to the number of iterations allowed in the algorithm. This allows the algorithm stop if epsg is not satisfied.

(iii) Step size (so-called 'df1'). It is the expected decrease in the gradient norm of the cost function during the first iteration and determines the first step size.

Both (i) and (ii) allow the algorithm to stop. The inversion is considered convergent if it reaches a certain amount of reduction in the norm of the gradient cost function before reaching the maximum number of iterations. Otherwise, the system cannot find any minimum (case non-convergent) in which case the criteria must be relaxed (should increase maxiter etc.).

It is important to note that the M1QN3 minimizer uses a line search before each iteration, so that each iteration consists of at least one simulation. This number of simulations (so-called nsim) can also be limited and taken as a criterion for performance. However, it is not considered in our analysis. More information about the M1QN3 minimizer and its parameters can be found in (Gilbert, Lemaréchal, and Simulator 2009) .

2.5.2 Variational parameters

Besides the optimization parameters, the inversions are mainly driven by the error statistics considered in the variational processes: the **B** and **R** matrices and their respective weight (Anderson et al. 2000). The **R** matrix is usually well prescribed by the observation errors provided with the satellite products. Concerning the **B** matrix, attributing the errors and covariances to the prior emissions at the grid-cell size is more complicated. Indeed, this information is usually not provided with the emissions inventories. Uncertainties are mainly estimated for national and annual budgets, given as a range of values, and how they redistribute spatially and temporally is difficult to assess. In practice, the errors values considered for **B** are assigned from a range of values from the inventories, the literature, and the expert knowledge. They may also define the range of “freedom” one gives to the system to reach a reasonable solution. In this thesis work, we tested this range of possible values to guide our configuration choice accounting both the performances of the inversion and the computation time. This means to define range of parameter values one wants to test for the errors in NO_x emissions and their spatial (or horizontal) correlations. These parameters are called variational parameters and are as follows:

(iv) Emission errors (**B** matrix variances).

(iv) Correlation of the emission errors (**B** matrix covariances). To consider these correlations, we did the assumption of Gaussian correlations of the errors and use the spatial correlation parameter defined in CIF as hcorr:sigma, which defines the horizontal correlation of grid cells by the sigma factor. It is formulated as follows;

$c(i,j) = \exp(-dist(i,j)/sigma)$ where c, the correlation between i-th grid cell and j-th grid cell, is a decreasing exponential of the distances over sigma. Thus, an increasing sigma increases the correlation between grid cells. We consider the sigma values as the variable parameter in the tests. Note that, vertical correlation in the uncertainty is ignored.

For sensitivity analysis, the following configuration is then taken as a baseline in this thesis: 70 % anthropogenic emission error, biogenic emission errors are ignored, the spatial correlation is assumed 50 km, the stopping criterion is taken as 0.05 with a maximum of 10 iterations are allowed with the first step size of 0.01.

2.5.2 Sensitivity to optimization parameters

Table 2 summarizes the tests performed by changing the technical parameters on the day of 08/02/2020. Test A is taken as a reference with anthropogenic NO_x emission error 70 %, no error in biogenic NO_x emissions, spatial correlation as 50 km, and technical parameters; maxiter, epsg, df1 are as 10, 0.05 and 0.01, respectively. It can be deduced that the reference Test A converges in 4 iterations with the result that the norm of the gradient of the cost function J decreases by 99 %. According to the M1QN3 user manual, a too small value of df1 causes the algorithm to stop at the first step; however, a large value of df1 is reduced after a few iterations. Therefore, we tested for larger values at df1 (e.g., 0.1, 0.2 and 0.3), but did not obtain significant changes (from tests B to D) in terms of emission corrections. It almost only increases the computation duration. Next, we tested to go further in the minimization algorithm by stressing the stopping criterion. The epsg parameter is then decreased to 0.01 in Test E and to 0.001 in Test F. The maximum number of iterations is increased to avoid the system stops before reaching the epsg criterion. Reference Test A and Test E give the same results as the reference test stops with a reduction of the gradient norm close to 99 %. Test F goes further in the reduction of the gradient norm and reaches the convergence but without any significant impact on the differences of emissions between the prior and the posterior (Figure 34) and increasing the computation time (Table 2). Due to the limited impact of the optimization parameters on the inversion results but their implication in terms of computation time, these parameters are taken as in reference Test A for the next inversion tests.

Table 2 Inversion tests with changed optimization parameters for the day of 08-02-2020.

				Optimization Parameters				
Name of the test	Anthropogenic emission error (%)	Spatial correlation (km)	Biogenic emission error (%)	**Max iteration**	**Stop: epsg**	**Step size: df1**	Duration (in hour)	Convergent? (Yes / No)
A	70	50	-	10	0.05	0.01	3	Y
B	70	50	-	10	0.05	0.1	2.5	Y
C	70	50	-	10	0.05	0.2	4	Y
D	70	50	-	10	0.05	0.3	10	Y
E	70	50	-	20	0.01	0.01	3	Y
F	70	50	-	20	0.001	0.01	5	Y

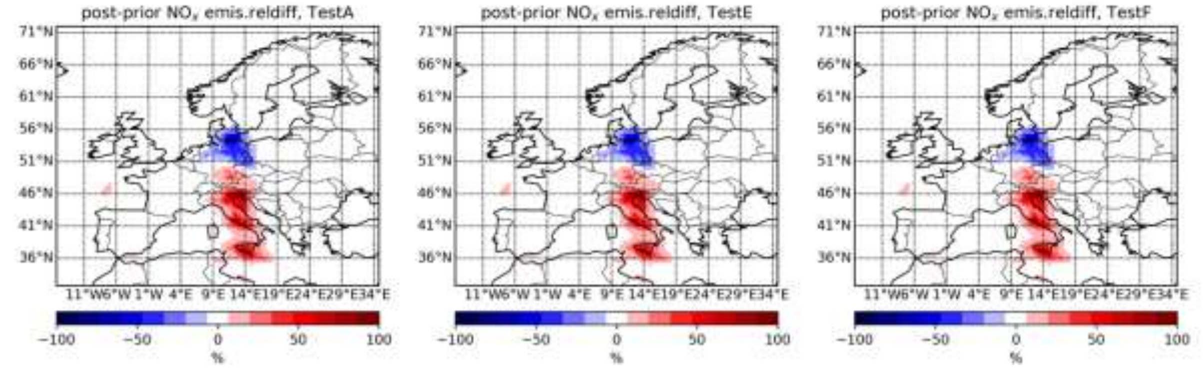


Figure 34 NO_x emission corrections obtained by Test A (left), Test E (middle), Test F (right). Note that, the corrections are the same.

2.5.3 Sensitivity to variational parameters

In this section, we tested different parameters related to the **B** error covariance matrix, which is a key parameter in variational inversion (Sections 2.1 & 2.2), with the aim of building **B** matrix and analyzing how the CIF-CHIMERE inversion system reacts to different estimations of these values. To do so, we need to know the uncertainties and correlations on emissions, but we have low knowledge about these uncertainties as discussed previously. For this reason, we perform different tests over a range of uncertainties.

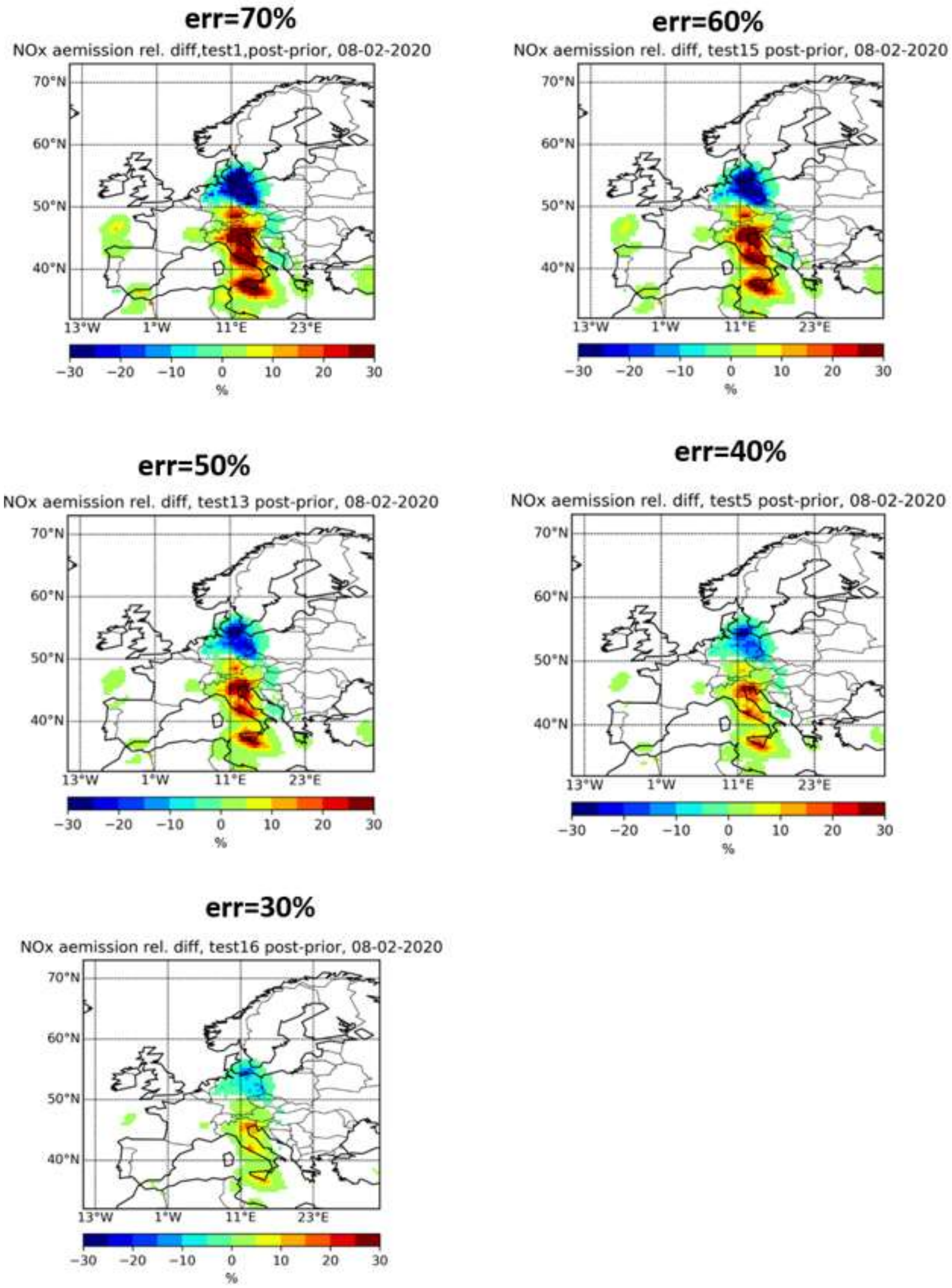
2.5.3.1 **B** variances

Variances of **B** are the diagonal terms of the matrix representing the different anthropogenic emission errors. In Europe, the uncertainty of national anthropogenic NO_x emission estimates stays at 20-30 % (Citepa 2021), but uncertainty values are more likely to be larger on a smaller scale (grid-cell scale). In literature, Müller and Stavrakou 2005 reported 50 % of uncertainties in their inversions for Europe. Souri et al. 2020 assumed these uncertainties to be 50 % for their inversions in Asia. Therefore, to evaluate the impact of the choice of the values in correcting NO_x emissions without pushing the limits of computational resources, we considered **B** variances for our tests between 30 % and 100 %. Table 3 summarizes the performed tests. It is important to note that these errors indicate how much confidence we have in the prior emissions and which freedom we allow the system to have. A higher error value gives the inversion system more freedom to correct emissions. Hereunder, test I, which has the largest error value we tested, is not convergent with the given parameters.

Table 3 Inversions tests with different errors in anthropogenic NO_x emissions for the day of 08-02-2020.

	Variational Parameters							
Name of the test	**Anthropogenic emission error (%)**	Spatial correlation (km)	Biogenic emission error (%)	Max iteration	Stop: epsg	Step size: df1	Duration (in hour)	Convergent? (Yes / No)
A	70	50	-	10	0.05	0.01	3	Y
I	100	50	-	10	0.05	0.01	12	N
J	60	50	-	10	0.05	0.01	2.5	Y
K	50	50	-	10	0.05	0.01	2	Y
L	40	50	-	10	0.05	0.01	2	Y
M	30	50	-	10	0.05	0.01	2	Y

Tests J to M have an emission error that decreases by 10%, respectively. They are all convergent. Tests with smaller errors require shorter computation time to converge compared to the reference test A. Figure 35 shows the spatial distribution of the emission corrections obtained from these tests. It is seen that the range of corrections decreases with respect to the error value. More precisely, these corrections range from ±90 %, ±70 %, ±50 %, ±30 % and ±20 %, corresponding to the errors of 70 %, 60 %, 50 %, 40 % and 30 %, respectively. While tests L and M are too restrictive, test K provides sufficient freedom when the error parameter **B** is 50 % and the corrections are within the 50 % range. This means that, the inversion system is well responding to the error changes in the anthropogenic NO_x emissions.


err=70%

err=60%

err=50%

err=40%

err=30%

Figure 35 Increments according to the different errors in prior NO_x emissions (From left top to bottom left, test A and tests J to M, respectively) (differences<1% is neglected).

2.5.3.2 **B** covariances

Defining the covariances or non-diagonal terms of **B** matrix is even more difficult to evaluate. First NO_x inversion studies neglected to spatial correlations of emissions errors making the assumption that at medium/coarse resolution the transport of NO_2 between neighbouring grid cells can be neglected (e.g., Martin 2003). Konovalov et al. 2006 discussed this assumption with a CHIMERE at 0.5° resolution and concluded that considering two layers of neighbouring grid cell were sufficient to represent the NO_2 transport. Based on this, we define the range of variation of the sigma parameter used to prescribe the horizontal correlation in our case as an exponentially decaying function as mentioned previously. We tested different correlation lengths based on the CHIMERE resolution (50 km) from 0 to 100 km. Table 4 summarizes these tests for different spatial error correlations of emissions.

Table 4 . Inversions tests with different spatial correlations for the day of 08-02-2020

	Variational Parameters							
Name of the test	Anthropogenic emission error (%)	**Spatial correlation (km)**	Biogenic emission error (%)	Max iteration	Stop: epsg	Step size: df1	Duration (in hour)	Convergent? (Yes/No)
A	70	50	-	10	0.05	0.01	3	Y
G	70	100	-	10	0.05	0.01	2.5	Y
H	70	-	-	10	0.05	0.01	2.5	Y

Figure 36 shows the difference between estimated anthropogenic NO_x emissions and prior emissions in %, from left to right, without considering any correlation, a correlation taken as 50 km (same as the CHIMERE model resolution) and 100km (correspond to the tests in the Table 4).

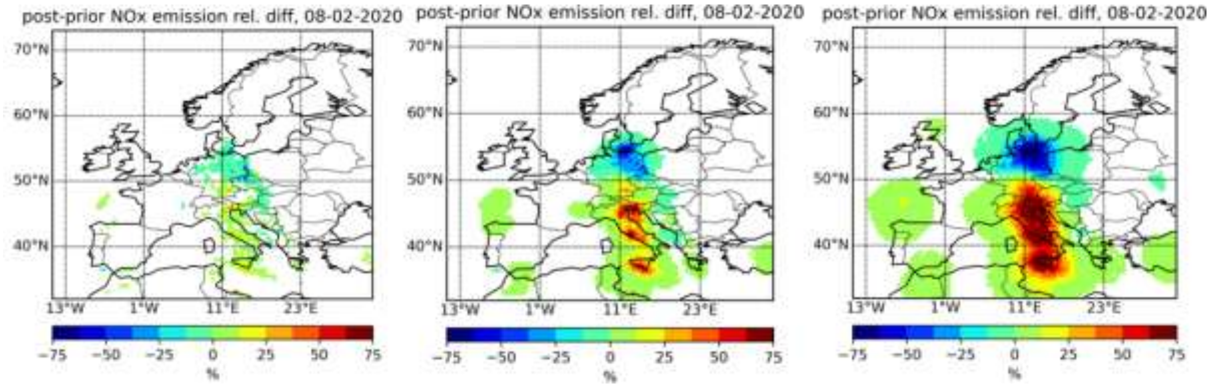


Figure 36 Anthropogenic NO_x emissions increments (in %) provided by inversions test from left to right; with no correlation (Test H), sigma=50 (Test A, reference test) and sigma=100 (Test G), respectively (differences<1% is neglected).

According to these tests, neglecting correlation in correction of the emissions (leftmost) leads to corrections only localized compared to the tests where correlations is considered (middle and rightmost). According to the study of Konovalov et al. 2006, this assumption is likely too strict for the considered CHIMERE resolution. Accounting for a spatial correlation with a Gauss function with the half width (sigma) equal to the CHIMERE resolution fills the recommendation of Konovalov et al. 2006 study of considering two layers of neighboring grid cells. Test G, with sigma equal to 100 km gives too diffusive corrections. Another “advantage” of considering horizontal correlation is somehow to counterbalance the lack of observations such it can be the case with OMI. Therefore, it seems reasonable to take spatial correlations with the correlation length derived from the model resolution as in the reference test A (middle).

2.5.3.3 Biogenic emission error

So far, we have performed inversion tests without taking into account the error in biogenic emissions. In winter months NO_x emission is dominated by anthropogenic activities but in summer months biogenic emissions (biomass burning, lightning and soil) are also a substantial source (Vinken et al. 2014). Like anthropogenic emissions, estimates of biogenic emissions are subject to uncertainties; e.g., lack of good quality datasets due to highly variable agricultural activities such as fertilization of soils, vegetation types and soil management practices (Skiba et al. 2021). These estimates are often larger than these for anthropogenic emissions are. Therefore, we consider to errors between 100 % and 200 % in biogenic emissions in our tests and we include a summer day where the impact could be more important (Table 5).

Table 5 Inversions tests with different errors in biogenic NO_x emissions for two days 08-02-2020 and 28-08-2019

		Variational Parameters				
Name of the test	**Day**	Anthropogenic emission error (%)	Spatial correlation (km)	**Biogenic emission error (%)**	Duration (in hour)	Convergent ? (Yes/No)
N	08/02/2020	50	50	100	2	Y
O	08/02/2020	50	50	200	2.5	Y
P	28/08/2019	50	50	100	2	Y

Note that the anthropogenic emission error is assigned to 50 % as discussed in 2.5.3.1, 2005 for Europe. Figure 37 shows the biogenic NO_x emission differences from tests N and O (respectively). According to the figure, we can conclude that, doubling the biogenic emissions errors results in only minor changes in the correction of the emissions – the corrections are smaller than 1 % in winter. On the other hand, on a summer day (Figure 38), when taking the same error (100 %) in biogenic emissions, the correction of the emissions is ten times more significant compared to a winter day but remains small with ±10 %. Therefore taking 100 % on error in biogenic emissions would be sufficient for our next inversion configuration.

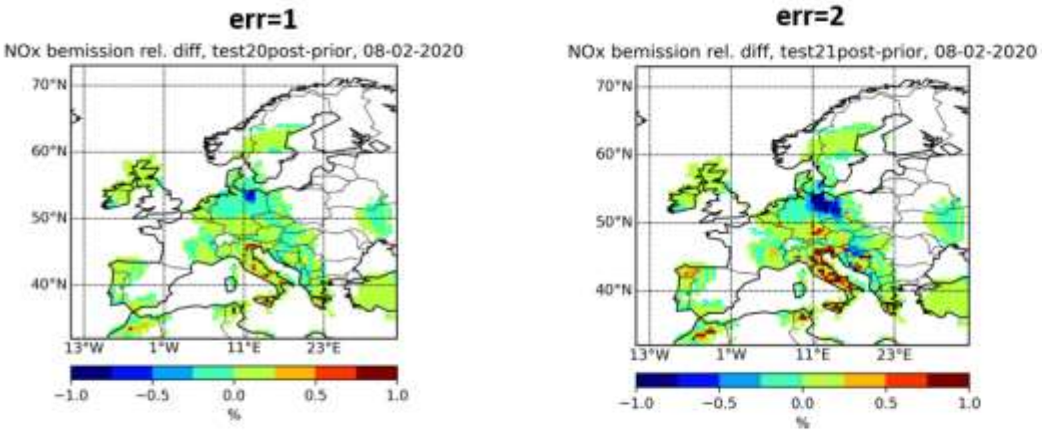


Figure 37 Increments according to the 100% (Test N, left) and 200% (Test O, right) errors in biogenic NO_x emissions (differences<1e-5% is neglected).

As a conclusion, these are preliminary tests to evaluate how the state-of-the-art CIF-CHIMERE inversion system responds to a number of changing parameters and to prepare for the inversions for Europe and China in the following chapters. These tests show that our choices based on expert knowledge are reasonable and consistent with the literature. For a comprehensive sensitivity study, more tests should be conducted, as one day is not representative. e.g., for seasonality.

The final configuration we choose is 50 % anthropogenic emission errors with 50 km spatial correlation and 100 % biogenic emission errors for Europe. For China, NO_x emissions uncertainties reported in literature are slightly larger than in Europe, we then decide to increase the **B** variances to 70%.

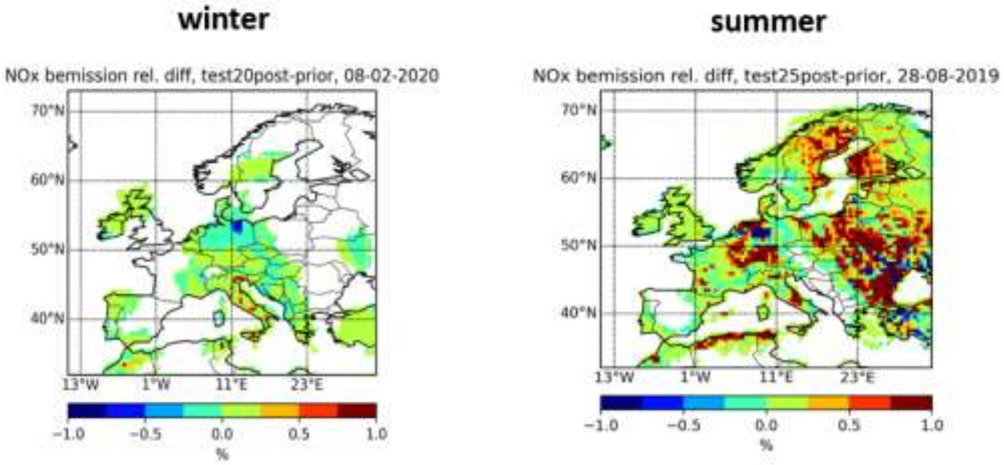


Figure 38 Increments using 100% biogenic error emissions in a winter day (Test N, left) and in a summer day (Test P, right) (differences<1e-5% is neglected).

Third part : Scientific Applications

Chapter 3. Inverse modeling of NO_x emissions over China

This chapter is an adapted and expanded version of the scientific paper with its supplementary elements published in MDPI's journal Atmosphere under the title "Anthropogenic NO_x Emission Estimates over East China for 2015 and 2019 Using OMI Satellite Observations and the New Inverse Modeling System CIF-CHIMERE" and can be found as Savas et al.,2023 on https://doi.org/10.3390/atmos14010154 (Last Access:08/04/2023). We acknowledge the open access license policy of the publisher MDPI.

Summary

China is the world's largest emitter of air pollutants (Lin et al. 2014). Over the last few decades, NO_x emission levels in China have further increased due to growth in industrial activities, urbanization and weak pollution regulations (Jin, Andersson, and Zhang 2016; Shaw and Van Heyst 2022). However, in March 2011, the Chinese government for the first time introduced regulations to reduce the level of NO_x pollutants and control these emissions, called the 12th Five-Year Plan (2011-2015). The target of this regulation was to reduce the annual Chinese NO_x emissions nationwide by 10 % by 2015 compared to the levels of 2010 (de Foy, Lu, and Streets 2016). Studies have shown that these regulations can reduce NO_x emissions and the Chinese government has achieved its emission reduction target by reducing total NO_x emissions by 18.6 % (Liu et al. 2016, 13th Five-Year Plan). Since then, the changes in NO_x emissions have been assessed using various approaches to evaluate the impact of the regulations (e.g., (Zheng et al. 2018; Xi Lu et al. 2020). Complementary to the previous studies, this study estimates anthropogenic NO_x emissions in 2015 and 2019 over Eastern China using as a reference the Hemispheric Transport of Air Pollution (HTAP) v2.2 emission inventory for 2010 and the new variational inversion system the Community Inversion Framework (CIF) interfaced with the CHIMERE regional chemistry transport model and OMI satellite observations. We also compared the estimated NO_x emissions with the independent Multi-Resolution Emission

Inventory for China (MEIC) v1.3, from 2015. The inversions show a slight global decrease in NO_x emissions (in 2015 and 2019 compared to 2010), mainly limited to the most urbanized and industrialized locations. In the locations such as Baotou, Pearl River Delta, and Wuhan, the estimations in 2015 compared to 2010 are consistent with the target reduction (10 %) of the 12th Five-Year Plan. Comparisons between our emission estimates and MEIC emissions in 2015 suggest that our estimates likely underestimate the emission reductions between 2010 and 2015 in the most polluted locations of Eastern China. However, our estimates suggest that the MEIC inventory overestimates emissions in regions where MEIC indicates an increase of the emissions compared to 2010.

3.1 Introduction

Nitrogen oxides (NO_x=NO_2+NO) are a family of highly reactive trace gases that cause air quality deterioration in the troposphere. They play a key role in the formation of ozone (O_3) through photochemistry and the formation of secondary organic aerosols (Seinfeld and Pandis 2016; Ng et al. 2007). Most tropospheric NO_x are emitted as NO and rapidly converted to NO_2 by various reactions (Atkinson 2000). Anthropogenic activities such as fossil fuel combustion are dominant NO_x emission sources, especially in industrialized regions (Houghton et al. 2001). On the other hand, natural and biogenic sources can be a major source of NO_x, such as soil NO_x emissions in agricultural areas (Almaraz et al. 2018).

Over the past few decades, NO_x emission levels in China increased due to growth in industrial activities and weak pollution regulations (Claire Granier et al. 2011; Shaw and Van Heyst 2022; Jin, Andersson, and Zhang 2016). The control of NO_x emissions was added to the regulations in March 2011 with the 12th Five-Year Plan (2011–2015) to reduce the annual Chinese NO_x emissions nationwide by 10 % compared to 2010 (de Foy, Lu, and Streets 2016). Studies have shown that the Chinese Government met its emission reduction target due to these regulations in this period and even exceeded the plan's target e.g., (CAA (Clean Air Asia) 2016; Anger et al. 2016). This plan is followed by a 13th Five-Year Plan (2016–2020) with a target of reducing Chinese NO_x emissions nationwide by 15 % in 2020 compared to 2015 ("Notice of the State Council on Printing and Distributing the '13th Five-Year' Ecological and Environmental Protection Plan, Government Information Disclosure Column" 2016).

Chemistry transport models (CTMs) have been widely used to study the efficiency of environmental regulations, assess their impacts, and make future predictions through different scenarios (Anger et al. 2016; Colette et al. 2012; Zhao, Zhang, and Nielsen 2014; Permadi, Kim Oanh, and Vautard 2018; Nellie Elguindi et al. 2020). Emission inventories are one of the main inputs of CTMs, and their spatial and temporal accuracies are essential for simulations and forecasts. One common way to build an emission inventory is the bottom-up approach. This approach quantifies emissions based on the local distribution of activities, emission factors, and annual self-declarations by emitters. Numerous studies estimated China's NO_x emissions using these activity rates and species' emission factors (Xia, Zhao, and Nielsen 2016; Shi et al. 2014). For example, Li, Zhang, et al. 2017 calculated the annual anthropogenic Chinese NO_x emissions for 2008 as 26.55 Tg using the Multiresolution Emission Inventory for China (MEIC) v1.0 and 20.66 Tg using Hemispheric Transport of Air Pollution (HTAP) v2.2 bottom-up emission inventories. For 2010, Li, Zhang, et al. 2017 and Zheng et al. 2018 estimated annual NO_x emissions as 26.5 Tg and 29.1 Tg, respectively, using two different versions of MEIC bottom-up inventory (v1.3 and v1.0, respectively). However, even though the information comes firsthand, significant uncertainties remain. Using Monte Carlo simulations, Zhao, Zhang, and Nielsen 2013 predicted Chinese annual NO_x emissions' uncertainties to be 35 % of estimated emissions in 2010. Crippa et al. 2018 estimated NO_x emission uncertainties in 2012 as 56.2 % for top emitting regions in China. It is worth noting that the uncertainties are given on an annual and national basis. They might be larger when considering grid points or smaller timescales. Moreover, the collection of data to build bottom-up inventories is time-consuming; then, updating these inventories may take several years (Ding et al. 2017).

On the other hand, using inverse modeling techniques, which constrain emissions by atmospheric observations, brings complementary information and, in some regions such as China, captures current emission trends better than traditional bottom-up inventories (Nellie Elguindi et al. 2020). In addition, unlike bottom-up inventories, these top-down inventories provide broad spatial coverage and have a fast update ability, as emissions can be estimated when observations are ready to interpret. Nevertheless, top-down inventories have some limitations and uncertainties related to the performances of CTMs on transport and chemical processes (de Foy et al. 2014), and the observations themselves (e.g. limited sensitivity to the surface and cloud coverage for satellite observations) (Boersma, Vinken, and Eskes 2016). Thus, different inversion techniques have been developed to estimate NO_x emissions. For example, earlier studies used a mass balance approach to constrain NO_x emissions on various spatial scales (Leue et al. 2001; Cooper et al. 2017; Han, Kim, and Song 2020; Itahashi et al. 2019; Martin 2003). Most of the studies assume a linear relationship between NO_x prior emissions and NO_2 columns, and do not consider horizontal emission transport. With the increase in computational capacity, advanced inversion techniques such as the Kalman filter (Napelenok et al. 2008) and variational data assimilations (Kurokawa et al. 2009) have been developed to take into account the transport of emissions and nonlinearities in chemical reactions. Ronald J. van der A et al. 2017 derived anthropogenic NO_x emissions on a 25 × 25km^2 provincial level using the DECSO v4 algorithm (Mijling and van der A 2012; Ding et al. 2015; Ding et al. 2017), which estimates emissions using Kalman filter, the regional CHIMERE v2013 CTM (Menut et al. 2013), MEIC v1.0 inventory for 2010 (Zheng et al. 2018; Li, Liu, et al. 2017) and Ozone Monitoring Instrument (OMI) NO_2 observations using DOMINO v.2 algorithm (Boersma et al. 2011) for the period 2005–2015 in Eastern China. They found an increasing trend in total NO_x emissions until 2012, except for 2009. Even though the total NO_x emission peaks occur in 2012, different peaks occur between 2011 and 2014 on a provincial level. Qu et al. 2022 co-assimilated anthropogenic and biomass burning NO_x, SO_2, and CO emissions from bottom-up inventory HTAP v2.2 2010 Janssens-Maenhout et al. 2015 in China using a sector-based variational algorithm and GEOS-Chem Adjoint model (Henze, Hakami, and Seinfeld 2007) for January 2005–2012. Unlike Ronald J. van der A et al. 2017, their inversion captures the NO_x emissions peak for 2011. Moreover, they found that all estimated NO_x emission top-down inventories (sector-based & species based) are smaller than HTAP v2.2 bottom-up inventory by 21–26 % in January 2010. Zheng et al. 2018 quantified anthropogenic NO_x emissions using a combination of MEIC v1.3 bottom-up inventory and index decomposition analysis. They compared their bottom-up emission estimates with other studies, which include OMI observations and top-down emission estimates between 2010 and 2015 for China and Eastern China. According to their study, all top-down NO_x emission inventories calculated by inverse modeling using OMI NO_2 columns show smaller relative differences between these two years than bottom-up estimations and OMI observations (Zheng et al. 2018)(p.14106, Table 2).

The previous studies show a significant range of uncertainties regarding the NO_x emissions from bottom-up and top-down approaches and a significant spread in the emissions derived from different approaches (Ding et al. 2017). The changes in the emissions such as the emission reduction targets of governmental mitigation strategies, which the inventories need to assess, are usually smaller than the uncertainties on the emissions and then an important challenge for the inventories. A variety of tools to estimate emissions is essential to assess the emissions and their related uncertainties better. Our study proposes the use of a new inverse modeling system, the Community Inversion Framework (CIF) (Berchet et al. 2021), interfaced with the CHIMERE chemistry-transport model (Menut et al. 2013) and its adjoint (Fortems-Cheiney et al. 2021) and OMI satellite observations to estimate Chinese NO_x emissions in 2015 and 2019. The objectives of this study are twofold: (i) testing the new CIF-CHIMERE over China and evaluating if the derived emissions and concentrations simulated using them are closer to independent inventory and measured concentrations, respectively; (ii) evaluating if the NO_x emission regulations planned in the two last Five-Year Plans are visible in the emissions derived with the new CIF-CHIMERE system. We choose to focus on the last year (2015) of the 12th Five-Year Plan. We do not consider the last year of the 13th Five-Year Plan, 2020 as it was perturbed by the COVID-19 pandemic. We analyze the NO_x emission estimates at the scale of Eastern China, at the scale of the most urbanized, and industrialized Chinese locations. We evaluate the estimated NO_x emissions by comparing with the independent inventory MEIC v1.3 and surface NO_2 observations from the Chinese air quality network. Section 3.2 describes the methodology and tools used in this study. Results and discussions are presented in Section 3.3.

3.2 Materials and methods

3.2.1 Inversion-related tools and data

3.2.1.1 Inverse modeling and the CIF system

The atmospheric inverse modeling approach applied to emission estimates aims at constraining a prior knowledge of emissions with atmospheric observations (concentrations or columns). A chemical transport model combined with an observation operator is used to map the emissions into the observation space. The emissions are usually constrained at the model grid point and at a temporal resolution better than annual. Due to the high dimension of the problem, sophisticated approaches (Bocquet et al. 2015) are necessary. The evaluation of the estimated or posterior emissions can only be performed indirectly by comparing concentrations simulated using these posterior emissions (named Conc_Post in the following) to independent measured concentrations. This step is often named “experiment run”. To evaluate if the inversion succeeds to better assess the emissions compared to our prior knowledge used, it is also necessary to perform a “control run” to simulate the concentrations using the prior emissions (named Conc_Prior in the following) and quantify if the Conc_Post are closer to the measured concentrations than the Conc_Prior. Figure 39 summarizes the different steps needed to perform atmospheric inversions and to evaluate them.

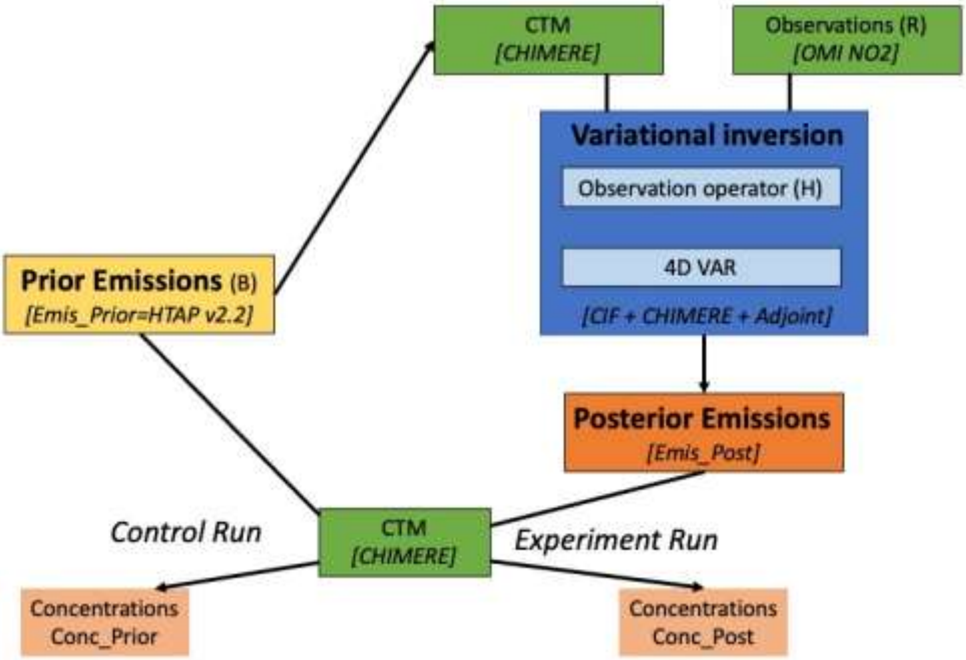


Figure 39 Schematic view of the atmospheric inversion and evaluation.

As the dimension of the problem to solve is high, we use a variational approach based on a Bayesian framework and normal distributions of the errors and uncertainties. The system aims to find an optimal solution of the control vector x by minimizing the cost function J (Equation 29).

$$J(x) = \frac{1}{2}(x - x_b)^T B^{-1}(x - x_b) + \frac{1}{2}(H(x) - y_o)^T R^{-1}(H(x) - y_o) \quad \text{(Eq.29)}$$

where x_b is a prior estimate of the control vector, B is the prior error covariance matrix, $H(x)$ is the observation operator matrix which links control variables to atmospheric measurements at specified locations and time using a CTM (CHIMERE in this study), y_o is the observation vector (OMI NO_2 columns in this study), and R is the covariance matrix of the observation errors. In our case, the control variables are not emissions, but rather daily multiplicative factors applied to NO_x emissions. During a given day, the corresponding multiplicative factor is used to scale homogeneously the hourly emission values, retaining the relative variations of the diurnal profile. This relationship between control variables and emissions can be stated as follows:

$E_{NO_x}(\text{time, altitude, latitude, longitude}) = x(\text{day, altitude, latitude, longitude}) \times$
$E_{NO_x,ref}(\text{time, altitude, latitude, longitude})$

with $E_{NO_x,ref}$ the reference emissions from the bottom-up inventory. Thus, the values of x_b are all set to 1 and the control covariance matrix B is expressed in relative terms and not in emission covariances. The dimension of x and x_b is $24 \times 16 \times 64 \times 60$ and corresponds to the hours in a day, the level of altitude and the number of latitudinal and longitudinal grid cells in the domain, respectively. The dimension of y is 1D (e.g., 780) and determined by the selection of the observations (selection criteria explained in the OMI observation section).

We use the CIF inversion system interfaced with the CHIMERE CTM and its adjoint and the OMI NO_2 columns to perform the inversion experiments. The CIF is an open-source unified system designed to allow the use of different atmospheric inversion approaches, atmospheric transport models, and observations with large flexibility (Berchet et al. 2021). It is designed to estimate the fluxes of various greenhouse gases and reactive species both at the global and regional scales. As we consider reactive species, the inversion problem is non-linear. We then use the M1QN3 limited-memory quasi-Newton minimization algorithm (Gilbert, Lemaréchal, and Simulator 2009) to minimize the cost function J (Equation 29). It is worth noting that computing the uncertainties on the posterior emissions is very challenging when using variational approaches due to the dimensionality of the problem and constraints on time and computing sources (Berchet et al. 2021; Fortems-Cheiney et al. 2021; Rayner, Michalak, and Chevallier 2019). These uncertainties are then not estimated here.

3.2.1.2 HTAP v2.2 emission inventory

The Hemispheric Transport Atmospheric Pollution (HTAP) inventory, version 2.2 (Janssens-Maenhout et al. 2015) is used to prescribe anthropogenic emissions of pollutants to the CHIMERE model. For the country of China, HTAP v2.2 is constructed using regional emission inventories in collaboration with the MICS Asia Scientific Community (MICS-Asia) and REAS experts with a horizontal resolution of $0.25^\circ \times 0.25^\circ$ converted to $0.1^\circ \times 0.1^\circ$ (Galmarini et al. 2017). Emissions are estimated from seven main sectors of anthropogenic activities: air, ships, energy, industry, transport, residential, and agriculture. The HTAP v2.2 emissions correspond to the year 2010 and used as the prior in the inversion runs (Figure 39), whatever the considered year. In this study, emissions will be called Prior when they are used as a prior in the inversion. In parallel, they are also used as a reference for the emissions of the year 2010, the starting point to evaluate the 12th and the 13th Five-Year Plans. In this study, they are called HTAP v2.2 when they are used as a reference in the simulations. In addition, the biogenic NO_x emissions are given by the Model of Emissions of Gases and Aerosols from Nature version 2.1 (MEGAN v2.1) model (Guenther et al. 2012).

In this study, we set prior uncertainties B by taking diagonal terms as 70 % for the anthropogenic NO_x emissions (Müller and Stavrakou 2005) and 100 % for the biogenic NO_x emissions. We use horizontal correlations that decay exponentially with the correlation length scale derived from the model (CHIMERE) resolution. Noting that, in this study, NO_x emissions are expressed as nitrogen oxides (NO_2 and NO) in kilotons (hereinafter, kt NO_x or Tg NO_x is referred to as kt or Tg).

3.2.1.3 OMI satellite observations

The Ozone Monitoring Instrument (OMI) is a nadir-viewing ultraviolet-visible spectrometer on National Aeronautics and Space Administration’s (NASA) Earth Observing System (EOS) Aura polar satellite, launched in July 2004 (Schoeberl et al. 2006). The instrument is a result of the contribution of the Netherlands’ Institute for Air and Space Development and the Finnish Meteorological Institute to the EOS Aura mission. It has been providing daily data since 9th August 2004 by making full global measurements of the solar radiation backscattered from the Earth’s atmosphere and surface, along with measurements of the solar irradiance with a spatial resolution of 24×13 km^2 at nadir (across × along-track) with a local overpass time 13:40 LT (Levelt et al. 2006).

This study uses OMI NO_2 tropospheric columns from the European Quality Assurance for Essential Climate Variables (QA4ECV) project’s NO_2 ECV precursor version 1.1 product (Boersma et al. 2018). This QA4ECV retrieval consists of an improved spectral fitting method and data assimilation scheme using the TM5-MP model in $1^{\circ} \times 1^{\circ}$ degree horizontal resolution to estimate stratospheric background NO_2 and to reinforce the air mass factor (Zara et al. 2018; Lorente et al. 2017). OMI tropospheric NO_2 column data is defined as the vertical integrations of the NO_2 molecules between the surface and the tropopause. The data set is collected from the Tropospheric Emission Monitoring Internet Service (TEMIS https://www.temis.nl/airpollution/no2.php Last access: 26/09/2022). We first select OMI observation data of which ; (i) the processing error flag equals 0 for a pixel, (ii) the solar zenith angle is lower than 80°, (iii) the snow ice flag is lower than 10 or equal to 255, (iv) the ratio of tropospheric air mass factor (AMF) over geometric AMF is higher than 0.2 to avoid situations in which the retrieval is based on very low (relative) tropospheric AMFs, and (v) the cloud fraction is lower than 0.5. Second, we filter out these selected daily OMI observation data with a maximum of 100 % error.

3.2.1.4 The regional CHIMERE CTM

CHIMERE is a state-of-the-art three-dimensional regional CTM that simulates hourly atmospheric concentrations of a number of pollutants of interest over a given domain (CHIMERE https://www.lmd.polytechnique.fr/chimere/ Last access:26/09/2022). External forcings are required to run a simulation, such as meteorological fields, primary pollutant emissions (anthropogenic and biogenic emissions), and initial and boundary conditions.

In this study, we use CHIMERE v2013 (Menut et al. 2013) with the following inputs to simulate NO_2 concentrations. Boundary conditions are given by a climatology from the global model LMDZ-INCA (Hauglustaine et al. 2004). Meteorological conditions come from the European Centre for Medium- Range Weather Forecasts (ECMWF) and are interpolated hourly. Biogenic emissions are estimated from the MEGAN v2.1 model (Guenther et al. 2012). Finally, anthropogenic emissions are those from the HTAP v2.2 bottom-up inventory (Janssens-Maenhout et al. 2015).

Our study configuration is based on a $0.5^{\circ} \times 0.5^{\circ}$ horizontal resolution with 17 vertical layers from surface to 200 hPa and 8 layers within the first three kilometers. It includes 65 (longitude) $\times$ 61 (latitude) grid cells that correspond to $18^{\circ}\mathrm{N} - 50^{\circ}\mathrm{N};\ 102^{\circ}\mathrm{E} - 132^{\circ}\mathrm{E}$ and covers eastern and southern China, seas and Korean Peninsula (Figure 40).

The gaseous chemical scheme used in CHIMERE is MELCHIOR-2 (Guenther et al. 2012; Menut et al. 2013). The aerosol module of CHIMERE is not considered in the simulations and inversions, as the adjoint of this module is not available (Fortems-Cheiney et al. 2021).

For the comparison between the observations and the model during the inversion process (Figure 39), super-observations at the model resolution ($0.5^{\circ} \times 0.5^{\circ}$) are considered. As the number of OMI observations lying in the model grid-cell is generally small and will not allow one to decrease observation errors by averaging, we choose to use the median of corresponding

OMI NO_2 columns. The averaging kernel A and the observation errors associated with the median columns are considered too. The observation errors are used to fill the R matrix.

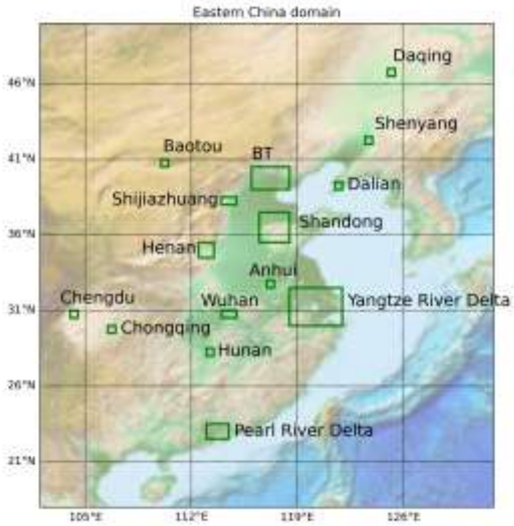


Figure 40 Study domain. Selected locations presenting the most urbanized and industrialized regions (see section 3.3.3) are indicated by the rectangles.

To compare with OMI super-observations, simulated NO_2 concentrations in the same grid cells as the super observations are selected. The hourly-simulated concentrations are interpolated in time and OMI's pressure levels. OMI averaging kernels are applied to these NO_2 vertical profiles using Equation 30 (Boersma et al. 2017).

$$y = \sum_{i=1}^{n} y_i^s A_i \qquad \text{(Eq. 30)}$$

where; y is the model equivalent of the satellite data, n is the number of levels of the satellite observations, y_i^s is the simulated concentration interpolated at the i-th level, A_i is the averaging kernel of the i-th level. Finally, modelled NO_2 tropospheric columns are calculated and compared to OMI NO_2 columns (the 2nd term of the cost function J, Equation 29).

3.2.1.5 Inversion experiments

This study performs CIF-CHIMERE inversions with a daily assimilation window for 2015 and 2019. The inversion is considered convergent when the gradient norm of the J function is reduced by more than 95 % (Fortems-Cheiney et al. 2021). By this criterion, there are 267 and 296 convergent days per year, respectively. Due to the variability of the daily spatial coverage of OMI observations, we consider monthly averaged corrections to build the final posterior emission inventories. Examples of monthly corrections are shown in the Figure 41 for the month August in 2015. Example of innovation and residue map for the same month is seen in Annex I: Innovation & Residue Map. It is seen that, the system goes towards to the observations where the simulations have higher monthly concentrations than OMI observations, in particular in the urbanized regions.

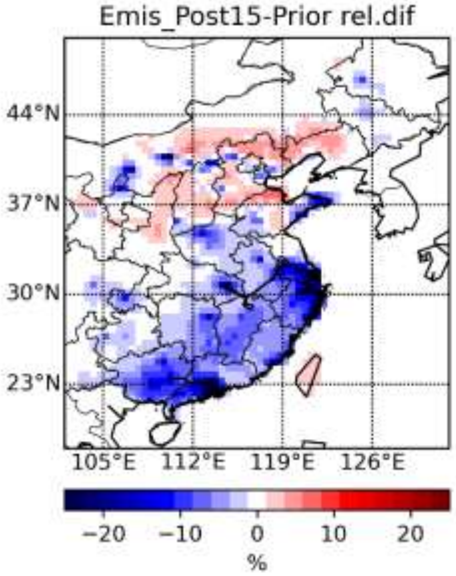


Figure 41 Monthly Corrections of NO_x emissions for August 2015

The main inversion run configurations made for 2015 and 2019 are summarized in Table 6.

Table 6 Summary of the inversions and simulations performed in this study. The output column gives the short names, which will be used in the following.

Type	Target Year	Emission Inventory	Objective	Output
Inversion	2015	HTAP v2.2 2010	Estimate monthly anthropogenic NO_x Emissions	Emis_Post15
Inversion	2019	HTAP v2.2 2010	Estimate monthly anthropogenic NO_x Emissions	Emis_Post19
Control	2015	HTAP v2.2 2010	Simulate hourly NO_2 Concentrations	Conc_Prior15
Control	2019	HTAP v2.2 2010	Simulate hourly NO_2 Concentrations	Conc_Prior19
Experiment	2015	Emis_Post15	Simulate hourly NO_2 Concentrations	Conc_Post15
Experiment	2019	Emis_Post19	Simulate hourly NO_2 Concentrations	Conc_Post19
Evaluation	2015	MEIC v1.3 for NO_x + HTAP v2.2 for other species	Simulate hourly NO_2 Concentrations	Conc_MEIC15

3.2.2 Evaluation data and the method

3.2.2.1 MEIC bottom-up emission inventory

MEIC (Multi-resolution Emission Inventory for China) is a bottom-up emission inventory developed and maintained at Tsinghua University since 1990 (http://meicmodel.org.cn/?p=1579&lang=en Last access: 26/09/2022). It is designed to track anthropogenic emissions over mainland China in various spatial resolutions. In this inventory, NO_x emissions are mainly presented in five sectors: energy, transportation, industry, agriculture, and residential. More details about inventory construction and sectors can be found in Li, Zhang, et al. 2017b; Zheng et al. 2018.

In this study, we use MEIC v1.3 from 2015 (Zheng et al. 2018; 2021) as an independent inventory to evaluate the estimated posterior emissions for the same year (Emis_Post15). It should be noted that, this version of MEIC contains the impacts of the implementation of China's Air Pollution Prevention and Control Action Plan. The MEIC inventory is also named MEIC15 in the following.

3.2.2.2 CNEMC ground-based observations

The China National Environmental Monitoring Centre (CNEMC) network, operating under the China Ministry of Ecology and Environment (MEE), has provided nationwide hourly in-situ concentrations on pollutants such as SO_2, O_3, NO_2, CO, PM2.5, and PM10 since 2014. These measurements are carried out using a commercial chemiluminescence analyzer equipped with a molybdenum converter.

This study considers surface NO_2 concentration measurements from 648 stations in 2015 and 626 stations in 2019, located in Eastern China.

3.2.2.3 Evaluation and analysis methodology

The emission estimates (Emis_Post15 and Emis_Post19) are evaluated indirectly by comparing the NO_2 concentrations simulated in the experiment runs defined in Table 6 with the surface measurements. The simulated and observed daily concentrations are calculated from the hourly simulated and observed concentrations, and compared. The simulated concentrations are bilinearly interpolated at the location of the stations. The normalized mean biases (NMB), the root-mean-square error (RMSE), and the Pearson correlation coefficient (*r*) are the metrics used for the evaluation (Annex D: Statistical Indicators). The measured concentrations are taken as the reference. As already mentioned, it is also necessary to perform a control run based on the prior emissions to evaluate if the inversions help to better assess the emissions compared to the prior (for simplicity, we will use the term of improvement in the following). In addition, calculating the ratios of the statistical indicators between the experiment run and the control run for the same year allows one to quantify the improvement brought by the inversions (Annex E: Ratio Formulas). Indeed, when the ratios of the mean bias and of the RMSE are smaller than one and/or the correlation ratio is higher than one, we can consider that the new emission estimates are improved compared to the prior knowledge of these emissions. In addition, we also performed a run for 2015 where the emissions from HTAP v2.2 are replaced with the NO_x emissions from MEIC version 1.3. This allows an evaluation of the emission estimates against an independent inventory.

It is worth noting that, for the comparison between models and surface stations, Lamsal et al., 2008 recommended to apply a correction factor to the measured NO_2 concentrations. Indeed, due to the measurement analyzer which catalyzes NO_2 into NO, other reactive oxidize nitrogen compounds (NO_z) such as peroxyacetyl nitrate (PAN), and nitric acid (HNO_3) can be also converted to NO to a certain extent, which results in an overestimation of ambient NO_2 concentrations (Lamsal et al. 2008). For this reason, Lamsal et al. 2008 suggested using the following correction factor (*CF*) formula to alleviate and avoid this overestimation of NO_2.

$$CF = \frac{NO_2}{NO_2 + \sum AN + 0.95PAN + 0.35HNO_3} \qquad \text{(Eq.31)}$$

We use CHIMEREv2013 to simulate NO_2, the sum of all alkyl nitrates (*AN*), *PAN*, and HNO_3 at a resolution of 0.5°×0.5° for 2015 and 2019, respectively. Calculated CFs are bilinearly interpolated according to the location of the stations (latitude, longitude) before being applied to the hourly measured ground-based NO_2 concentrations.

To avoid representativeness issues during the comparison of simulated concentrations at the coarse model resolution and the surface measurements, we first classify the stations according to their environment type. The types of the stations follow the classification done in Lachatre et al. 2020 using the method of Flemming, Stern, and Yamartino 2005. In the method they use ozone and its diurnal variability to evaluate whether stations represent rural, suburban, urban, or traffic environments. In our analyzes, we include stations in rural, suburban and urban environments and exclude stations in the traffic environment, as they are near sources and are not comparable to the coarse model resolution.

In addition to the evaluation of the emission estimates using surface measurements, we also analyze the emissions Emis_Post15 and Emis_Post19 in terms of annual, monthly, and regional emission budget. The emission budgets are calculated as the sum of emissions in each grid cell of the region / domain and over the time period considered. Comparison with the independent MEIC inventory for 2015 on an annual and monthly basis over Eastern China and selected urbanized locations are also performed and analyzed in Section 3.3.

3.3 Results and Discussion

3.3.1 Evaluation of the CIF-CHIMERE system and NO_x emission estimates with ground-based measurements

The emission estimates evaluation is not direct and is based on the experiment and control runs described in Section 3.2. Then, we compared simulated surface NO_2 concentrations using prior (Conc_Prior), posterior (Conc_Post), and MEIC (Conc_MEIC) NO_x emissions for 2015 and 2019 with NO_2 ground-based measurements over Eastern China. Although the global north / south gradient observed by the ground-based measurements is rather well captured by the simulations using prior and MEIC inventories (Figures 42b and 42c), the model tends to overestimate concentrations in the North China Plain (NCP) and underestimate them in the western, northeastern, and southern regions except in the Pearl River Delta (PRD) (Figure 42d) for 2015. It is worth noting that this behavior is similar in simulations using posterior emissions (Annex F: Annual concentrations 2015).

In 2019, we observe the same north / south pattern with an underestimation in the south (except in the PRD) and an overestimation in the NCP region (Figure 43).

The statistics of the comparison between the simulations and the surface measurements are reported in Table 7.

For 2015, the annual mean of surface NO_2 concentrations calculated from hourly concentrations at the stations is 27.4 $\mu g/m^3$ over East China. The annual mean value of NO_2 Conc_Prior15 and

NO_2 Conc_Post15 are higher than the stations' mean value, whereas NO_2 Conc_MEIC15 is smaller, suggesting that while NO_2 Conc_Prior15 and NO_2 Conc_Post15 tend to overestimate measured surface concentrations, NO_2 Conc_MEIC15 underestimates them. Note that, the annual mean value of NO_2 Conc_Post15 is slightly closer to the stations (~4.7 %) than NO_2 Conc_Prior15 (~ 6.6 %) and NO_2 Conc_MEIC15 (~ −7 %). In 2019, the annual mean value of NO_2 Conc_Post19 (29.5 µg/m^3) is also slightly closer to the stations (24.8 µg/m^3) than NO_2 Conc_Prior19 is (30.3 µg/m^3).

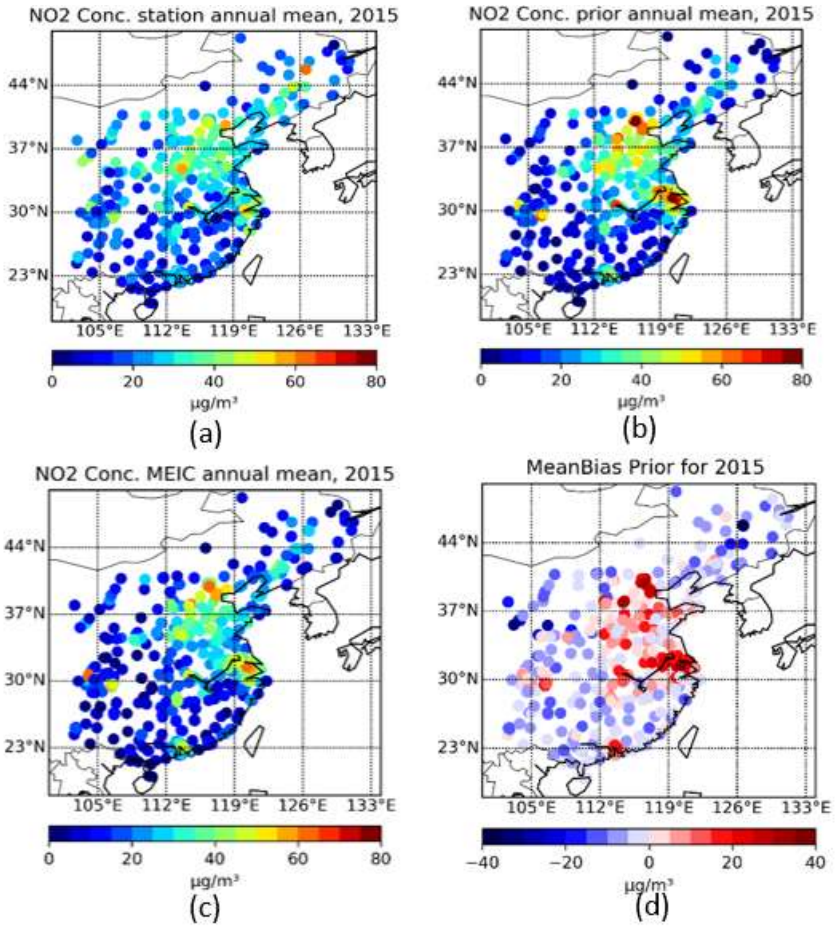


Figure 42 Spatial distribution of annual mean surface NO_2 concentrations: (a) Stations; (b) CHIMERE using prior emissions (Conc_Prior15); (c) CHIMERE using MEIC emissions (Conc_MEIC); (d) Mean bias calculated by the mean differences between Conc_Prior15 and stations for 2015

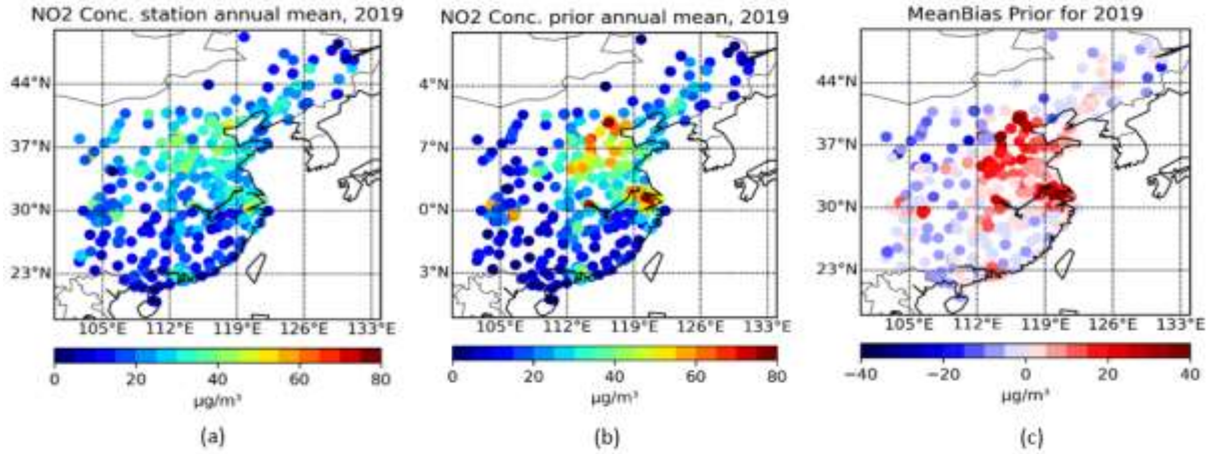


Figure 43 Spatial distribution of annual mean surface NO_2 concentrations: (a) Stations; (b) CHIMERE using prior emissions (Conc_Prior19); (c) Mean bias calculated by the mean differences between Conc_Prior19 and stations for 2019

Table 7 Annual statistics for the comparison between simulated and measured NO2 hourly surface concentrations for 2015 and 2019. Mean = Annual Mean, SD = Standard Deviation (SD), RMSE = Root Mean Square Error are in µg/m^3. r = Pearson correlation coefficient. NMB = normalized mean bias in % (Formulas used here are given in the Annexes A and D), (SDs are calculated using annual means of the concentrations)

	Station		Conc_Prior					Conc_Post					Conc_MEIC				
Year	Mean	SD	Mean	SD	r	NMB	RMSE	Mean	SD	r	NMB	RMSE	Mean	SD	r	NMB	RMSE
2015	27.4	11.6	29.2	18.6	0.63	7	22.1	28.8	18.1	0.64	5	21.6	25.5	15.8	0.62	-7	20.6
2019	24.8	9.8	30.3	19.2	0.67	22	21.9	29.5	18.2	0.68	19	20.8					

The other statistical indicators in Table 7 show also that they are globally slightly better for NO_2 Conc_Post than for NO_2 Conc_Prior for the two years and for NO_2 Conc_MEIC for 2015. The larger impact of the use of the posterior emissions to simulate NO_2 concentrations is visible on the NMB, which decrease by about 30 % and 15 % compared to the prior in 2015 and 2019, respectively.

These first results suggest that the CIF-CHIMERE system constrained with OMI NO_2 observations behaves as expected, bringing the simulated NO_2 concentrations closer to the truth – given by the surface measurements – when the emissions estimates are used in a CTM. However, it is worth noting that the improvement remains small.

To complete this global evaluation of Chinese annual means, we also evaluated the statistics, station by station, to see how homogeneous they are over the domain and if spatial features appear. As described in section 3.2.2, we calculated the ratios of the experiment and control run statistical indicators (r ratios, mean bias ratios, and RMSE ratios) at each station (Figure 44 for 2019, Annex G: Statistical Indicators' Ratios for 2015). The r ratios are mainly larger than one for most stations both in 2015 and 2019, showing a better agreement with the surface measurements in terms of temporal correlation when the Emis_Post is used to simulate NO_2 concentrations. The improvement is more significant in the south of the domain, especially in the highly urbanized locations of the PRD and the Yangtze River Delta (YRD). The improvement is relatively small, but it might be explained by the fact that the correlation coefficient is calculated on hourly concentrations, whereas the emissions are corrected on a monthly basis (Section 3.2.1). Concerning the mean bias and the RMSE, the situation is more contrasted with ratios smaller than one (showing an improvement), mainly above the most urbanized area of the PRD, Wuhan, and YRD. This is particularly the case in 2019. In these regions, the reduction of the mean bias of the simulated NO_2 concentrations compared to the surface measurements can reach up to 30-40 % when the posterior emission estimates are used.

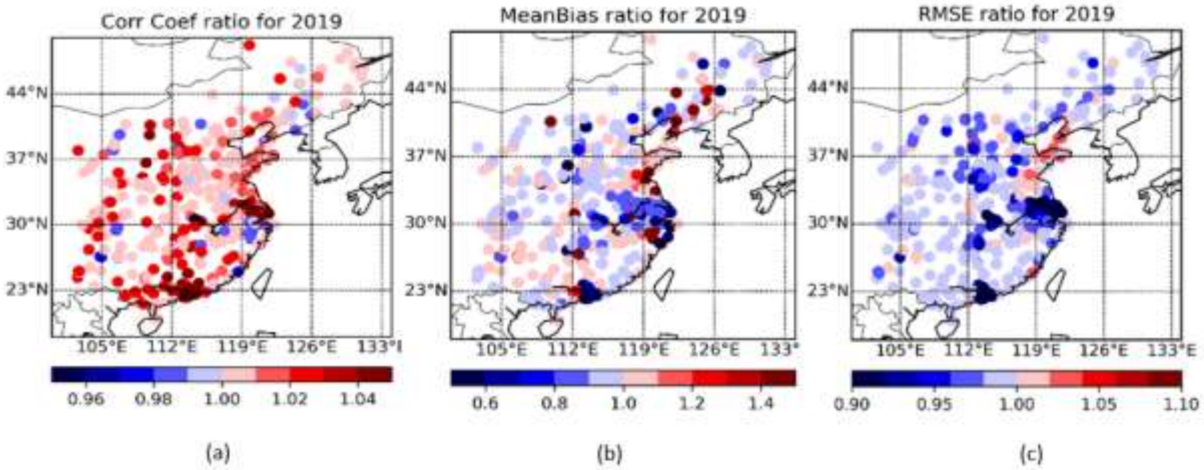


Figure 44 Ratios of the Conc_Prior19 and Conc_Post19 statistical indicators: Correlation Coefficient, Mean Bias and RMSE, respectively (a-c), for 2019.

As a result of these analyses, the inversions using OMI observations and CIF-CHIMERE seem to correct NO_x emissions towards the truth as the NO_2 concentrations simulated using these emissions are closer to the surface measurements. The improvements compared to the surface measurements remain small over the whole domain, but are much more significant in large urbanized regions suggesting that the OMI NO_2 observations allow a better constrain on polluted regions than on background regions.

Figure 45 show an example of the comparison of monthly averaged NO_2 surface concentrations from simulations and stations for 2015 and 2019 for Wuhan, where the emission corrections are the largest. The stations selected for this comparison correspond to grid cells in Wuhan with minimum 60 kt emissions (Figure 46).

In both years, simulations using prior NO_x emission inventory overestimate NO_2 surface concentrations compared to the stations. Using posterior NO_x emission inventory decreases the differences between simulated NO_2 surface concentrations and measurements, especially from May to October 2015 and March to October 2019. In both years, the highest decrease is in the summer months (−22 % in August 2015 and −42 % in August 2019).

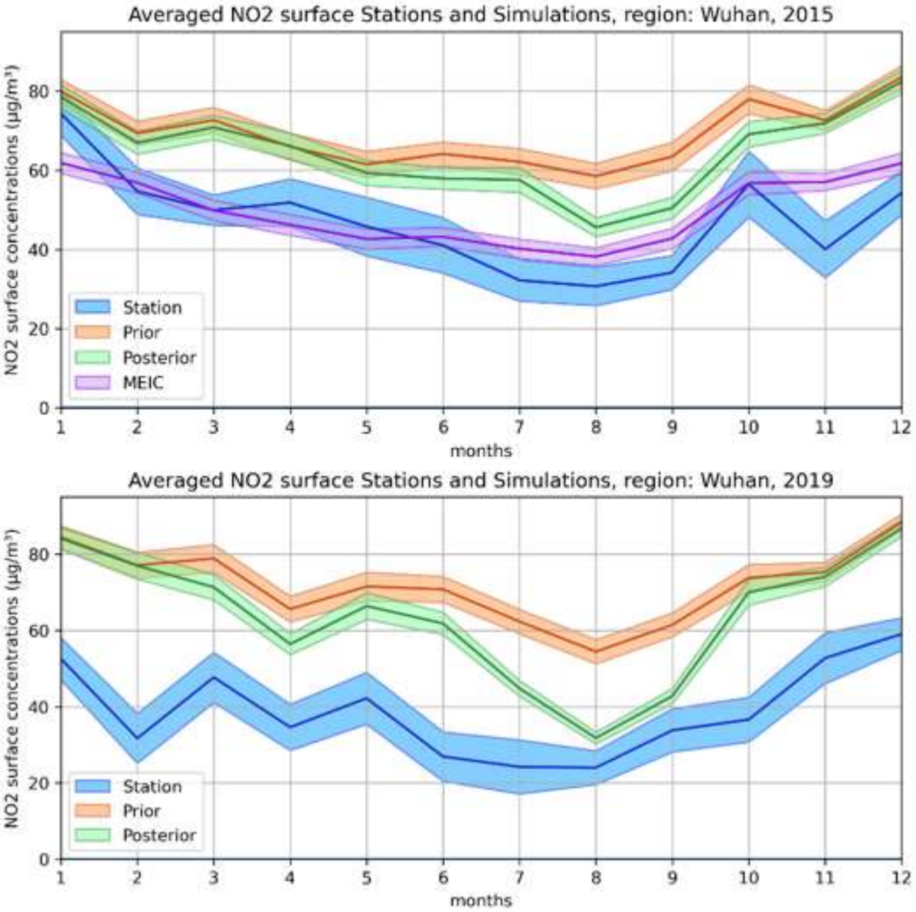


Figure 45 Monthly averaged NO_2 surface concentrations from stations and simulations for Wuhan in (top) 2015 and, (bottom) 2019.

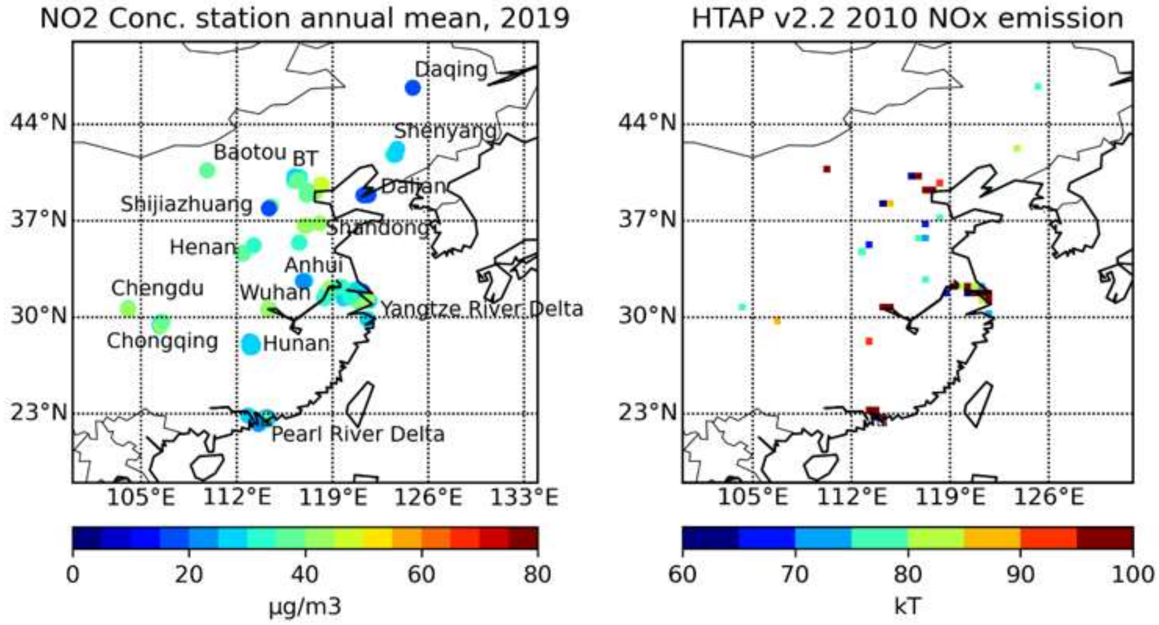


Figure 46 (Left) Ground-Based Stations according to the selected locations, (Right) HTAP v2.2 NO_x Emissions after applying the 60kt threshold (2010) to determine the selection of the locations. The remaining 0.5° × 0.5° grid cells correspond to the selected location.

3.3.2 Analysis of the estimated Chinese annual anthropogenic NO_x emissions in 2015 and 2019

The annual NO_x emission budget over East China and including other countries in the domain of Figure 40 are reported in Table 8 for several inventories. In 2015, the MEIC inventory gives NO_x emissions 12.8 % smaller than our estimation for East China. Liu et al. 2018 estimates NO_x emissions for 2015 using the DECSO algorithm (Mijling and van der A 2012) also combining CHIMERE v2013b and OMI but a different version (DOMINO v2 (Boersma et al. 2011)), with a finer resolution (0.25° × 0.25°). According to this study, their annual anthropogenic NO_x budget estimation (21.5 Tg) is calculated over the entire domain of Figure 40, also including e.g., the Korean Peninsula, a part of Japan, and is 5.2 % larger than our calculations.

Table 8 The annual NO$_x$ emission budget (in Tg) over East China and the entire domain in the Figure 40 with respect to different inventories.

Inventory	**East China**	**Domain – Figure 40**
HTAP 2010	18.6	20.5
Emis_Post15 (this study)	18.5	20.4
Emis_Post19 (this study)	18.4	20.3
MEIC15	16.4	-
DECSO15 (Liu et al. 2018)	-	21.5

The different emission inventories are in rather good agreement, especially when comparing to the typical emission uncertainties (30-50 %). This is particularly the case for two based on top-down approaches involving OMI observations and the CHIMERE model.

The global annual NO$_x$ emission budget over Eastern China does not show significant changes from 2010 to 2019. However, significant spatial differences between HTAP v2.2 for 2010 and Emis_Post15 and Emis_Post19 inventories are visible as shown in Figure 47.

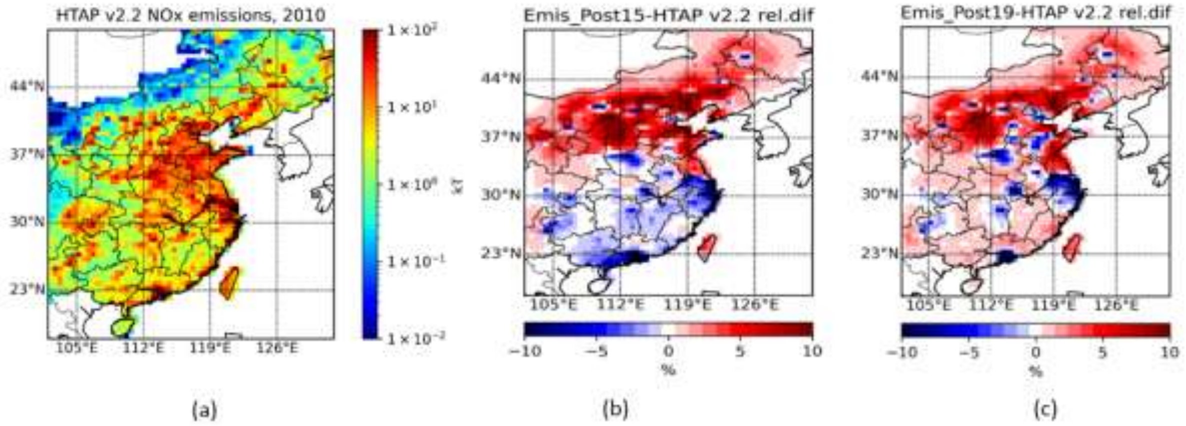


Figure 47 (a) NO$_x$ emissions at 0.5° × 0.5° resolution for 2010 from the HTAP v2.2 inventory in kt; Relative differences of NO$_x$ emissions for between 2010 and (b) 2015 and (c) 2019, respectively.

In 2015 and 2019, a global increase compared to 2010 is estimated in the northern part of China and a decrease in the south and large urbanized and industrialized locations. When the emissions are reduced compared to 2010, this reduction remains moderate and varies from region to region. For some locations, such as PRD, the reduction can reach up to −12 % in 2015, which is in fairly good agreement with the 12th Five-Year Plan's nationwide target of a 10 % reduction for 2015 compared to 2010 (de Foy, Lu, and Streets 2016). In 2019, the estimated reduction is more confined around urbanized locations than in 2015, especially in the south.

To evaluate how important these reductions in NO_x emissions are between 2015 and 2019, we analyzed the relative differences between the posterior emissions for 2019 and 2015 (Figure 48).

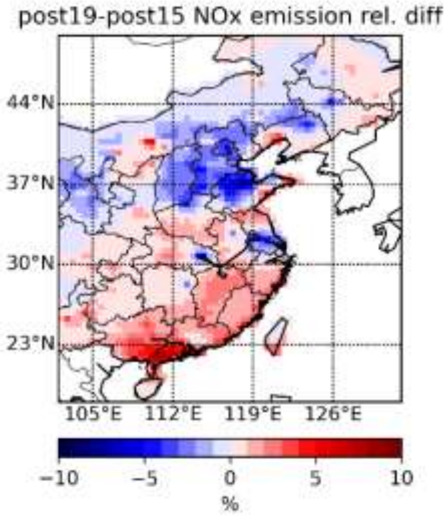


Figure 48 Relative differences between Emis_Post19 and Emis_Post15 NO_x emissions.

The differences between these two years range from ±8 % at the pixel scale. In the south, the emissions tend to increase slightly in 2019 compared to 2015; on the contrary, in the north and the locations such as Wuhan, YRD, and NCP, the emission reduction continues between 2015 and 2019 showing that the regulation effects are still ongoing. However, the reductions are smaller than 5 % below the expected levels in the 13th Five-Year Plan in which the emissions should decrease by 15 % compared to 2015 (“Notice of the State Council on Printing and Distributing the ‘13th Five-Year’ Ecological and Environmental Protection Plan_Government Information Disclosure Column” 2016). It is important to note that the changes in emissions inferred from our inversions as well as the target reductions of the 12th and 13th Five-Year Plans are smaller than the commonly assessed uncertainties on the emissions (30-50 %).

3.3.3 Analysis of the estimated annual anthropogenic NO_x emissions at selected urbanized and industrialized locations

3.3.3.1 Evaluation of the estimated annual anthropogenic NO_x emissions in 2015 and 2019

As discussed in section 3.3.1, the inversions seem more relevant in polluted regions. We analyze emission changes focusing on the most urbanized and industrialized locations. For this, we select only the locations with a minimum of NO_x emissions of 60 kt for 2010, which should represent the most urbanized and industrialized locations, where we expect stronger emissions control (Figure 46). According to this selection, we have detected 15 locations reported from the north to the south, in Table 9. They are also reported in the Figure 40.

Table 9 Definition of the locations and their annual NO_x budgets concerning the different emission inventories and years.

Location	**Latitude & Longitude**	**HTAP v2.2 (kt)**	**Emis_Post15 (kt)**	**MEIC15 (kt)**	**Emis_Post19 (kt)**
Daqing	46–47°N, 124–125.5°E	78.9	71	72.5	70.1
Shenyang	41–44°N, 123–125°E	82.8	84.9	114.5	81.1
Baotou	40–42°N, 108–111°E	136.3	116.2	116.4	122.3
Beijing-Tianjin	38–41°N, 115–118.5°E	450.8	447.3	381.2	431.7
Dalian	38–41°N, 120–122°E	74.2	77.4	88.4	76.3
Shijiazhuang	37–39°N, 113–115°E	151.1	152.2	107.2	144.5
Shandong	35–38°N, 115–119°E	296.6	295.8	245.4	279.2
Henan	34–37°N, 112–114°E	145.5	136	84.2	132.2
Anhui	32–33°N, 116–118°E	76.4	71.8	39.8	70.1
Yangtze River Delta	29–32°N, 118–122°E	1229.4	1162.1	963.8	1129.8
Wuhan	29–32°N, 113–116°E	260.7	237.8	123.2	222
Chengdu	30–32°N, 103–105°E	79.9	76.7	109.2	77.9
Chongqing	29–30°N, 106–107°E	87.4	82.6	72.3	84.5
Hunan	27–28°N, 112–113°E	94.1	87.5	99.9	85.6
Pearl River Delta	22–24°N, 112–115°E	420.8	370.9	390	375

In 8 out of 15 locations (Daqing, Beijing-Tianjin (BT), Shandong, Henan, Anhui, YRD, Wuhan, and Hunan), there is a decrease in NO_x annual emission budget between 2010, 2015, and 2019 (Table 10).

Table 10 Relative Differences of annual NO_x emission budgets at the 15 selected locations.

Location	Emis_Post15 – HTAP v2.2 (%)	Emis_Post19 – HTAP v2.2 (%)	MEIC15 – HTAP v2.2 (%)	MEIC15 – Emis_Post15 (%)
Daqing	-10	-11.1	-8	2.1
Shenyang	2.5	-2.1	38.3	34.9
Baotou	-14.8	-10.3	-14.6	0.2
Beijing-Tianjin	-0.8	-4.2	-15.4	-14.8
Dalian	4.4	2.8	19.2	14.2
Shijiazhuang	0.8	-4.3	-29.0	-29.5
Shandong	-0.3	-5.8	-17.2	-17
Henan	-6.6	-9.2	-42.2	-38.1
Anhui	-5.9	-8.2	-47.9	-44.6
Yangtze River Delta	-5.5	-8.1	-21.6	-17.1
Wuhan	-8.8	-14.9	-52.7	-48.2
Chengdu	-4.1	-2.5	36.7	42.5
Chongqing	-5.5	-3.3	-17.2	-12.4
Hunan	-7	-9.0	6.2	14.2
Pearl River Delta	-11.9	-10.9	-7.3	5.2

The highest decrease is observed in Baotou in 2015, with a 14.8 % decrease, and in Wuhan in 2019, with a 14.9 % decrease. Four locations (Baotou, Chengdu, Chongqing, and PRD) show a reduction in 2015 compared to 2010 but an increase in 2019 compared to 2015. At two locations, Shenyang and Shijiazhuang, NO_x emissions increase in 2015 but decrease in 2019 compared to 2010. Dalian is the only location where estimated NO_x emissions for 2015 and 2019 are higher than HTAP v2.2 emissions in 2010, but a decrease is observed between 2015 and 2019. Finally, there are no locations that show a monotonic increase between 2010, 2015 and 2019.

The annual emission budgets derived from the independent MEIC inventory for each location are also reported in Table 10 for comparison. The estimated posterior NO_x emissions for 2015 are closer to MEIC emissions at all locations, except for Shijiazhuang, Chengdu, and Hunan. Note that the correction of the emissions (Emis_Post - HTAP v2.2) is relatively small in Shijiazhuang, with only 0.75 % corrections. In Chengdu and Hunan, MEIC NO_x emissions are higher than HTAP v2.2 emissions, whereas the inversion tends to decrease them. These results also suggest that the corrections made to the emissions during the inversion are consistent.

3.3.3.2 Monthly evaluation of the anthropogenic NO_x emissions

Figure 49 shows the time series of monthly anthropogenic NO_x emissions for the different emission inventories considered in this study. In terms of seasonal variations, NO_x emissions are larger during wintertime, especially in December, at all the locations. However, a drop is observed for most of the locations in February, likely related to the Chinese New Year. This temporal evolution is consistent with the study of Ding et al. 2017 in which a 10 % reduction in February is estimated because of the Spring Festival and a peak is observed in December.

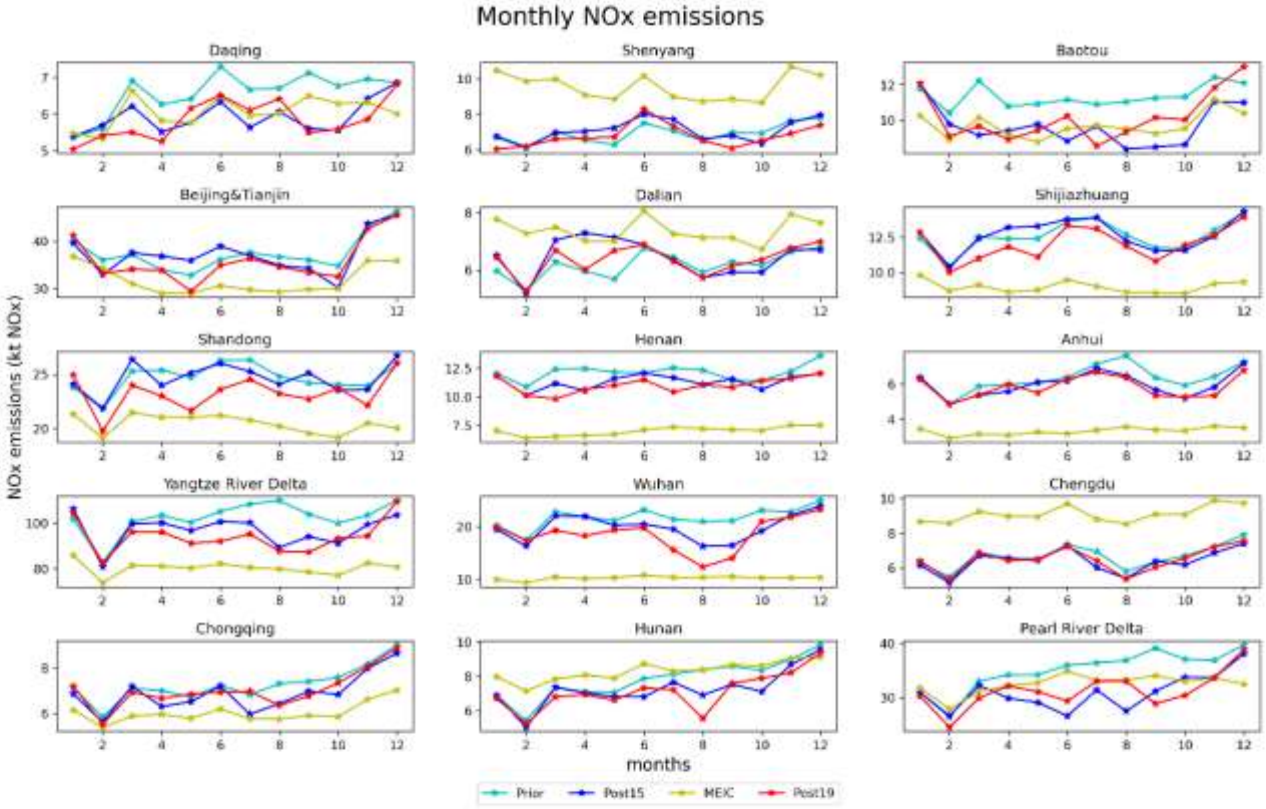


Figure 49 NO_x monthly budget comparisons at the selected locations.

For the winter months, very few emissions changes between the years (2010, 2015 and 2019) are observed. These very small changes may be explained either because prior emissions do not need to be corrected or because the emissions remained stable between the years or because few or no observations are available to constrain the inverse modeling system (2010 being also used as prior emissions). For example, the latter is the case in Daqing for January and December 2015, where no OMI observation data are available over the location.

At Daqing, Baotou, Henan, YRD, and PRD, posterior emissions decrease compared to 2010 throughout both years, with the largest reduction in YRD and PRD during summertime, 20 % and 26 %, respectively.

At Wuhan, Chongqing, Hunan, and Anhui, there are no significant changes in the NO_x emissions between 2010, 2015 and 2019 in the first half of the year, but a decrease compared to 2010 occurs mainly in the second half of the years, with the highest reduction observed in Wuhan (22 % in 2015 and 41 % in 2019) in August.

In the northern locations, particularly around the Yellow Sea, such as Dalian, BT, and Shijiazhuang, the 2015 emissions are increased during spring compared to 2010. The emissions decrease between 2015 and 2019 and come back to the 2010 levels for Dalian and BT and below the 2010 levels for Shijiazhuang.

In the southern locations (below 33°N), such as Anhui, YRD, Wuhan, Chongqing, Hunan, and PRD, 2015 and 2019 emissions are reduced compared to 2010, especially during the summer months. It is worth noting that the summer months correspond to the months where we have more days with OMI observations.

Concerning the MEIC inventory for 2015, MEIC shows lower emissions than the 2010 HTAP inventory over the year 2015 in almost all locations, except in Shenyang, Dalian, Chengdu, and Hunan where MEIC shows larger emissions in 2015 compared to 2010 (HTAP). Note that as already noticed on the global analysis, the Emis_Post15 emission estimations tend to get closer to the MEIC inventory.

3.3.3.3 Comparison with in-situ measurements

The NO_2 surface concentrations simulated using Emis_Post for 2015 and 2019 as well as the ones simulated using MEIC15 are compared with ground-based measurements at each of the 15 selected locations. The statistics of the comparison are displayed in Annex H: In-Situ Comparison. As for section 3.3.1, the comparison results of the concentrations simulated in the control runs and the surface measurements are also reported in the same Annex H. In 2015, the Emis_Post15 emissions allow the improvement of the simulated NO_2 surface concentrations compared to these simulated in the control runs for nine locations in terms of NMB, RMSE and/or r. In 2019, this number increases to thirteen locations but the NMB tends to be larger than in 2015. This is consistent with the results discussed in section 3.3.1, showing that the CIF-CHIMERE system constrained by OMI observations performs as expected.

By comparison, the use of MEIC15 emissions to simulate NO_2 concentrations in 2015 gives better statistics for eight locations compared to the emission estimated in this study for 2015. The better performances when using MEIC emissions correspond to locations where the MEIC emissions are significantly smaller than the Emis_Post15 and HTAP v2.2 (Beijing-Tianjin, Shijiazhuang, Shandong, Henan, Anhui, YRD, and Wuhan). This points out that the posterior emission corrections derived from OMI observations and the CIF-CHIMERE are likely underestimated. This might be partly explained by the limited daily spatial coverage of OMI measurements, which does not allow sufficient corrections of the emissions. TROPOMI satellite instrument observations with their better spatial resolution and coverage should help resolving this issue. On the contrary, when MEIC15 emissions are larger than Emis_Post15 or HTAP v2.2, a degradation of the statistics is observed compared to Conc_Post15 (Dalian, Chengdu, Hunan, PRD), suggesting that MEIC overestimate emissions at these locations. These results illustrate the complementariness of bottom-up and top-down approaches to assess NO_x emissions and the need of robust ground-based measurement network to evaluate the emission inventories.

3.4 Conclusions

This chapter estimates anthropogenic NO_x emissions over East China for 2015 and 2019 using the new inverse modeling system CIF-CHIMERE and OMI NO_2 satellite observations. One of the objectives was to evaluate if this new inversion system constrained by OMI observations can infer emissions closer to the truth. As true emissions are unknown, ground-based measurements of NO_2 surface concentrations are considered as the reference to evaluate the emission estimates indirectly. The analyses performed in this chapter show that using the NO_x emissions estimates to simulate NO_2 surface concentrations improves the comparison with ground-based measurements, as is expected when atmospheric inversion approaches are applied. The results also suggest that the OMI observations provide a better constraint to the NO_x emissions in the polluted areas compared to the background areas.

The annual anthropogenic NO_x emission budgets estimated in this study for East China are in fairly good agreement with the annual emission budgets derived from independent inventories such as MEIC (~13 %) and DECSO (~5 %) when considering the un-certainties on NO_x

emissions (30-50 %). Nevertheless, the analysis at the finer spatial scale of the most populated and urbanized locations of East China shows situations more contrasted. In terms of the evolution of the emissions between 2010, 2015, and 2019, according to our emissions' estimates, some locations such as Baotou, PRD, and Wuhan show a reduction of emissions between 2010 and 2015 close to the target reduction defined in the 12th Five-Year Plan. The evolution of the emissions between 2015 and 2019 is rather small. Our estimates show very slightly increased emissions, mainly in the southern locations, and a small decrease (<5 %) in the NCP or locations such as Wuhan and the YRD. The comparison of our emission estimates with MEIC emissions in 2015 stresses that our estimates likely underestimate the emission reductions between 2010 and 2015 in the most polluted locations of East China. In parallel, our emission estimates, when used to simulate NO_2 surface concentrations, lead to a better agreement with surface measurements in regions where MEIC indicates an increase of the emissions compared to 2010 – our estimates often suggest reduced or stable emissions. These results illustrate the benefit of having a variety of inventories derived from different approaches and of using their complementariness to assess NO_x emissions better. They also stress the need for a robust ground-based measurement network to evaluate the emission inventories.

In this chapter, we estimated NO_x emissions using daily OMI observations, to do that, we calculated emission corrections on a monthly basis due to the lack of daily coverage of OMI. However, new satellite observations with higher spatial and temporal coverage on a daily basis, such as TROPOMI on Sentinel 5p, could bring more information to the inversion system; and thus lead to more accurate emission estimates.

Chapter 4. NO_x emissions in Europe derived from satellite observations

4.1 Context of the study

The work presented in this chapter has been done in the framework of the ARGONAUT ANR project. The ARGONAUT project (pollutAnts and gReenhouse Gases emissiOns monitoring from spAce at high resolUTion) has been devised to help address the question of coordinated monitoring and mitigation strategies to achieve optimal reduction policies for both air quality and climate change related emissions. One of the main objectives of the project was to develop the inverse modeling capabilities of the French community for highly reactive pollutants (e.g., NO_x, NMVOCs), carbon monoxide, and GHG such as CO_2 at high resolution (0.1° × 0.1°) in France to exploit the new era of satellite observations at high resolution (the current TROPOMI and the future CO2M). These developments should help, in the end, to better take advantages of high resolution and co-emissions of species to constrain anthropogenic and biogenic sources separately. Considerations of the error statistics and the most adapted metrics, especially to solve the transport errors at a very high resolution of the point sources (2 km resolution), are also questions addressed in this project. One of the main outputs expected from the project is the provision of new estimates of French anthropogenic emissions of the primary air quality pollutants (NO_x, CO and NMVOCs) and potentially CO_2 at high resolution (0.1° × 0.1°), based on the state-of-the-art atmospheric inversion CIF using CHIMERE and highly resolved TROPOMI products. A stepwise strategy has been developed in the project, starting from inversions at a moderate resolution over the European domain (0.5° × 0.5°) to inversions at a higher resolution over France (0.1° × 0.1°). Doing inversions over the European domain allows one to compare the performances of the inversions constrained by observations at medium resolution with OMI and high resolution with TROPOMI. Indeed, the OMI resolution is not adapted to do inverse at 0.1° resolution over France. In this framework, the PhD work was to perform the inversion tests with the new TROPOMI observations for the different domains and to ensure the consistencies of the inversions by evaluating them with surface NO_2 measurements. The tests were done for the COVID-19 lockdown period in the spring of 2020, where sudden changes in anthropogenic NO_x emissions were expected and well reported from TROPOMI observations (Barré et al. 2021; Souri et al. 2021, and Figure 50). The tests presented in this chapter focus mainly on the first week of April 2020.

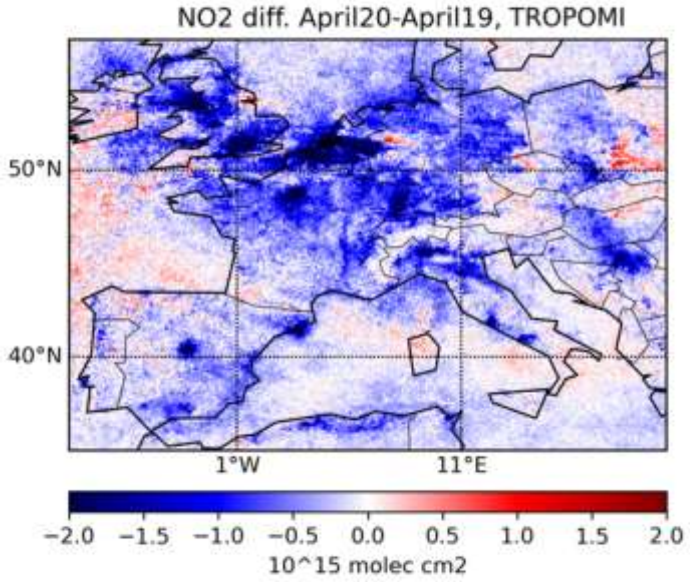


Figure 50 April 2020 and 2019 average NO_2 tropospheric column differences seen by TROPOMI (presented in EUMETSAT virtual conference in 2021)

4.2 Comparison of inversions using OMI and TROPOMI over Europe

As seen previously (Chapters 2 and 3), the CIF-CHIMERE inverse modeling system using OMI NO_2 satellite observations allows the estimation of NO_x emissions. We then use the set-up defined in Chapter 2 to perform emissions inversion using OMI and TROPOMI observations on a daily basis on the European domain (0.5° × 0.5°). We recall here that we use v1.3 data and super-observations at the model resolution calculated as the median of satellite pixels within the model grid-cell and the corresponding averaging kernel. In addition to the quality flags recommended by the respective user manuals of OMI and TROPOMI, we apply an additional filter based on the observation errors. Pixels with errors on the NO_2 tropospheric column larger than 100 % for OMI and 60 % for TROPOMI are filtered out. Figure 51 compares the NO_2 tropospheric columns observed by OMI and TROPOMI on April 1st, 2020. Despite using super-observations at 0.5° × 0.5° resolution, the best performances of TROPOMI, including a much higher spatial resolution, also lead to the significant difference in terms of sampling capabilities of the two instruments, TROPOMI leading to a much better spatial coverage capability. Note that our choice of filtering TROPOMI observations with 60% of observations errors rather than 100% like OMI is based on this better spatial coverage allowing one to be more restrictive in filtering the data while maintaining an improved coverage compared to OMI.

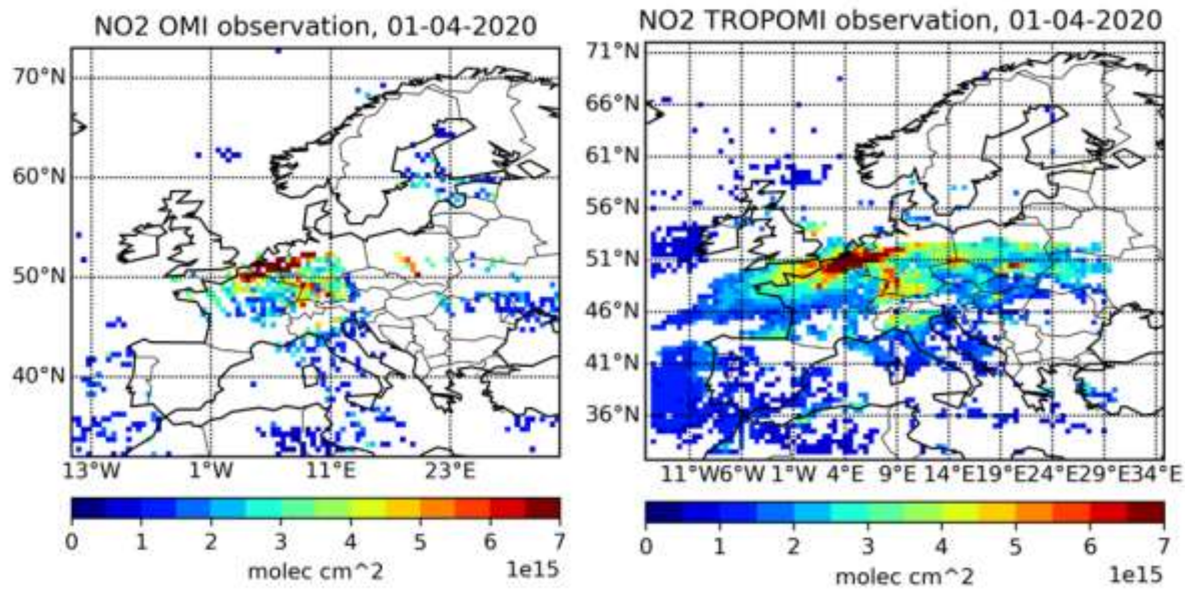


Figure 51 Comparison of NO_2 columns from OMI (QA4ECV) and TROPOMI (v1.3) and their coverage for the 01/04/2020

Figure 52 shows the results of the inversions using OMI and TROPOMI for April 1st, 2020. Note that we had to reduce the errors on prior anthropogenic emissions in the **B** matrix to 35% to ensure the convergence of the system with TROPOMI. In this figure, the posterior (corrected by inversion) emissions are compared to the prior emissions corresponding to the year 2016 (the CAMS-REG 2016 inventory is used as prior, see Chapter 2). Here again, the impact of the improved coverage of TROPOMI compared to OMI is well illustrated. Using the TROPOMI observation allows correction of emissions over a larger area than when using OMI. The inversion with TROPOMI shows a large reduction of the NO_x emissions compared to the prior. This reduction is much more limited when using OMI observations to constrain the inversion. Interestingly the patterns of increased emissions prescribed by the inversions (in red in the figure) are consistent between the two instruments.

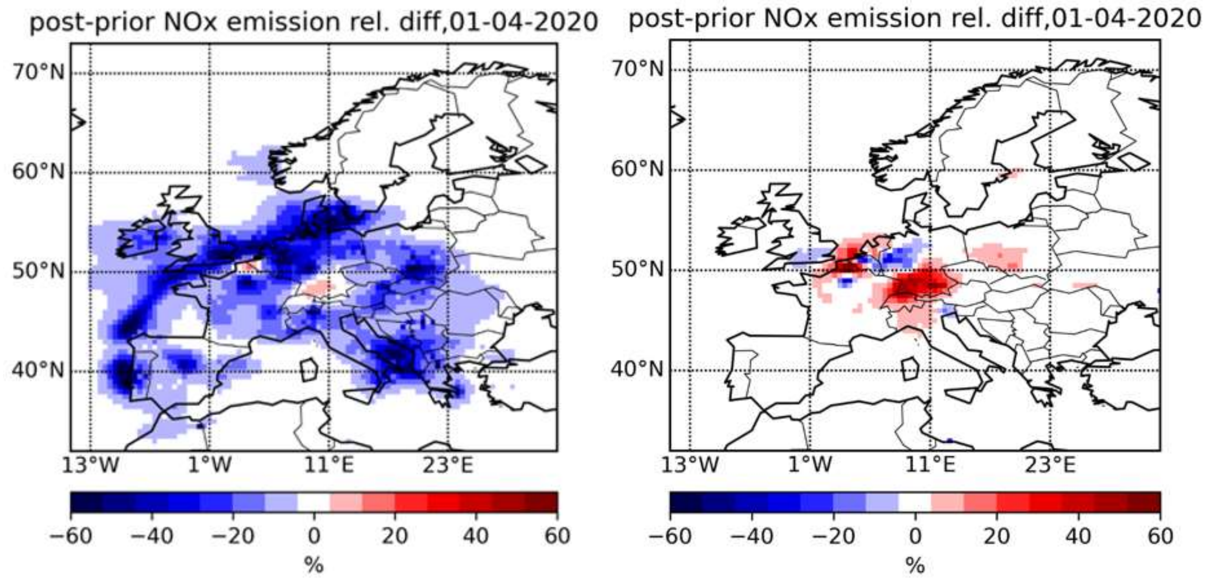


Figure 52 NO_x emission increments obtained by inversions using TROPOMI (left) and OMI (right) observations for 01/04/2020

Figure 53 shows a comparison of the hourly NO_2 concentrations simulated with CHIMERE using either the prior emissions (light blue) or the posterior emissions obtained with OMI (blue) and TROPOMI (violet) with surface measurements in the Paris area (example of one urban background station with (48.5°N, 2°E)) for April 1st, 2020. The concentrations simulated using prior emissions largely overestimate the observed concentration all over the day. The concentrations simulated with posterior emissions derived from OMI observations are in quite good agreement with the surface measurements during the morning peak, but they still overestimate the measurement during the evening peak. It is worth noting that, as we use a daily assimilation window, the same correction factor is applied to each hourly emission of the day (see Chapter 2). However, OMI observations are performed in the early afternoon, and they may not well represent the emissions later in the day and are more convenient to constrain emissions before the measurement. Note also that we make the assumption in the inversion that the diurnal profile of the emissions is correct in the model, which is likely not the case, especially during the COVID-19 lockdown when the practices of the population in terms of transport were completely unusual. Concerning the concentrations simulated with the posterior emissions corrected with TROPOMI, they mainly underestimate the measured concentrations at the station, especially during the morning peak. This would suggest that the corrections of the emissions prescribed by TROPOMI are too large.

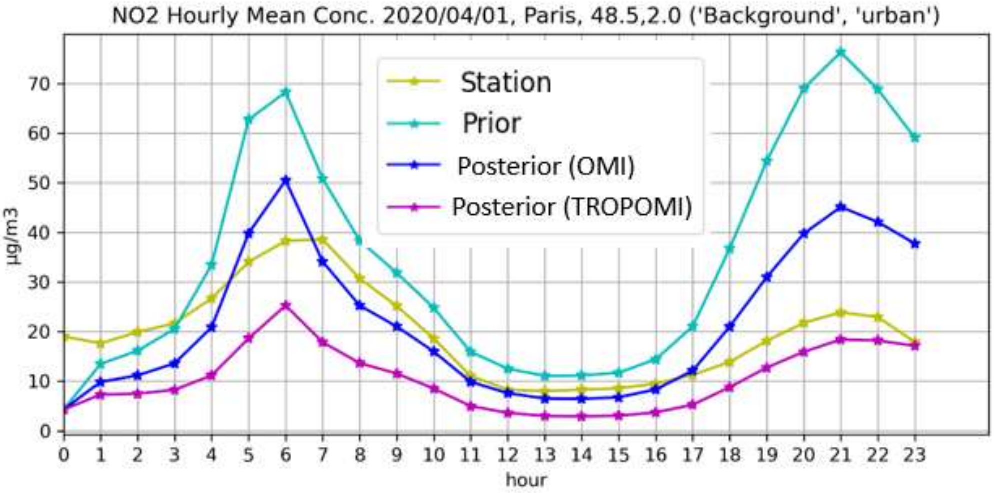


Figure 53 Comparison of the hourly NO_2 concentrations simulated with CHIMERE using prior and estimated emissions and observations at the Parisian background urban station

4.3 Difficulties in using TROPOMI observations for inversions

The underestimation of NO_x emissions obtained by inversion when TROPOMI observations are used is likely partly related to a systematic negative bias reported in TROPOMI observations. Indeed, the results presented in the previous subsection have been obtained with the TROPOMI observations v.1.3, available at the time of the study. Validation studies have shown that this version of the TROPOMI NO_2 product was biased negatively: around -51 % and -37 % in highly and slightly polluted areas, respectively (Verhoelst et al. 2021; Judd et al. 2020). This last negative bias was partly attributed to a degraded retrieval of the cloud pressure, which has been improved in the more recent version of the NO_2 TROPOMI product (van Geffen et al. 2022). Figure 54 shows an example of the differences between the product v1.3 and the PAL product v2.3, now available and used in the following (Pseftogkas et al. 2022).

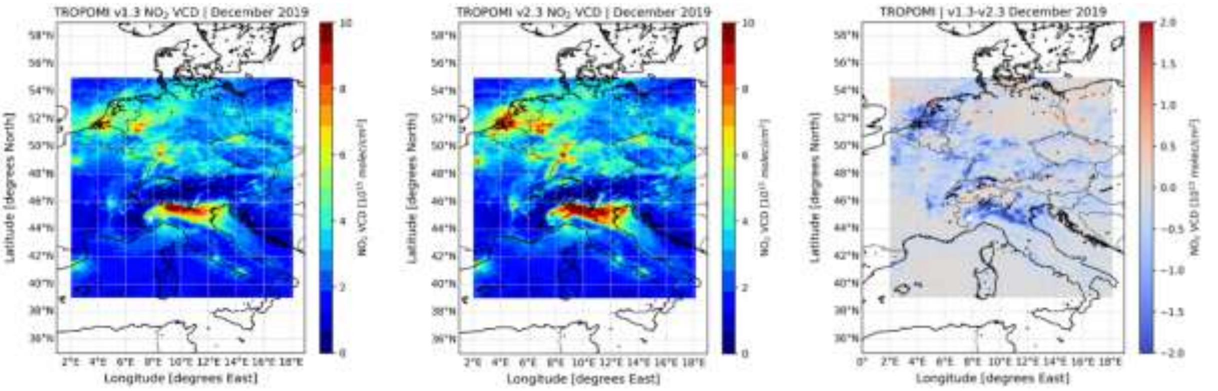


Figure 54 Comparison of averaged TROPOMI tropospheric columns in December 2019 for the v1.3 (left) and PAL v2.3 (middle), absolute differences (right). (The figure is taken from Pseftogkas et al., 2022)

Note that previous studies, such as, for example, Souri et al. 2021 chose to scale up TROPOMI NO_2 observations by 25% before including them in their inversion model to cope with this biases problem.

From our side, since CIF-CHIMERE is a new inversion system and TROPOMI is a quite new instrument (it is always a long process to get stabilized version of satellite products), using TROPOMI observations within the system was challenging throughout the PhD. In addition to this bias problem, we encounter strong convergence problems of the inversion system when using NO_2 TROPOMI observations. The inversions were convergent only for very few days. Applying a scaling factor to the NO_2 columns as Souri et al. 2021 did not help to solve the problem. The new TROPOMI PAL v2.3 product available since spring 2022 also did not solve our convergence problem. Several problems have been identified in the inversions process: (i) the inversion can lead to negative emissions, and (ii) the range of variations of the emissions and the NO_2 columns (several order of magnitude each) may give troubles to the inversion system to find an effective solution of the inverse problem. To solve these issues, new functionality has been included in the CIF by the developing team at LSCE, including the possibility of treating the control vector in a logarithmic scale. This allows one to solve our convergence problem during the inversion using TROPOMI as a constraint.

4.4 New set-up of the inversion

A new set-up of the inversion based on the use of the logarithm in the control vector has been defined for inversions over the European and France domains. The configurations used are based on the experience gained with the inversions performed with OMI (Chapters 2 and 3). They are detailed for the two domains.

4.4.1 Inversion over the European Domain

The configuration used for the European domain is close to the one we used for OMI inversion. To recall the configuration parameters:

- TROPOMI PAL v2.3 super observations (and their AKs) calculated by median were considered with quality flag higher than 0.75 and observations errors lower than 60 %.

- Anthropogenic and biogenic emission errors in **B** matrix were taken as 50 % and 100 %, respectively, with a 50 km horizontal correlation.

- Convergence criteria were taken as a 95 % reduction of the gradient norm.

- Daily inversions were performed for the week of 1-7 April 2020 (a week with the COVID-19 lockdown). The computation duration of one day of inversion was large (around 12 hours) and limits our capabilities to perform inversion over a longer period than a week, which would have been better to be more representative of the specific period of the lockdown and lead to more consolidated results.

- Daily inversions were also realized for the same week as above but for 2019. However, we only have two days of inversion. Changes in the super calculator (e.g., operating system changes in Très Grand Centre de Calcul du CEA (TGCC)) used to perform the inversions led to technical issues running the CIF-CHIMERE system, which could not have been solved by the time of this study. This prevents a meaningful comparison between emissions in 2019 and 2020 as initially planned.

4.4.2 Inversion over the France (ARGFR) Domain

To estimate NO_x emissions for France at high resolutions (0.1° × 0.1°), the following configuration of the parameters was taken:

- TROPOMI PAL v2.3 super observations (and their AKs) calculated by median were considered with quality flag higher than 0.75 and observations errors lower than 60 %.

- Anthropogenic and biogenic emission errors in **B** matrix were taken as 30 % and 100 %, respectively, with no horizontal correlation (for simplicity).

- Convergence criteria were taken as a 95 % reduction of the gradient norm.

- ARGFR domain includes 121 lon × 165 lat grid cells and corresponds to -5°N – 10°N and 39.75°E – 54.25°E and covers France.

- Daily inversions were performed for more than 20 hours. This coupled with technical issues on the super calculator (see above) limits our testing capabilities over the domain. Only preliminary results are shown.

4.5 Evaluation of inversion coherence of log and linear scaled emissions: China + OMI case

Before applying the new set-up of the inversion, including the logarithmic scale of the control vector, we wanted to test the consistency of the inversion obtained when using or not using this logarithmic scale. This was also important to check that the results presented in Chapter 3 would not have been much affected by using this new control vector. To do this, we used our previous work, where convergent inversion was obtained using the OMI observations for Eastern China. For this comparison, the day 23/05/2019 was chosen because OMI observations have good spatial coverage over China (it was the best of all days in 2019). Figure 55 presents the comparisons of the NO_x emission corrections in percentage using them as they are (left side) and in a logarithmic scale (right side). The negative and positive patterns of corrections of the emissions are similar for the two inversions. The magnitude of the corrections is mostly similar

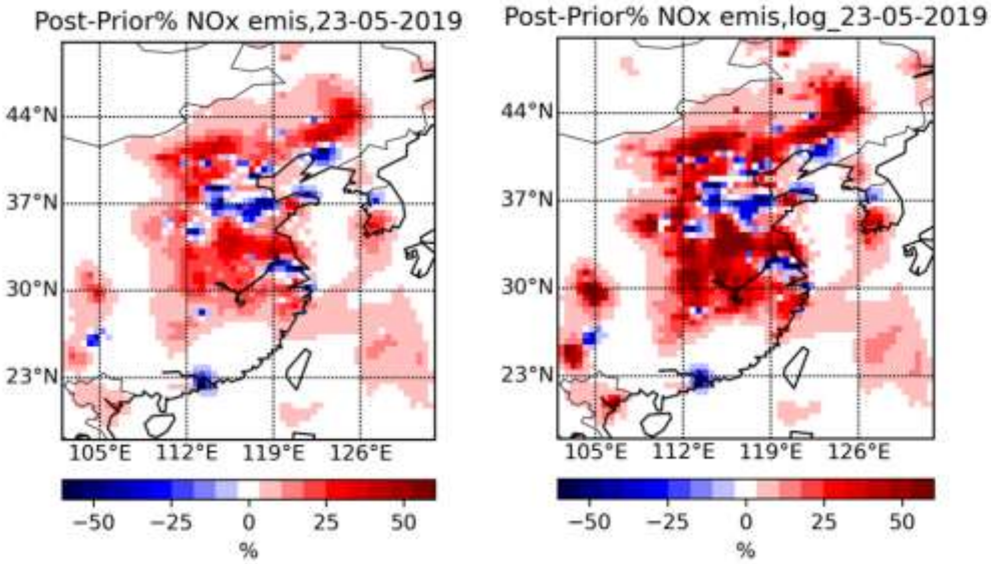


Figure 55 Comparison of NO_x emission corrections in % by performing inversion with the emissions in linear scale (left) and log scale (right) for the day 23/05/2019.

for the negative corrections in the most urbanized regions but tends to be significantly larger for the positive corrections. These positive corrections concern mostly moderated or low emissions and should have less impact. It would be interesting to extend the study to a larger period to correctly quantify the impact of using a logarithmic scale on emission inversion and assess if it improves the performance.

4.6 European NO_x emissions during April 2020 compared to 2016

4.6.1 Emission comparisons

Inversions covering the period April 1-7, 2020, were successfully performed using TROPOMI PAL v2.3 observations and prior emissions from 2016. Figure 56 shows the distribution of prior NO_x emissions for the same week over Europe (1-7 April), where the total amount for this week is calculated as 164 kt.

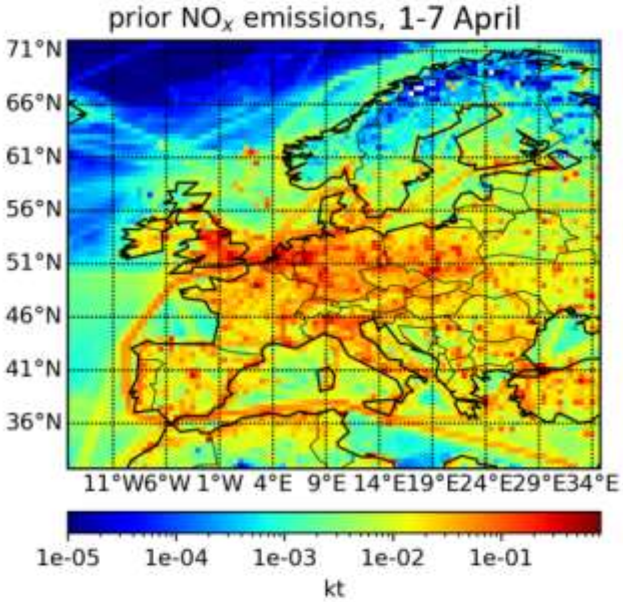


Figure 56 Distribution of total NO_x emissions for the week of 1-7 April from a prior inventory

The inversions we have performed for 2020 estimate a NO_x emission budget of 134 kt for this period, 19% lower than for the prior, representative of 2016. Figure 57 shows the emission corrections calculated by these inversions. According to this figure, the reduction of emissions for the first week of April is mainly seen in the northern part of Western Europe: south of the UK, north of France, Benelux, Germany, Poland, and north of Italy. The reduction against prior (2016) emissions reaches up to 67% locally. In other countries and regions, reductions are limited to major cities such as Barcelona and Madrid in Spain. Note that the system does not estimate any increase in NO_x emissions. Table 11 compares NO_x budgets from prior and estimated inventories by country. Accordingly, the highest reductions in emission budgets were achieved in Belgium and the Netherlands, with 48.2 % and 43.8 %, respectively.

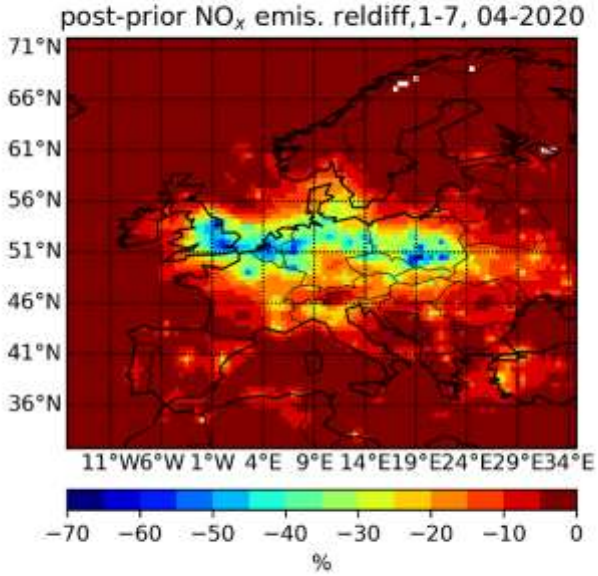


Figure 57 Increment in NO_x emissions (in %) for the week of 1-7 April

Table 11 Prior and Estimated NO_x emission budgets by some countries for the week of April 1-7, 2020, ordered from the highest to lowest difference in budget change.

Country	Prior (kt)	Posterior (kt)	%	weekly coverage of TROPOMI (%)
Belgium	2.7	1.4	-48.2	81
the Netherlands	3.2	1.8	-43.8	81
the UK	10.4	6.3	-39.4	30
Poland	9.2	5.7	-38	61
Germany	15.9	10.1	-36.5	79
France	11.6	9.1	-21.6	56
Italy	8	7	-12.5	62
Spain	7.1	6.5	-8.5	32

These results suggest that the emission reduction compared to the prior is larger in the northern European countries. However, it is important to check if the difference in emissions between the north and the south does not link with a sampling issue, as we consider only a limited number of days for this evaluation. Figure 58 shows the day-by-day TROPOMI observations for the studied week. The sampling of the northern countries is rather similar all over the week but more variable in the southern part of Europe, such as in Spain or Greece.

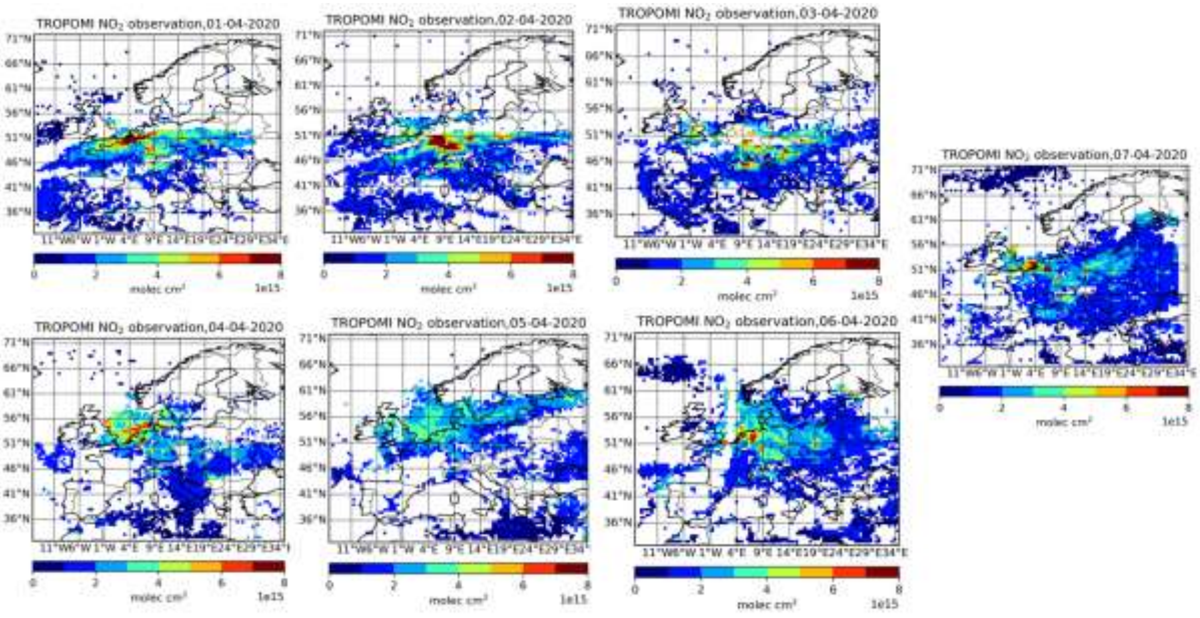


Figure 58 Day by day evolution of TROPOMI NO$_2$ columns (above figures are from the 1st to 3rd, below figures are from the 4th to 6th, and rightmost figure is for the 7th of April)

In terms of daily corrections of the emissions, Figure 59 shows that the corrections are mostly concentrated in the countries and regions listed previously. These corrections are rather homogeneous and negative (around -50 % in the Paris area, -60% in southern England, -55% in Benelux) from one day to another except for an increase of emissions prescribed in the south of Germany on April 2nd, and north of Italy on April 3rd. According to this sampling issue, the values of reductions (and the mean of weekly satellite coverage per country) given in Table 11 for Spain and Italy are likely, not representative.

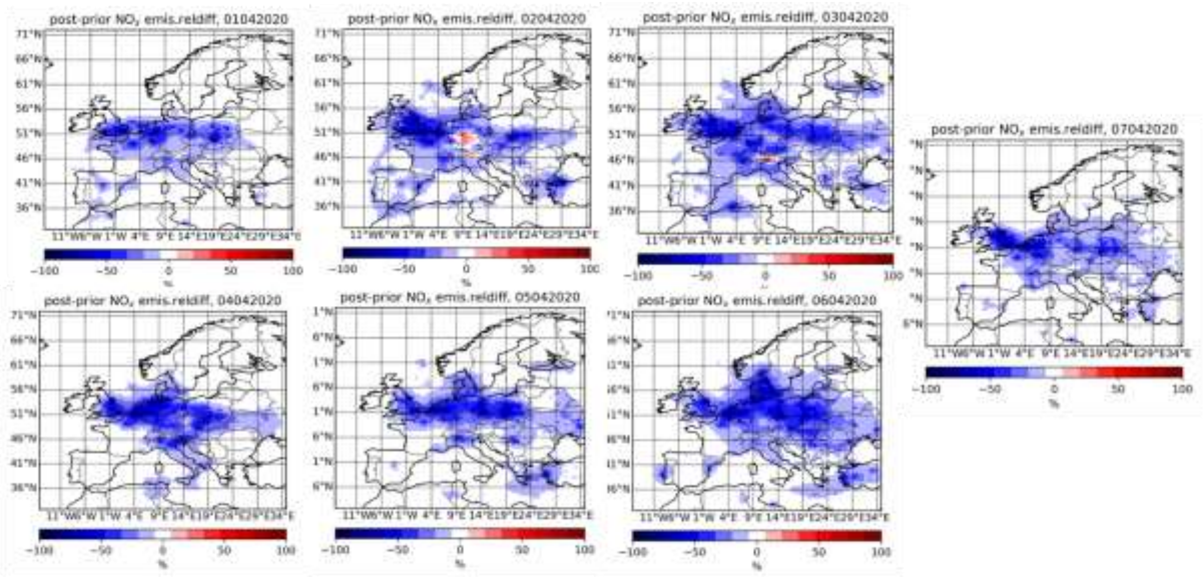


Figure 59 Daily variation of NO$_x$ emission corrections from 1st to 7th of April 2020

The comparison between the prior and the estimated emissions shows mainly a reduction of emissions during the first week of April when lockdowns were applied all over Europe and when one expects a significant drop in the NO_x emissions due to these restrictions. However, it is important to note that the reason for this decline between prior and posterior is not only due to the COVID-19 lockdown but also integrates the downward trends in NO_x in recent years as we compared with the prior emissions, which are representative of 2016 (e.g., -14 % decrease in France, 12 % decrease in the Netherlands in 2019 compared to 2016 (European Union Emission Inventory Report 2021)). Another explanation of the prescribed reductions may also arise from negative biases in the TROPOMI data, which may not be completely resolved in the PAL version of the TROPOMI product used. van Geffen et al. 2022 validated NO2 PAL v2.2 data against ground-based instruments and found that using PAL v2.2 improves negative biases with the ground-based data compared to v1.3, but still, PAL v2.2 underestimates NO_2 concentrations. PAL v2.2 and PAL v2.3 have minor differences, which do not affect the column densities during retrieval or data quality.

4.6.2 Evaluation with surface measurements

As for the Chinese study in Chapter 3, we conducted an evaluation of the estimated emissions, comparing NO_2 surface measurements to NO_2 concentrations simulated with the prior emissions and with the estimated emissions. For this comparison, surface data are corrected with the factors explained in Section 2.4.

Figure 60 presents the spatial distribution of stations, as well as a comparison of the weekly average of NO_2 surface concentrations for stations, prior and posterior simulations.

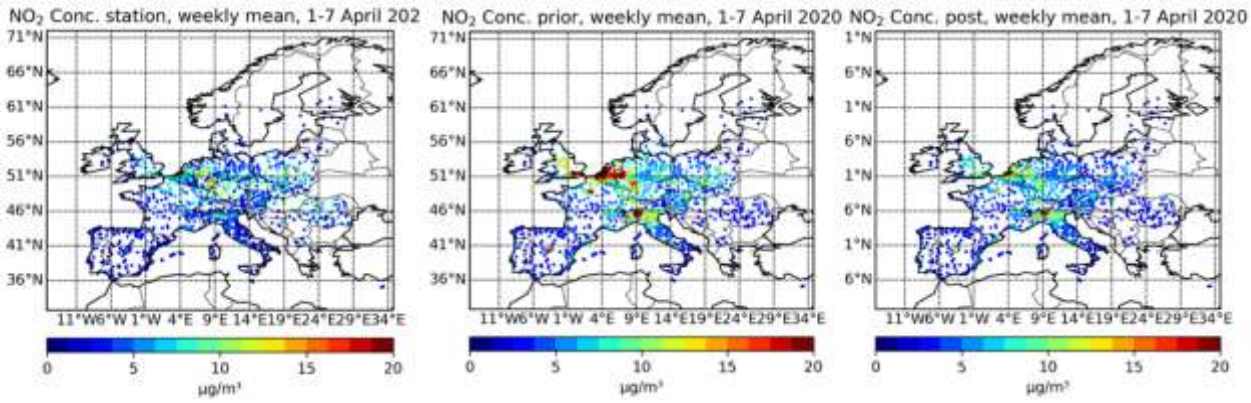


Figure 60 Weekly mean of NO_2 surface concentrations from stations (left), and prior (middle) and posterior simulations (right).

A total of 1927 stations are stated in the figures. The stations are rather well distributed all over Europe, with a larger number of stations in western and middle Europe, such as in Italy, Germany and Belgium/Netherlands. Averaged surface NO_2 concentrations calculated from hourly measurements from the stations and from prior and posterior simulations are 4.9 µg/m^3, 7.9 µg/m^3 and 5.9 µg/m^3, respectively. NO_2 concentrations simulated with the prior emissions largely overestimate measured surface NO_2 concentrations (about 60% of bias, Table 12). This overestimation drops to 21 % when posterior emissions are used to simulate NO_2 concentrations. The RMSE is also improved when posterior emissions are used. Nevertheless, we did not observe any improvement in the correlation, as this might be understandable because daily corrections may not influence hourly correlations. A comparison showing the differences station by station between the prior simulations and stations and the posterior simulations and stations is presented in Figure 61. The overestimation of NO_2 concentrations is mainly localized in the most urbanized and industrialized regions: Belgium, the Netherlands, the Ruhr in Germany, Paris, north Italy, and Madrid. This overestimation is reduced when using the posterior emissions. On the other hand, an underestimation of the measured NO_2 surface concentrations by the simulations is observed mainly in the eastern part of Europe and slightly reinforced when posterior emissions are used. This may suggest that the reduction of emissions prescribed by TROPOMI could also be attributed to a residual negative bias in the data.

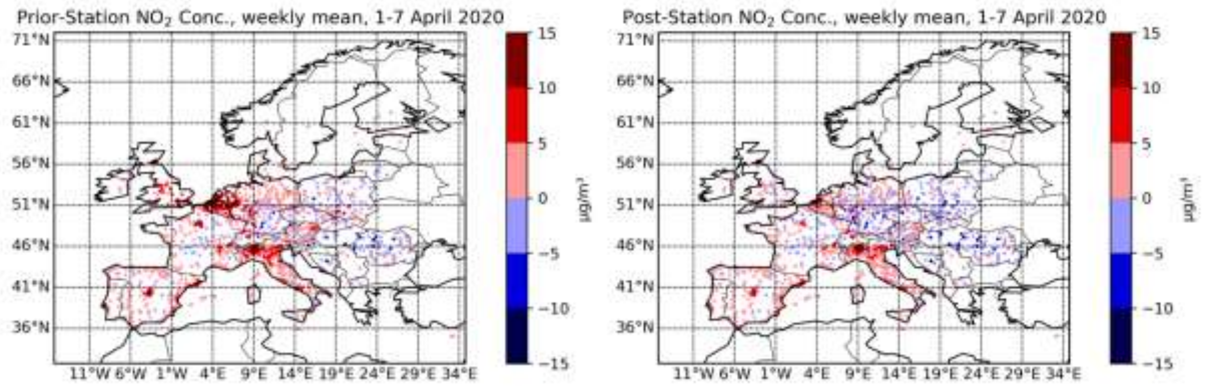
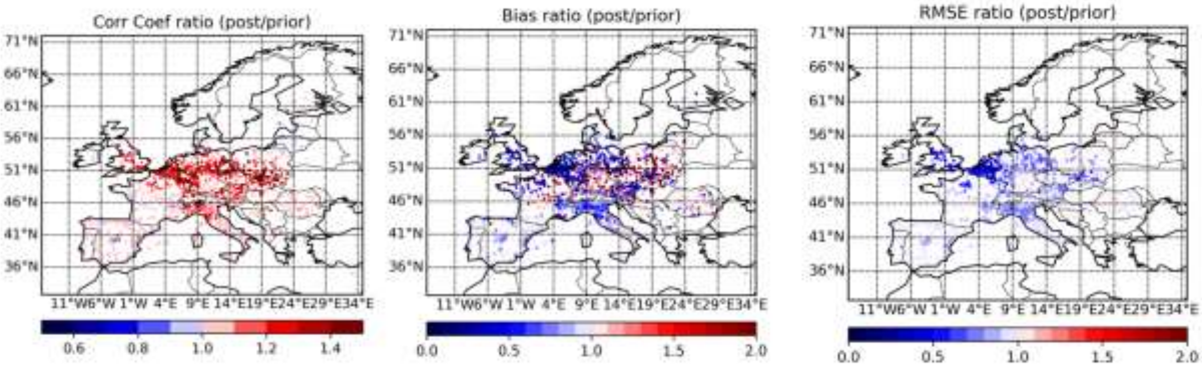


Figure 61 Differences between prior simulations and stations (left) and between posterior simulations and stations (right).

Table 12 Statistics for the comparison between simulated and measured NO_2 hourly surface concentrations for 1-7 April 2020. Mean = annual mean, SD = standard deviation, RMSE = root mean square error are in µg/m^3. r = Pearson correlation coefficient. NMB = normalized mean bias in % (Formulas used here are given in the Annexes A and D)

Station		Prior					Posterior				
Mean	SD	Mean	SD	*r*	NMB	RMSE	Mean	SD	*r*	NMB	RMSE
4.9	6.7	7.9	9.8	0.39	60.7	10	5.9	7.1	0.38	21.1	7.7

To push further the analysis, we also calculated the bias, RMSE and correlation ratios station by station as we did for the Chinese study (Annex E : Ratio Formulas). Figure 62 shows the ratios of these statistical indicators calculated for each station.

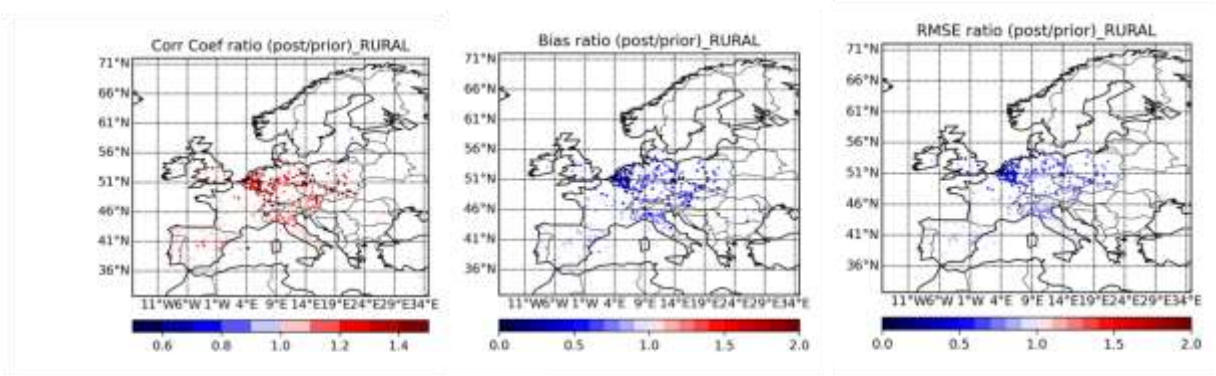

Corr Coef ratio (post/prior)

Bias ratio (post/prior)

RMSE ratio (post/prior)

Figure 62 Ratios of the statistical indicators: correlation coefficient (left), mean bias (middle) and RMSE (right) of posterior and prior per station. (To recall, correlation ratios bigger than one and bias and RMSE rations smaller than one meant to be an "improvement")

According to the figure, posterior simulations show good improvement over Benelux, the UK, Ile-de-France and Northern Italy (with increased correlation coefficient ratio and decreased mean bias and RMSE ratio). The correlation and the RMSE are rarely or very slightly degraded with the posterior. On the contrary, the biases at the stations are more largely degraded in the south of Germany and Poland. To analyze if some specific features could arise from these results, the comparison has also been made, separating the type of stations (rural and urban). Figures 63 and 64 show the ratios of the different statistical indicators separately for the rural and urban stations. The simulations of the NO_2 concentrations at rural stations are improved systematically when posterior emissions are used. On the other hand, the simulations at urban stations show both improvement and degradation of the biases. The improvements seem to occur mostly in large cities or industrial areas.


Corr Coef ratio (post/prior)_RURAL

Bias ratio (post/prior)_RURAL

RMSE ratio (post/prior)_RURAL

Figure 63 Ratios of the statistical indicators (correlation coefficient (left), bias (middle), RMSE (right)) for rural type of stations

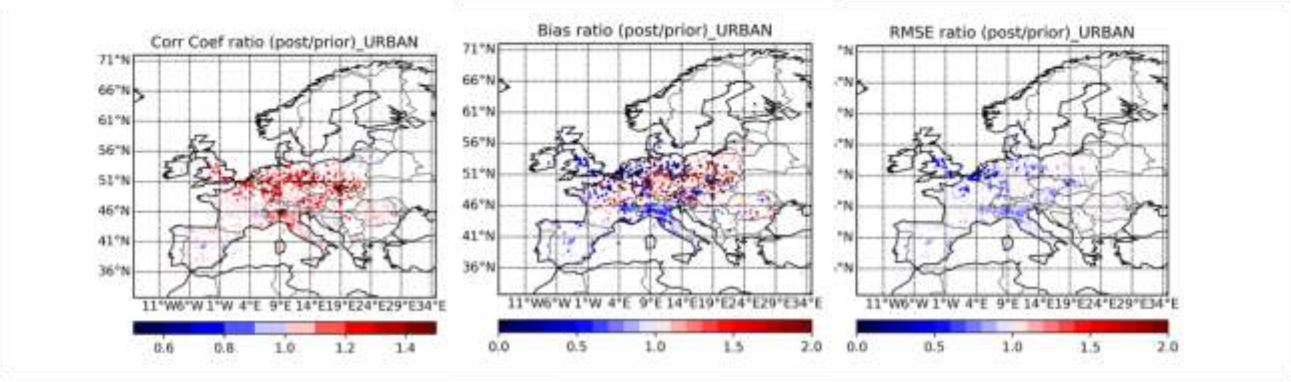


Figure 64 Ratios of the statistical indicators (correlation coefficient (left), bias (middle), RMSE (right)) for urban type of stations

These results suggest that when estimated emissions from CIF-CHIMERE system constrained by TROPOMI PAL v2.3 observations are used in a CTM, it brings simulated NO_2 concentrations closer to the “truth” given by the surface measurements and then we could state that the estimated NO_x emissions are closer to the “truth”.

4.7 Preliminary emissions for April 2019 and open questions

As mentioned previously, the reduction observed between the prior and the estimated emissions during the first week of April corresponds to a potential reduction between 2016 (prior) and 2020 (posterior). It likely includes a reduction due to the restrictions applied during the COVID-19 lockdown but also continuous reductions of NO_x emissions due to the mitigation. Residual negative biases in the TROPOMI product used for the inversion may also contribute to this prescribed reduction. To do a proper evaluation and assessment of the emission reductions induced by lockdowns, comparing with emissions inversed the year before is likely more relevant. To do so, we applied the CIF-CHIMERE to inverse emissions during the same period in April 2019 using TROPOMI observations. Unfortunately, technical issues on the CIF-CHIMERE after changes in the super calculator have limited these inversions to only two days and not the entire week as planned initially. The TROPOMI observations available for those two days are displayed in Figure 65. If the sampling of April 1st, 2019 covered, at least partly, similar regions to the ones mostly sampled during the first week of April 2020 (North of France, parts of Belgium and Germany, and North of Italy), the sampling of April 2nd, 2019 covers only the eastern part of Europe, quite differently from the regions sampled in 2020.

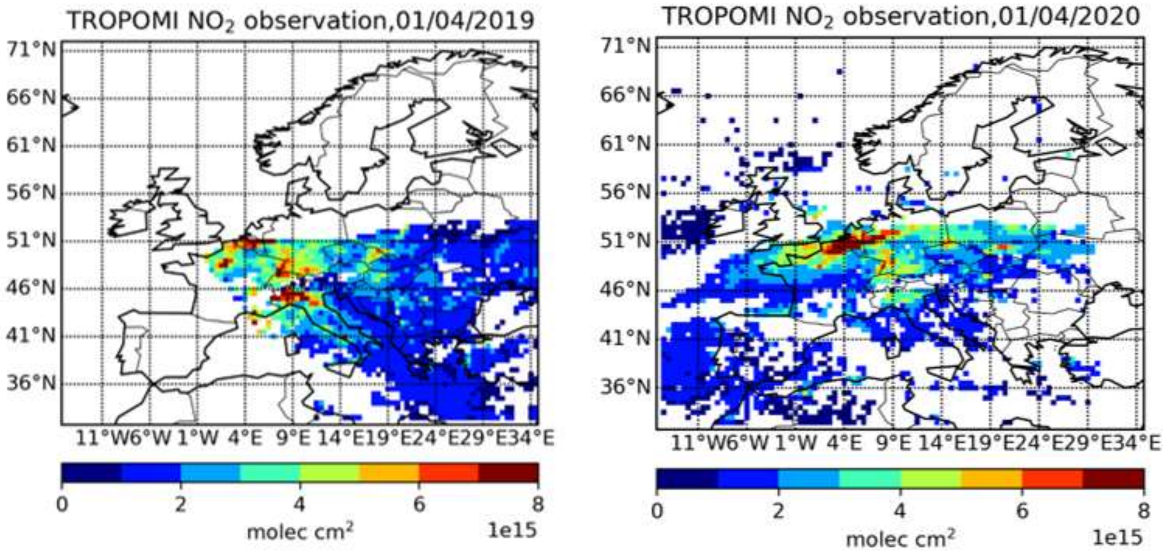


Figure 65 TROPOMI NO$_2$ observations for 1st of April in 2019 (left) and in 2020 (right)

The emissions estimates based on these observations and compared to the same prior emissions than previously (2016) are shown in Figure 66 for the two days. Interestingly, one of the main features, which comes out of this comparison, is that the inversion also led to a reduction of emissions in 2019 compared to the prior. The prescribed reduction is of the same order of magnitude compared to 2020, especially in the northern part of western Europe (the UK, Benelux, the Netherlands) on April 1st, 2019 when the TROPOMI sampling is close to the one of 2020. Compared with April 2nd, 2019 is less meaningful due to the difference in sampling. On April 1st, 2019, the inversion system increased the emissions in the south of Germany and North of Italy, and a feature also presents for some days in the first week of April 2020. The observed reduction of emissions between 2019 test days and prior emissions representing 2016 are much larger than the expected decrease of emissions stated from the continuous NO$_x$ emissions regulations applied in Europe during this period (-15% decrease in Germany, 10 % decrease in Italy in 2019 compared to 2016 (European Union Emission Inventory Report 2021). This might suggest that a residual negative bias remains in the last TROPOMI product used for this study and affect our inversions.

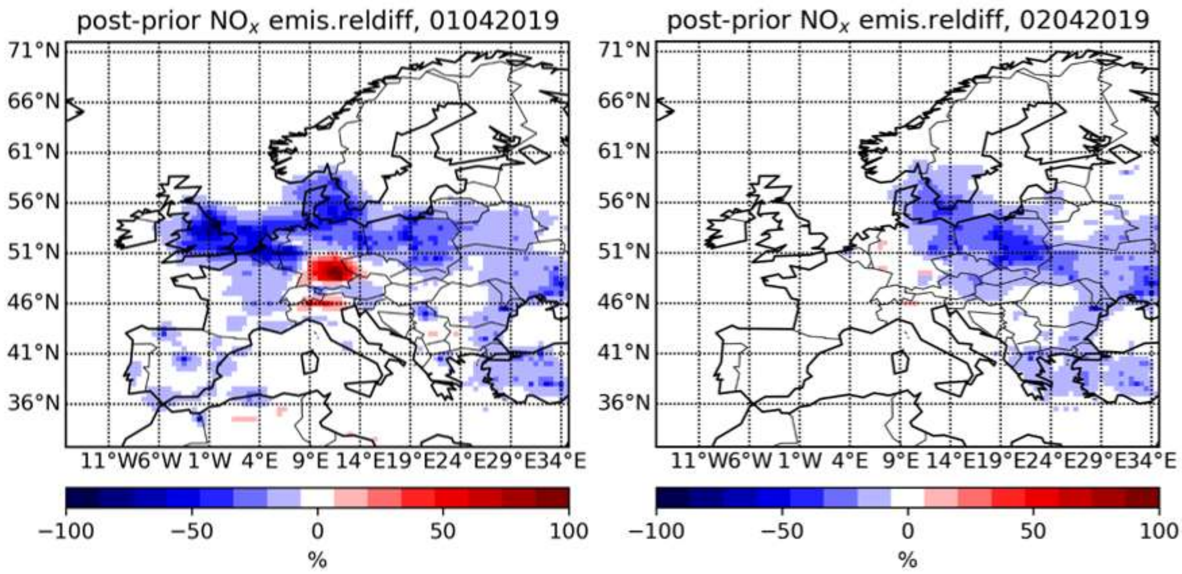


Figure 66 Increments in NOx emissions for 01/04/2019 (left), 02/04/2019 (right)

We also compare the emissions estimates of April 1st and 2nd, 2019 to the corresponding emissions of 2020 (Figure 67). Suppose one assumes that these days are representative of the general situation and that we can compare April 1st, 2019 and 2020 in the regions where the sampling is similar (note that both days are weekdays). In that case, reductions of about 40%, and 30-50% for Ile-de-France and South of Germany and North of Italy, respectively, are deduced from the comparison and might be attributed to lockdown restrictions. Guevara et al. 2021 estimated a 50 % reduction in the total budget of NO$_x$ emissions in Italy and France for the period March 23-April 26, 2021, which is consistent with our estimate. For Germany, it is lower than our estimate of 25%.

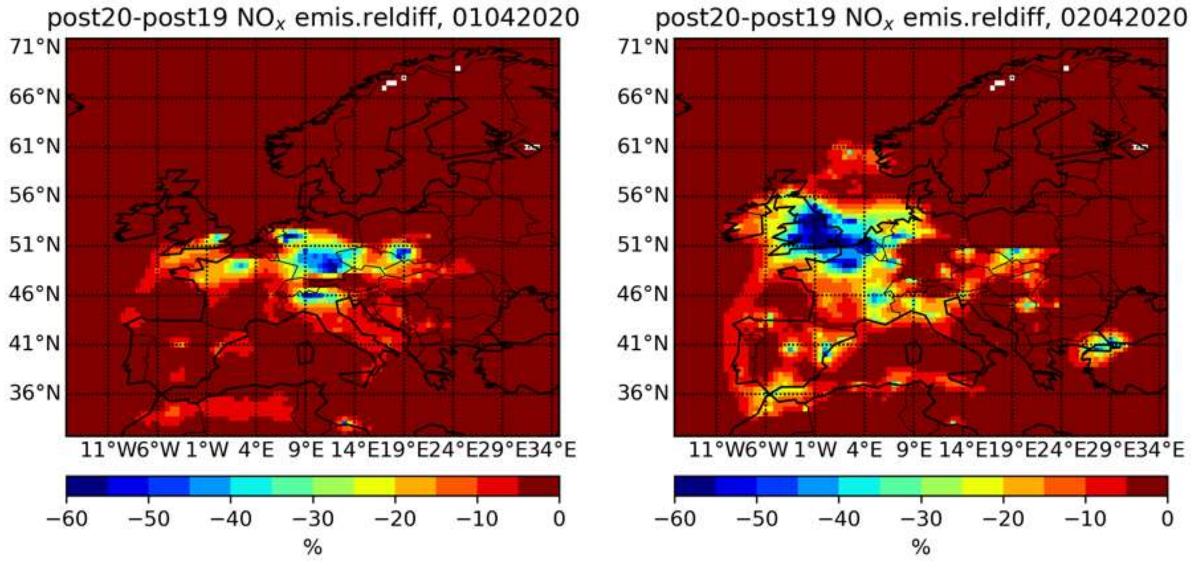


Figure 67 Differences between two estimated emissions for 2020 and 2019 for the days 1st (left) and 2nd (right) of April

Note that the reduction observed between 2020 and 2019 emissions for April 2nd cannot be exploited because the TROPOMI sampling for that two days is very different. The differences observed in Figure 67 (right panel) over Eastern Europe represent only the difference between 2020 and the prior emissions and not between 2019 and 2020, as no emissions corrections were derived from the inversion for that day in this region due to sampling issue (Figure 66, right panel).

4.8 First tests of inversion at high resolution over France

As one of the objectives of the ARGONAUT project and of this PhD work was to set up inversions at high resolution over France, we tested the estimation of French emissions with a 10 km resolution inversion to take advantage of the high resolution TROPOMI observations using the PAL v2.3 product. For these inversions, we encountered the same problem of convergence of the system and using a logarithmic scale for the control vector was necessary. Due to the computational duration of the inversion (almost one day per inversion) and the technical issue encountered at the TGCC super calculator, only a few days of inversions have been performed with the configuration given in Section 4.4. Figure 68 shows an example of one test for the day of 01/04/2020. The gain in the spatial representation of the emission corrections is well visible in the figure. The correction is localized mainly along the main roads and the urban centers. Details on the changes in emissions on the ship tracks are also available. These preliminary results look promising. Analyzing the inversion on longer periods with a detailed evaluation at the station will be very interesting, especially to see if the high resolution allows improvements for both urban and rural stations.

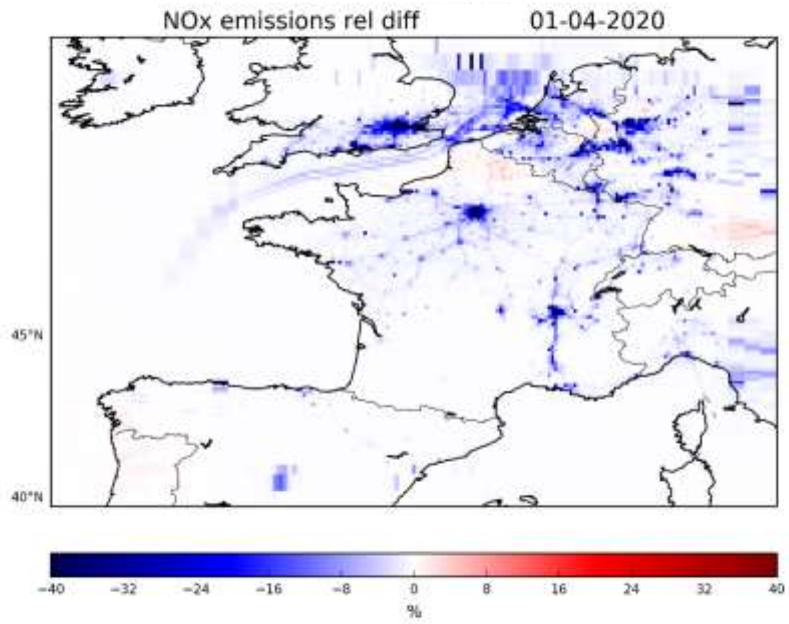


Figure 68 Increments in NO$_x$ emissions over France for 01/04/2020

General Conclusions and Perspectives

In this study, we examined the capabilities of the newly developed inversion system, CIF, coupled with the chemistry transport model, CHIMERE and its adjoint using observations from different satellites at different resolutions for various urbanized regions in the Northern Hemisphere. In particular, we focused on estimating NO_x emissions to test the 4D-Var mode of the CIF-CHIMERE system. NO_x are primary air pollutants involved in the production of secondary pollutants such as ozone and secondary aerosols. Governments or policymakers have started introducing regulations to control and mitigate them during the last decades, with successful reductions in Europe (-60 % since 1990) and China (-10 to -20 % since 2010). One key issue is to monitor the emissions and their changes, independently from the official declarations of the emitters. Satellite observations coupled with inverse modeling approaches are one of the most promising tools to achieve this monitoring. Indeed, the strong development of satellite observations with instruments providing highly resolved NO_2 tropospheric columns with increased confidence in the data at a global scale once a day, such as TROPOMI, offers new opportunities to monitor NO_x emissions at a high spatial and temporal resolution when used in the inverse modeling systems. Therefore, one of the objectives of this PhD work was to test and prepare the daily inversion of NO_x emissions using the new CIF-CHIMERE and TROPOMI observations.

To start with, we set up the configuration of the new CIF-CHIMERE system and tested some of its parameters that could be critical for our inversion applications. We performed a baseline test with the configuration adapted from literature uses or based on expert knowledge to inverse NO_x emissions. We used as emission constraints NO_2 columns from the OMI satellite over the domain of Europe at medium resolution (50 km). In each test, one parameter (from optimization and variational parameters) of this baseline test was changed. Different parameters related to the optimization were found to have less importance while estimating the emissions than variational parameters. Indeed, the main drivers of the inversion are the error statistics considered for the inversion, especially the errors on the background emissions and their correlations. The different tests performed showed that the new CIF-CHIMERE system responded well to error statistics prescribed in the background covariance matrix. These tests allow us to define a configuration of use for the inversion system as well as a suited range of variations of the different parameters for the appropriate use of the system. This is helpful to define adapted configuration when other domains are considered. In addition, during the PhD, new functionality was implemented in the CIF-CHIMERE system by its developing team to overcome convergence problems we encountered when the variables were constrained and the constraints covered several orders of magnitude. We now have the possibility to use a logarithmic scale for the control vector, which helps the optimization and the convergence of the system. We tested the influence of using the logarithmic scale rather than the linear scale for the inversion. The test was made on only one day, but the results suggest that the patterns of correction of the emissions are not changed by changing the scale. The magnitude of the corrections is consistent for the largest emissions but can be significantly different in the background regions. This should not affect that much the result in terms of budget or air quality simulation performances. Nevertheless, this question should be addressed for a longer period in the future to be more representative.

Next, the new CIF-CHIMERE system is used to estimate Chinese NO_x emissions for the years 2015 and 2019. The Chinese government first introduced regulations on NO_x emissions in its 12th Five-Year Plan in 2011, followed by new ones in the 13th Five-Year Plan in 2016. Our target years were chosen as the last years of the two five-year plans, excluding 2020, due to the COVID-19 pandemic. To constrain Chinese NO_x emissions, NO_2 column data from OMI satellite observations (13 km × 25 km) are used at 50 km resolution. Due to the low daily OMI coverage, monthly corrections of the NO_x emissions were calculated. It is found that using these NO_x emission estimates to simulate NO_2 surface concentrations slightly improves the comparison with ground-based measurements. The estimated annual budgets were in good agreement (within the uncertainties) with independent inventories such as MEIC and DECSO in 2015. Nevertheless, this agreement varies depending on the Chinese regions and the year period. The results suggest that OMI observations provide a better constraint to the NO_x emissions in the polluted areas compared to the background areas. Polluted areas corresponding to the main urbanized regions in China satisfy the target of the 12th Five-Year Plan (-10 %). On the other hand, the evolution of estimated emissions from 2015 to 2019 remains small, suggesting that reduction targets of the 13th Five-Year Plan were not fulfilled. This statement is not supported by recent literature and needs further investigation.

This PhD work is a part of the ARGONAUT project. One of the aims of the project is to estimate French NO_x emissions at high-resolution (10 km) using high-resolution satellite observations such as TROPOMI (5.5 km × 3.5 km). As OMI's moderate resolution is not suitable for inversions at 10 km over France, we first adapted the CIF-CHIMERE system to 50 km resolution covering a larger area, Europe, to be able to compare inversion performances of TROPOMI and OMI. Our results show the benefit of the high resolution and improved instrumental performances of TROPOMI to provide a larger daily coverage and improve the daily inversion capabilities. Due to the computation time, we first performed inversions at a moderate resolution over the European domain to evaluate the sudden changes in NO_x emissions during the COVID-19 lockdown period. We focus on the first week of April 2020 and compare the new emission estimates to the prior emissions, which are representative of the year 2016. The evaluation of the inversion results by comparing the NO_2 concentrations simulated with these new emissions to surface concentration measurements from the air quality network shows that the new emission estimates improve the comparison statistics compared to the prior emissions. The bias is reduced from 60 % to 20 %, and the RMSE is decreased by about 25 %. We showed a systematic improvement in the comparison for rural stations. For urban stations, the improvement is larger for city clusters (e.g., Ile-de-France, North of Italy), but some degradation of the bias is observed in the south of Germany and in Poland mainly.

In terms of emissions changes compared to the prior for this first week of April 2020, a reduction of emissions is seen in the northwestern European countries with reductions up to 67 % at the grid-cell scale or around 35-45 % for countries such as Belgium, the Netherlands, UK, Germany or Poland. Countries in the South (France, Italy and Spain) show a minor reduction in their emissions during this week. The difference was attributed to the sampling of the observations, which is more constant and homogeneous from day to day in the north than in the south, instead to a difference in applying restrictions during the lockdown.

The methodology to assess the COVID-19 lockdown impact is usually to compare 2020 with 2019. However, due to the technical issue with the super calculator used to perform the inversion, only a few days have been inversed for April 2019. Nevertheless, the preliminary comparison we did for these days would suggest a reduction of 40 %, and 30-50 % for Ile-de-France, South of Germany and North of Italy respectively, but this needs to be confirmed on a larger time period to be representative. We also made the first inversion attempts at high resolution for one day (April 1st, 2020) over France. The results look promising to assess and monitor emissions at high spatial resolution.

In general, except for China, our work has been done on limited time periods and days. An extension of the studied periods is needed to get results that are more representative and consolidated. Nevertheless, the results of this PhD work stress the promising possibilities offered by the new CIF-CHIMERE system and the TROPOMI data to monitor NO_x emissions and their changes from moderate to high resolution. This work also raised some questions and perspectives, which could be addressed in the future.

From a methodological point of view, since the CIF-CHIMERE inversion system does not take aerosol into account (the adjoint of the CHIMERE- aerosol module is not coded in the adjoint model), limitations in the simulated NO_2 concentrations may arise due to the photochemical role of aerosols. Moreover, this work stresses several issues. One of them is the sampling impact of the observations on the inversion and the achievable temporal resolution. Despite improvements with TROPOMI, cloud coverage is still a limiting factor in some situations to achieve meaningful daily inversion. This raises the question of the temporal resolution to which the inversions can be done. Statistical evaluation of the coverage of TROPOMI per region and grid cell could help define this temporal resolution. Also, indicators of the coverage, which could represent the relevance and the confidence to the inversion, could be provided with the inversion results at higher resolution to allow a proper use of the results and avoid misinterpretation. Technically, this also raises the question of the best way to perform inversion at the suited resolution. If a weekly resolution is recommended, for example, is it more suitable to perform a 4D-Var inversion increasing the assimilation window to the week considering the individual observation day by day, or is it more suitable to aggregate the observations and treat them as an average over the week?

Another methodological issue that needs to be addressed in the future with the CIF-CHIMERE is the separation of the anthropogenic and biogenic emissions or the emissions by sectors. This implies technical developments in the CIF to adapt the control vector. Note that to do so, it is necessary to evaluate if enough information is given by the observations during the inversion. The high resolution of TROPOMI could help with this task. Also, co-assimilate pollutants having common and specific emission sources could help go further in our capabilities to separate the sources. This also leads to the creation of more complex **B** matrices instead of simple homogeneous and isotropic **B** matrices, for example, correlations can be selected by the type of sources.

From a scientific point of view, our study of China was a first step and raised the question about the effectiveness of reducing the emissions during the 13th Five-Year Plan. Extending of study to the entire decade would be interesting to evaluate if our findings are specific to 2019. Using TROPOMI for recent years would also be interesting, but we need to ensure the consistency between the OMI and TROPOMI observations to be able to calculate trends. In Europe and France (and also other parts of the world), assessing the impacts of the pandemic on the NO_x emissions at high resolution, including all the lockdowns and restrictions, would be very interesting too.

Note that one important concern in achieving these different developments and studies is to ensure that the observations are not biased; however, TROPOMI data still consists of a residual negative bias. Therefore, debiasing techniques could be considered to be applied before the assimilation of the data in the inversion system. Furthermore, in addition to using observations from low-orbit satellites in the inversion system, observations from geostationary satellite instruments can also be used to provide high temporal resolution information to the inversion system (e.g. TEMPO, Sentinel-4, GEMS).

Annexes

Annex A : Statistical formulas

The statistical formulas are recalled for the estimation and relation of the background and observations errors. Let *u* and *v* be a scalar variable with n dimension:

- Average:

$$\boldsymbol{\bar{u} = \frac{1}{n}\sum_{i=1}^{n} u_i}$$

- Variance

$$\boldsymbol{var(u,v) = \frac{1}{n}\sum_{i=1}^{n} (u_i - \bar{u})^2}$$

- Standard Deviation:

$$\boldsymbol{\sigma(u) = \sqrt{var(u)}}$$

- Covariance:

$$\boldsymbol{cov(u,v) = \frac{1}{n}\sum_{i=1}^{n} (u_i - \bar{u})\,(v_i - \bar{v})}$$

- Correlation:

$$\boldsymbol{\rho(u,v) = \frac{cov(u,v)}{\sigma(u)\sigma(v)}}$$

Annex B : 4D-Var cost and the gradient evolution

Let be the J cost function,

$$J(\boldsymbol{x}) = \underbrace{\frac{1}{2}(\boldsymbol{x}-\boldsymbol{x}^{\boldsymbol{b}})^{\mathrm{T}}\mathbf{B}^{-1}(\boldsymbol{x}-\boldsymbol{x}^{\mathbf{b}})}_{J^{b}(x)} + \underbrace{\frac{1}{2}\sum_{\mathrm{i}=0}^{\mathrm{n}}(\boldsymbol{y}_i-\mathbf{H}_i(\boldsymbol{x}_i))^{\mathrm{T}}\mathbf{R}_i^{-1}(\mathbf{y}_i-\mathbf{H}_i(\boldsymbol{x}_i))}_{J^{o}(x)} \qquad \text{(Eq.21)}$$

where the state $\boldsymbol{x}$ at time i is obtained iteratively by the linear model $\mathbf{M}$,

$$\boldsymbol{x}_i = \mathbf{M}_{0\to i}(\boldsymbol{x})$$

$$= \mathbf{M}_{i-1,\mathrm{i}}\left(\mathbf{M}_{i-2,i-1}\left(\dots\mathbf{M}_{1,2}\left(\mathbf{M}_{0,1}(\boldsymbol{x})\right)\right)\right)$$

$$= \mathbf{M}_{\mathrm{i}}\left(\mathbf{M}_{\mathrm{i}-1}\left(\dots\left(\mathbf{M}_2\left(\mathbf{M}_1(\boldsymbol{x})\right)\right)\right)\right)$$

$$\overset{\text{(to simplify)}}{=} \mathbf{M}_i\mathbf{M}_{i-1}\dots\mathbf{M}_2\mathbf{M}_1(\boldsymbol{x}).$$

Then the cost function J can be written as a function of the model iteration as follows,

$$\boldsymbol{J}(\boldsymbol{x}) = \boldsymbol{J}^{\mathbf{b}}(\boldsymbol{x}) + \frac{\mathbf{1}}{\mathbf{2}}\sum_{\mathbf{i}=\mathbf{0}}^{\mathbf{n}}(\mathbf{y}_i-\mathbf{H}_i\mathbf{M}_i\mathbf{M}_{i-1}\dots\mathbf{M}_2\mathbf{M}_1(\boldsymbol{x}))^{\mathbf{T}}\mathbf{R}_i^{-1}(\mathbf{y}_i-\mathbf{H}_i\mathbf{M}_i\mathbf{M}_{i-1}\dots\mathbf{M}_2\mathbf{M}_1(\boldsymbol{x})) \qquad \text{(Eq.B1)}$$

The minimization of Eq.B1 requires its gradient as follows:

$$\nabla J(\boldsymbol{x}) = \mathbf{B}^{-1}\left(\boldsymbol{x}-\boldsymbol{x}^{\mathbf{b}}\right) - \sum_{\mathrm{i}=0}^{\mathrm{n}}\mathbf{M}_1^{\mathrm{T}}\mathbf{M}_2^{\mathrm{T}}\dots\mathbf{M}_{i-1}^{\mathrm{T}}\mathbf{M}_i^{\mathrm{T}}\mathbf{H}_i^{\mathrm{T}}\mathbf{R}_i^{-1}(\boldsymbol{y}_i-\mathbf{H}_i\mathbf{M}_i\mathbf{M}_{i-1}\dots\mathbf{M}_1(\boldsymbol{x}))$$

where innovation vector $\mathbf{d}_i = \boldsymbol{y}_i - \mathbf{H}_i\mathbf{M}_i\mathbf{M}_{i-1}\dots\mathbf{M}_1(\boldsymbol{x})$. Then the gradient of the $J^{\boldsymbol{o}}(\boldsymbol{x})$,

$$-\nabla J^{\mathrm{o}}(\boldsymbol{x}) = \sum_{\mathrm{i=0}}^{\mathrm{n}} \mathbf{M}_1^{\mathrm{T}}\mathbf{M}_2^{\mathrm{T}} \ldots \mathbf{M}_{i-1}^{\mathrm{T}}\mathbf{M}_i^{\mathrm{T}}\mathbf{H}_i^{\mathrm{T}}\mathbf{R}_i^{-1}\mathbf{d}_i$$

$$= \mathbf{H}_0^{\mathrm{T}}\mathbf{R}_0^{-1}\mathbf{d}_0 + \mathbf{M}_1^{\mathrm{T}}\mathbf{H}_1^{\mathrm{T}}\mathbf{R}_1^{-1}\mathbf{d}_1 + \mathbf{M}_1^{\mathrm{T}}\mathbf{M}_2^{\mathrm{T}}\mathbf{H}_2^{\mathrm{T}}\mathbf{R}_2^{-1}\mathbf{d}_2 + \cdots + \mathbf{M}_1^{\mathrm{T}} \ldots \mathbf{M}_{n-1}^{\mathrm{T}}\mathbf{M}_n^{\mathrm{T}}\mathbf{H}_n^{\mathrm{T}}\mathbf{R}_n^{-1}\mathbf{d}_n$$

$$= \mathbf{H}_0^{\mathrm{T}}\mathbf{R}_0^{-1}\mathbf{d}_0 + \mathbf{M}_1^{\mathrm{T}}\left(\mathbf{H}_1^{\mathrm{T}}\mathbf{R}_1^{-1}\mathbf{d}_1 + \mathbf{M}_2^{\mathrm{T}}(\mathbf{H}_2^{\mathrm{T}}\mathbf{R}_2^{-1}\mathbf{d}_2 + \cdots + \mathbf{M}_n^{\mathrm{T}}\mathbf{H}_n^{\mathrm{T}}\mathbf{R}_n^{-1}\mathbf{d}_n)\right).$$

So if we put it all together, we obtain the $\nabla J(\boldsymbol{x})$ as follows,

$$\nabla J(\boldsymbol{x}) = \mathbf{B}^{-1}(\boldsymbol{x} - \boldsymbol{x}^{\mathbf{b}}) - \left(\mathbf{H}_0^{\mathrm{T}}\mathbf{R}_0^{-1}\mathbf{d}_0 + \mathbf{M}_1^{\mathrm{T}}\left(\mathbf{H}_1^{\mathrm{T}}\mathbf{R}_1^{-1}\mathbf{d}_1 + \mathbf{M}_2^{\mathrm{T}}(\mathbf{H}_2^{\mathrm{T}}\mathbf{R}_2^{-1}\mathbf{d}_2 + \cdots + \mathbf{M}_n^{\mathrm{T}}\mathbf{H}_n^{\mathrm{T}}\mathbf{R}_n^{-1}\mathbf{d}_n)\right)\right) \qquad \text{(Eq.B2)}$$

In the main text, Eq.B2 is the same as Eq.23 in which $\mathbf{d}_i$ is considered.

This factorization makes calculation of J followed by ∇J defined by;

$$\boldsymbol{x}_i^* = \mathbf{M}_{i+1}^{\mathrm{T}}\boldsymbol{x}_{i+1}^* + \mathbf{H}_i^{\mathrm{T}}\mathbf{R}_i^{-1}\mathbf{d}_i\,, \quad i = n-1, 0 \qquad \text{(Eq.B3)}$$

$$\boldsymbol{x}_n^* = \mathbf{H}_n^{\mathrm{T}}\mathbf{R}_n^{-1}\mathbf{d}_n$$

Annex C : Non-linear operators

In the case of the non-linearity of the model $\boldsymbol{\mathcal{M}}$ and observation operator $\boldsymbol{\mathcal{H}}$, thanks to the tangent linear hypothesis (TLH) and adjoint models (Lewis and Derber, 1985) these non-linear operators can be replaced with their tangent linear operators by using the below Equations. This procedure force cost function *J* to be a convex function and 4D-Var algorithm can be used to calculate the result approximately.

$$\boldsymbol{\mathcal{M}}_{0\rightarrow \mathrm{i}}(\mathbf{x}_0) - \boldsymbol{\mathcal{M}}_{0\rightarrow \mathrm{i}}\left(\mathbf{x}_0^{\mathbf{b}}\right) \approx \mathbf{M}_{0\rightarrow \mathrm{i}}(\mathbf{x}_0 - \mathbf{x}^{\mathbf{b}})$$
$$\boldsymbol{\mathcal{H}}_{\mathrm{i}}(\mathbf{x}_{\mathrm{i}}) - \boldsymbol{\mathcal{H}}_{\mathrm{i}}\left(\mathbf{x}_{\mathrm{i}}^{\mathbf{b}}\right) \approx \mathbf{H}_{\mathrm{i}}(\mathbf{x}_{\mathrm{i}} - \mathbf{x}_{\mathrm{i}}^{\mathbf{b}})$$

Annex D : Statistical indicators

Pearson's Correlation Coefficient (*r*), Normalized Mean Bias (NMB) and Root Mean Square Error (RMSE) Formulas from top to down (*n* = number of available hourly measurements in the relevant year, $\overline{simulation}$ and $\overline{station}$ are the averaged values of simulated concentrations and stations, respectively.)

$$r = \frac{\sum_{i}^{n}(simulation_i - \overline{simulation})(station_i - \overline{station})}{\sqrt{\sum_{i}^{n}(simulation_i - \overline{simulation})^2 (station_i - \overline{station})^2}}$$

$$NMB(\%) = \frac{\sum_{i}^{n}(simulation_i - station_i)}{\sum_{i}^{n} station_i} x100$$

$$RMSE\left(\mu g/m^3\right) = \sqrt{\frac{1}{n}\sum_{1}^{n}(simulation_i - station_i)^2}$$

Annex E : Ratio formulas

r ratios, MB ratios and RMSE ratios formulas are as follows:

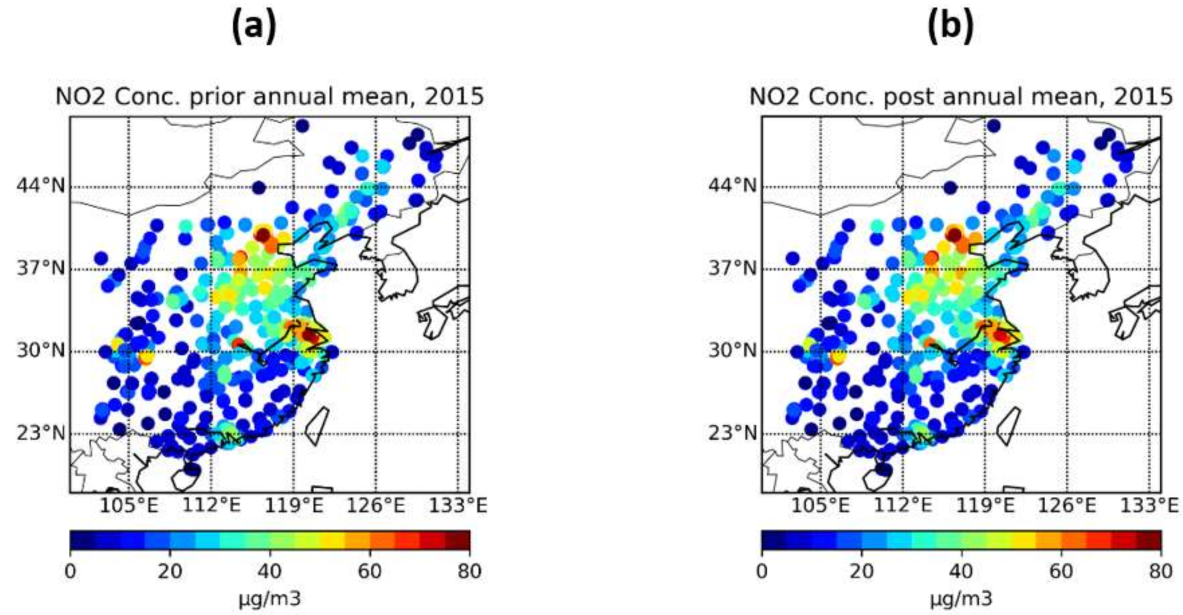
$$\boldsymbol{r_{ratio} = \frac{r_{prior}}{r_{post}}}$$

$$\boldsymbol{MB_{ratio} = \frac{MB_{prior}}{MB_{post}}}$$

$$\boldsymbol{RMSE_{ratio} = \frac{RMSE_{prior}}{RMSE_{post}}}$$

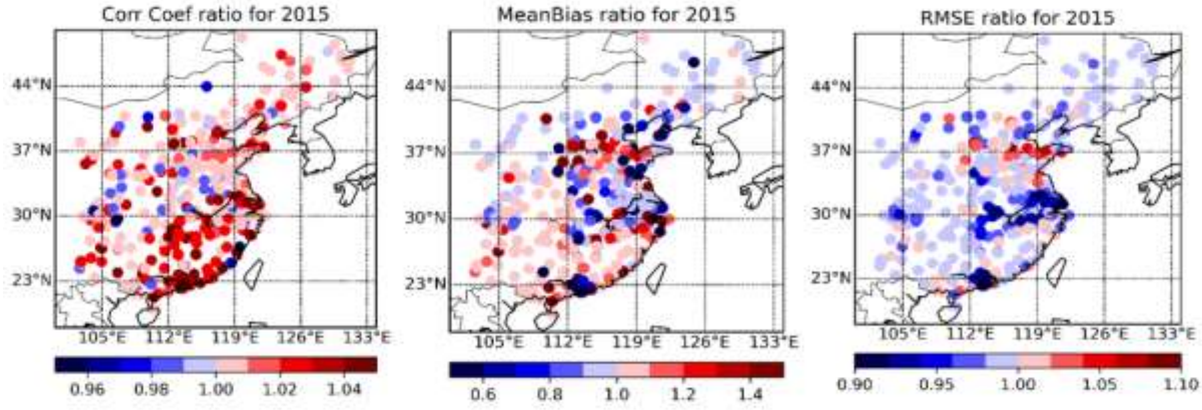
Annex F : Annual concentrations 2015

Spatial distribution of annual mean surface NO_2 concentrations: (a) CHIMERE using prior emissions (Conc_Prior15); (b) CHIMERE using posterior emissions (Conc_Post15)



Annex G : Statistical indicators’ ratios 2015

Ratios of the Conc_Prior and Conc_Post’s statistical indicators: Correlation Coefficient, Mean Bias and RMSE, respectively, for (a-c) 2015.



(a) (b) (c)

Annex H : In situ comparison

Table H : NO$_2$ Annual Mean surface concentrations and statistics of the 15 selected locations for (top) 2015 and, (bottom) 2019.

Location	**Annual Mean (μg/m^3)**				***r***			**NMB (%)**			**RMSE (μg/m^3)**		
	Station	**Prior**	**Post**	**MEIC**	**Prior**	**Post**	**MEIC**	**Prior**	**Post**	**MEIC**	**Prior**	**Post**	**MEIC**
Daqing	24.9	25.4	23.9	24.4	0.68	0.69	0.66	2.1	-3.9	-1.8	12.8	12.6	13.6
Shenyang	41.8	41	41.7	41.4	0.65	0.65	0.59	-1.9	-0.2	-0.9	21.3	21.4	23.1
Baotou	36.7	33.4	30.3	26.4	0.55	0.54	0.55	-9	-18	-28	22.2	23	23.9
Beijing-Tianjin (M*)	46.7	66.2	67	55.9	0.66	0.67	0.65	41.9	43.6	19.7	36	36.4	28.8
Dalian (P)	28.3	27.6	28.4	29.5	0.58	0.59	0.53	-2.7	0.4	4.1	17.6	17.5	19.1
Shijiazhuang (M*)	39.1	65	65.8	56.3	0.58	0.58	0.57	66	68.5	44.1	40.1	40.6	34
Shandong (M*)	42.3	52.9	53.4	41.4	0.65	0.66	0.63	25	26.4	-2	25.4	25.6	22.3
Henan (P, M)	38.8	52.7	50.7	38	0.59	0.59	0.59	35.8	30.8	-2.1	28.3	27	22.7
Anhui (P*, M*)	25	35.8	34.8	29.1	0.44	0.43	0.43	43.4	39.4	16.4	24.5	24.2	21.5
Yangtze River Delta (P, M)	37.9	62.5	59.7	50.7	0.58	0.6	0.58	65.2	57.8	33.8	37	34.5	28.5
Wuhan (P, M*)	47.1	69.5	64.9	49.8	0.62	0.63	0.57	47.7	37.9	5.8	34.5	31.4	26.7
Chengdu (P*)	43.2	49.4	47.9	62.9	0.5	0.49	0.54	14.3	11	45.8	24.8	24.4	31
Chongqing (P*, M)	38.5	59.1	56.8	50.9	0.4	0.39	0.41	53.4	47.6	32.3	34.1	32.8	27.5
Hunan (P)	33.5	46.4	44	49.2	0.54	0.55	0.55	38.8	31.6	47	24.2	22.5	26.3
Pearl River Delta (P)	30.1	33.9	30.5	35.3	0.54	0.55	0.59	12.5	1.3	17.3	20.6	19.5	20.8

Location	**Annual Mean (μg/m^3)**			***r***		**NMB (%)**		**RMSE (μg/m^3)**	
	Station	**Prior**	**Post**	**Prior**	**Post**	**Prior**	**Post**	**Prior**	**Post**
Daqing (P*)	19	23.7	22.1	0.72	0.71	24.6	16.3	12.2	11.7
Shenyang (P*)	33	41.5	41.3	0.74	0.73	25.7	25	19.6	19.5
Baotou	36.7	35.5	33.5	0.6	0.61	-3.2	-8.6	21.8	21.7
Beijing-Tianjin (P)	38	67.5	66.7	0.65	0.65	77.9	75.5	42.7	41.8
Dalian	21.3	27.5	28.1	0.68	0.68	29.2	31.9	15.6	15.9
Shijiazhuang (P)	29	67.8	66.7	0.62	0.63	134.2	130.3	47.5	46.5
Shandong (P)	38.3	57.2	55.6	0.73	0.74	49.2	45	28.7	27.6
Henan (P)	34	57.1	54	0.69	0.69	67.9	59	32.1	29.7
Anhui (P)	23.9	37.1	35.6	0.69	0.69	55.5	49.1	20.6	19.5
Yangtze River Delta (P)	34.4	62.3	58.4	0.62	0.65	81.2	69.6	38.3	34.3
Wuhan (P)	39.3	72.5	64.7	0.66	0.67	84.8	64.8	41	34.8
Chengdu (P)	39.5	54	53.2	0.6	0.6	36.8	34.8	26	25.5
Chongqing (P)	35.5	60.9	59.7	0.52	0.52	71.5	68.3	35.1	34.3
Hunan (P)	27	48.5	45.7	0.67	0.67	79.5	69.4	27.5	25.4
Pearl River Delta (P)	25.7	35.3	31.9	0.57	0.61	37.1	23.8	21.9	19.1

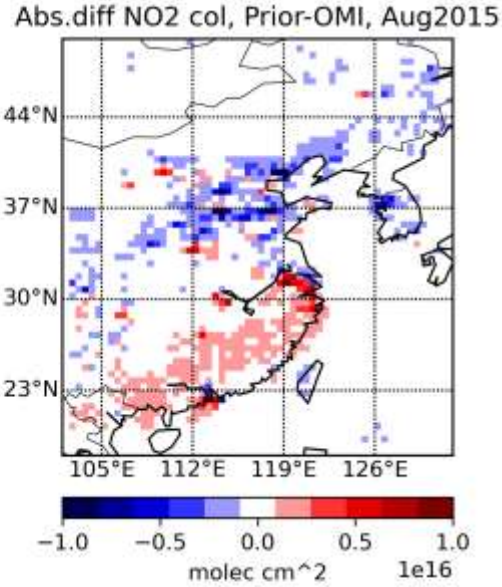
P : Posterior better than Prior ; P* : Posterior better than Prior when excluding *r*

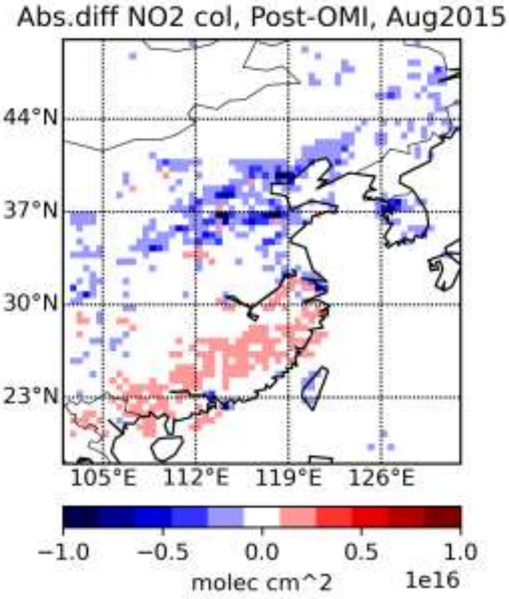
M : MEIC is better than Prior ; M* : MEIC is better than Prior when excluding *r*

(Note that; “better” means lower NMB and RMSE, and higher *r*)

Annex I : Innovation & Residue Map

NO_2 monthly mean columns differences between Prior and OMI observations (below) and Posterior and OMI observations (bottom) for 2015.





Annex J : Résumé substantiel en français

L'Organisation mondiale de la Santé estime que l'exposition à la pollution de l'air ambiant provoque 4,2 millions de décès prématurés dans le monde chaque année. Ce problème environnemental est apparu au 20ème siècle en raison d'une intense industrialisation et urbanisation, incitant les gouvernements à mettre en place des politiques pour réduire les émissions anthropiques de polluants atmosphériques. Depuis 1990, les pays industrialisés, en particulier en Europe, ont pris des initiatives pour réduire les émissions de polluants grâce à des stratégies d'atténuation. Les pays en développement, comme la Chine, ont également commencé à mettre en place des politiques d'atténuation plus récemment.

Les oxydes d'azote (NO_x) sont l'un des principaux polluants, causant des impacts sanitaires directs tels que des infections respiratoires, de l'asthme et des maladies pulmonaires chroniques, et endommageant les écosystèmes. Les NO_x sont des précurseurs d'autres polluants atmosphériques tels que l'ozone et les aérosols secondaires, contribuant également au changement climatique en tant que forceurs climatiques à courte durée de vie. Lorsqu'ils se dissolvent dans l'eau, ils créent des acides qui nuisent à la végétation, aux bâtiments et aux matériaux, et contribuent à l'acidification et à l'eutrophisation des écosystèmes. Par conséquent, les NO_x, en particulier le NO_2, sont des polluants hautement réglementés.

En Europe, selon l'Agence européenne pour l'environnement, les émissions de NO_x ont été réduites de près de 60 % entre 1990 et 2019, le transport routier et l'énergie étant les principaux secteurs émetteurs. En Chine, la production d'électricité et l'industrialisation, fortement dépendantes de la consommation de charbon, sont les principaux contributeurs aux émissions de NO_x, suivis des véhicules routiers. Un développement économique rapide a entraîné une augmentation significative des émissions de NO_x entre les années 90 et la fin des années 2000. Pour lutter contre ce problème, la Chine a mis en place des réductions d'émissions de NO_x avec des objectifs de réduction de 10 % en 2015 et de 15 % en 2020 dans leurs 12ème et 13ème plans quinquennaux gouvernementaux, respectivement.

Les capacités d'observation des polluants, en particulier les concentrations de NO_2, ont augmenté au cours des trois dernières décennies avec le développement de réseaux de mesure de la qualité de l'air. Dans le même temps, les observations par satellite des colonnes troposphériques de NO_2 ont montré des améliorations sans précédent : depuis l'instrument satellite GOME avec une résolution spatiale de quelques centaines de kilomètres au milieu des années 90 jusqu'à OMI et GOME-2 (résolution de quelques dizaines de kilomètres) au milieu des années 2000 et TROPOMI (résolution de quelques kilomètres) depuis 2018. Les capacités de modélisation atmosphérique ont également progressé, permettant aux modèles de transport chimique de simuler les distributions tridimensionnelles des concentrations de polluants à des échelles locales à globales avec une haute résolution spatiale et temporelle. Des systèmes opérationnels tels que le Service de surveillance de l'atmosphère Copernicus (CAMS pour Copernicus Atmospheric Monitoring Service en anglais) en Europe et la plateforme nationale Prev'Air en France, ont fourni des estimations des polluants atmosphériques à différentes échelles en exploitant la complémentarité de la modélisation et des observations et le développement de puissants systèmes d'assimilation de données.

Un problème clé est d'évaluer la réalité et l'efficacité des politiques de réduction des émissions. Les méthodes d'évaluation traditionnelles sont souvent basées sur des inventaires d'émissions officiels dérivés des déclarations d'activités des émetteurs et des données statistiques, appelées approche "bottom-up". Ces inventaires impliquent de fortes incertitudes et ne prennent pas en compte la variabilité spatiotemporelle élevée des flux d'émissions. En revanche, l'approche d'inversion atmosphérique, ou approche "top-down", utilise des observations atmosphériques disponibles, telles que des observations satellitaires, pour contraindre les émissions à travers la modélisation atmosphérique et un système d'inversion (assimilation de données). Cette approche estime les émissions tout en exploitant la variabilité spatiotemporelle élevée des observations satellitaires et la puissance de calcul avec une capacité de mise à jour rapide, apportant des informations complémentaires aux inventaires "bottom-up".

Au début de ce travail de thèse, le nouvel instrument satellitaire TROPOMI commençait tout juste à fournir des colonnes troposphériques de NO_2 à une résolution spatiale sans précédent de quelques kilomètres. Au niveau français, un nouveau système d'inversion était en cours de développement, le Community Inversion Framework (CIF) (Berchet et al. 2021) basé sur une version précédente moins polyvalente, PYVAR (Fortems-Cheiney et al. 2021), qui était d'abord dédiée à l'inversion et à la surveillance des gaz à effet de serre (CO_2, CH_4) à l'échelle mondiale. Ses capacités à appliquer des approches "top-down" aux espèces hautement réactives à l'échelle régionale et à haute résolution étaient limitées. Par conséquent, dans le cadre du projet ANR ARGONAUT, cette thèse visait à développer la modélisation inverse régionale des espèces réactives en France en utilisant le CIF, le modèle régional national CHIMERE et les nouvelles observations TROPOMI. Pour ce faire, nous avons concentré nos tests sur l'outil CIF-CHIMERE pour les émissions de NO_x en raison de la maturité des produits satellitaires de NO_2 et de leur bonne résolution temporelle. Nous avons considéré comme zones d'étude l'Europe, la France et la Chine. L'Europe et la Chine sont traitées à résolution moyenne (50 km) et la France à haute résolution (10 km). Les inversions sont contraintes soit en utilisant les données de OMI (13 km x 25 km) pour la Chine et l'Europe ou de TROPOMI (5,5 km x 3,5 km) pour l'Europe et la France.

Comme il s'agit d'un nouveau système, nous avons d'abord testé les capacités de l'outil CIF couplé au modèle CHIMERE. Les premiers tests du système CIF-CHIMERE ont été effectués pour tester la configuration du nouveau système et déterminer ses paramètres critiques pour l'inversion. Pour ce faire, un test de base, incluant une configuration adaptée des données de la littérature ou basées sur des connaissances d'experts, a été défini pour inverser les émissions de NO_x sur le domaine de l'Europe à résolution moyenne (50 km) en utilisant les colonnes NO_2 du satellite OMI. Le système a été testé en modifiant un paramètre à la fois, et il a été constaté que les paramètres variationnels, liés aux statistiques d'erreurs des émissions a priori, étaient plus significatifs que les paramètres d'optimisation dans l'estimation des émissions. Les statistiques d'erreur, en particulier celles sur les émissions de fond et leurs corrélations, sont ainsi des éléments déterminants pour l'inversion. L'étude a permis de définir une configuration et une plage de variations des différents paramètres adaptées au système d'inversion et qui permettent de définir rapidement une configuration adaptée à d'autres domaines d'étude.

Dans un second temps, le système CIF-CHIMERE a été utilisé pour estimer les émissions de NO_x chinoises pour 2015 et 2019, en utilisant les données de colonne de NO_2 des observations satellites OMI pour quantifier les réductions des émissions de NO_x chinoises en lien avec les stratégies d'atténuation du gouvernement chinois mises en place. Les années cibles 2015 et 2019 ont été choisies car elles correspondent aux dernières années des deux plans quinquennaux. Nous n'avons pas considéré 2020 en raison de la pandémie de COVID-19 qui aurait pu fausser les conclusions. La faible couverture quotidienne de OMI ne permet pas de dériver des émissions journalières, des corrections mensuelles des émissions de NO_x ont donc été calculées. Nous avons pu constater que les observations de OMI fournissent une meilleure contrainte sur les émissions de NO_x dans les zones polluées par rapport aux zones de fond. L'utilisation de ces estimations d'émissions de NO_x pour simuler les concentrations de NO_2 en surface avec un modèle améliore légèrement la comparaison des simulations avec les mesures au sol. Les budgets annuels estimés sont en bon accord (dans les incertitudes) avec les inventaires indépendants tels que l'inventaire MEIC ou DECSO en 2015. Les résultats suggèrent que l'objectif du douzième plan quinquennal a été atteint (-10 %) dans les zones polluées correspondant aux principales régions urbanisées de Chine. En revanche, l'évolution des émissions estimées de 2015 à 2019 reste faible, et les objectifs de réduction du treizième plan quinquennal n'ont pas été atteints. Ces derniers résultats ne semblent pas complètement cohérents avec la littérature récente et nécessitent des investigations supplémentaires.

Pendant la thèse, une nouvelle fonctionnalité a été implémentée dans le système CIF-CHIMERE par l'équipe de développement pour surmonter les problèmes de convergence des inversions que nous avons rencontrés. Nous avons maintenant la possibilité d'utiliser une échelle logarithmique pour le vecteur de contrôle, ce qui facilite l'optimisation et la convergence du système. Nous avons testé l'influence de l'utilisation de l'échelle logarithmique plutôt que de l'échelle linéaire pour l'inversion sur une seule journée, les résultats suggèrent que les distributions spatiales des corrections des émissions obtenues par inversion ne sont pas modifiées en changeant d'échelle. L'amplitude des corrections est cohérente pour les plus grandes émissions, mais peut être significativement différente dans les régions de fond. Cela ne devrait pas affecter beaucoup le résultat en termes de budget ou de performances de simulation de la qualité de l'air. Néanmoins, cette question devrait être abordée pour une période plus longue à l'avenir pour être plus représentative.

Nous avons également effectué des tests préliminaires pour appliquer le système CIF-CHIMERE en Europe et en France, principalement pour évaluer les changements soudains dans les émissions de NO_x pendant la période de confinement due au COVID-19. Ce travail fait partie du projet ARGONAUT, qui vise à estimer les émissions de NO_x en France à haute résolution (10 km) en utilisant des observations par satellite à haute résolution telles que TROPOMI. Dans un premier temps, le système CIF-CHIMERE a été adapté à une résolution de 50 km, couvrant une zone plus grande que la France, l'Europe, pour comparer les performances d'inversion de TROPOMI et d'OMI. L'accent a été mis sur la première semaine d'avril 2020, et les nouvelles estimations d'émissions ont été comparées aux émissions a priori correspondant à l'année 2016. L'évaluation des résultats d'inversion a montré que les nouvelles estimations d'émissions ont amélioré les statistiques de comparaison par rapport aux émissions a priori, réduisant le biais de 60 % à 20 % et diminuant le RMSE d'environ 25 %. L'amélioration est systématique pour les stations rurales et significative pour les stations urbaines dans les grandes régions urbanisées, mais une dégradation du biais est observée pour certaines stations urbaines, notamment dans le sud de l'Allemagne et en Pologne. En ce qui concerne les changements d'émissions par rapport au prior pour la première semaine d'avril 2020, une réduction des émissions a été observée dans les pays d'Europe du Nord-Ouest, avec des réductions allant jusqu'à 67 % à l'échelle du point de grille du modèle ou environ 35-45 % pour des pays tels que la Belgique, les Pays-Bas, le Royaume-Uni, l'Allemagne ou la Pologne. Les pays du Sud (France, Italie et Espagne) ont montré une réduction de leurs émissions plus faible au cours de cette semaine, ce qui est attribué à l'échantillonnage des observations, qui est plus constant et homogène de jour en jour dans le nord que dans le sud, plutôt qu'à une différence dans l'application de restrictions pendant le confinement. Bien que la méthodologie pour évaluer l'impact du confinement consiste généralement à comparer 2020 avec 2019, des problèmes techniques avec le supercalculateur utilisé pour effectuer l'inversion ont entraîné l'inversion de seulement quelques jours pour avril 2019 ne permettant pas de présenter des résultats totalement représentatifs. Néanmoins, une comparaison préliminaire suggère une réduction de 40 % pour l'Ile-de-France et de 30 à 50 % pour le sud de l'Allemagne et le nord de l'Italie. Une étude élargie sur une période plus longue est nécessaire pour confirmer ces premiers résultats. Ces résultats montrent toutefois que la haute résolution de TROPOMI et ses meilleures performances instrumentales fournissent une plus grande couverture quotidienne et donc des capacités d'inversion des émissions à l'échelle journalière améliorées.

Enfin, les premières tentatives d'inversion à haute résolution pour une journée (1er avril 2020) sur la France semblent prometteuses pour évaluer et surveiller les émissions à haute résolution spatiale.

Cette thèse de doctorat a permis de discuter le potentiel du nouveau système CIF-CHIMERE et des données TROPOMI pour surveiller les émissions de NO_x à différentes résolutions, de modérées à élevées. Cependant, le travail a été réalisé sur des périodes de temps et un nombre de jours limités, une perspective à court terme de ce travail serait d'étendre la durée des périodes étudiées pour obtenir des résultats plus représentatifs et consolidés. A plus long terme, d'autres questions se posent suite à ce travail. Par exemple, nous avons noté que, malgré les améliorations apportées par TROPOMI, la couverture nuageuse restait un facteur limitant dans certaines situations, la question de la résolution temporelle à laquelle les inversions peuvent être effectuées doit donc être encore approfondie. La question de la séparation des émissions anthropiques et biogéniques ou des émissions par secteur reste également une question ouverte pour laquelle la haute résolution spatiale apportée par TROPOMI pourrait aidée mais cela nécessite des développements techniques dans le CIF pour adapter le vecteur de contrôle. D’autre part, l'étude sur la Chine a soulevé la question de l'efficacité de la réduction des émissions au cours du 13e plan quinquennal, une extension de l'étude à l'ensemble de la décennie serait intéressante pour évaluer si les résultats que nous avons obtenus sont spécifiques à l'année 2019 ou généralisables. L'utilisation de TROPOMI pour les années récentes serait également intéressante, mais la cohérence entre les observations OMI et TROPOMI doit être vérifiée pour pouvoir calculer des tendances. En Europe et en France, il serait également très intéressant d'évaluer l'impact de la pandémie sur les émissions de NO_x à haute résolution, en incluant tous les confinements et les restrictions qui ont eu lieu pendant la durée de la pandémie. A noter que pour réaliser ces différents développements et études, il est important de s'assurer que les observations ne sont pas biaisées; cependant, les données TROPOMI semblent encore présenter un biais négatif résiduel. Par conséquent, des techniques de débiaisage pourraient être envisagées avant d'assimiler les données dans le système d'inversion.

Bibliography

A, R. J. van der, D. H. M. U. Peters, H. Eskes, K. F. Boersma, M. Van Roozendael, I. De Smedt, and H. M. Kelder. 2006. "Detection of the Trend and Seasonal Variation in Tropospheric NO2 over China." *Journal of Geophysical Research: Atmospheres* 111 (D12). https://doi.org/10.1029/2005JD006594.

A, Ronald J. van der, Bas Mijling, Jieying Ding, Maria Elissavet Koukouli, Fei Liu, Qing Li, Huiqin Mao, and Nicolas Theys. 2017. "Cleaning up the Air: Effectiveness of Air Quality Policy for SO2 and NOx Emissions in China." *Atmospheric Chemistry and Physics* 17 (3): 1775–89. https://doi.org/10.5194/acp-17-1775-2017.

Abbe, Cleveland. 1901. "THE PHYSICAL BASIS OF LONG-RANGE WEATHER FORECASTS." *Monthly Weather Review* 29 (12): 551–61. https://doi.org/10.1175/1520-0493(1901)29[551c:TPBOLW]2.0.CO;2.

Abram, Nerilie J., Helen V. McGregor, Jessica E. Tierney, Michael N. Evans, Nicholas P. McKay, and Darrell S. Kaufman. 2016. "Early Onset of Industrial-Era Warming across the Oceans and Continents." *Nature* 536 (7617): 411–18. https://doi.org/10.1038/nature19082.

Almaraz, Maya, Edith Bai, Chao Wang, Justin Trousdell, Stephen Conley, Ian Faloona, and Benjamin Z. Houlton. 2018. "Agriculture Is a Major Source of NOx Pollution in California." *Science Advances* 4 (1): eaao3477. https://doi.org/10.1126/sciadv.aao3477.

Anderson, Erik, Mike Fisher, Rosemary Munro, and Anthony McNally. 2000. "Diagnosis of Background Errors for Radiances and Other Observable Quantities in a Variational Data Assimilation Scheme, and the Explanation of a Case of Poor Convergence: BACKGROUND ERRORS FOR OBSERVABLE QUANTITIES." *Quarterly Journal of the Royal Meteorological Society* 126 (565): 1455–72. https://doi.org/10.1002/qj.49712656512.

Anenberg, Susan C, Arash Mohegh, Daniel L Goldberg, Gaige H Kerr, Michael Brauer, Katrin Burkart, Perry Hystad, Andrew Larkin, Sarah Wozniak, and Lok Lamsal. 2022. "Long-Term Trends in Urban NO2 Concentrations and Associated Paediatric Asthma Incidence: Estimates from Global Datasets." *The Lancet Planetary Health* 6 (1): e49–58. https://doi.org/10.1016/S2542-5196(21)00255-2.

Anger, Annela, Olivier Dessens, Fengming Xi, Terry Barker, and Rui Wu. 2016. "China's Air Pollution Reduction Efforts May Result in an Increase in Surface Ozone Levels in Highly Polluted Areas." *Ambio* 45 (2): 254–65. https://doi.org/10.1007/s13280-015-0700-6.

Asch, Mark, Marc Bocquet, and Maëlle Nodet. 2016. *Data Assimilation: Methods, Algorithms, and Applications*. Fundamentals of Algorithms 11. Philadelphia: Society for Industrial and Applied Mathematics.

Atkinson, Roger. 2000. “Atmospheric Chemistry of VOCs and NOx.” *Atmospheric Environment* 34: 2063–2101.

Bannister, R. N. 2017. “A Review of Operational Methods of Variational and Ensemble-variational Data Assimilation.” *Quarterly Journal of the Royal Meteorological Society* 143 (703): 607–33. https://doi.org/10.1002/qj.2982.

Barré, Jérôme, Hervé Petetin, Augustin Colette, Marc Guevara, Vincent-Henri Peuch, Laurence Rouil, Richard Engelen, et al. 2021. “Estimating Lockdown-Induced European NO2 Changes Using Satellite and Surface Observations and Air Quality Models.” *Atmospheric Chemistry and Physics* 21 (9): 7373–94. https://doi.org/10.5194/acp-21-7373-2021.

Bayes. 1763. “LII. An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, F. R. S. communicated by Mr. Price, in a letter to John Canton, A. M. F. R. S.” *Philosophical Transactions of the Royal Society of London* 53 (December): 370–418. https://doi.org/10.1098/rstl.1763.0053.

Berchet, Antoine, Espen Sollum, Rona L. Thompson, Isabelle Pison, Joël Thanwerdas, Grégoire Broquet, Frédéric Chevallier, et al. 2021. “The Community Inversion Framework v1.0: A Unified System for Atmospheric Inversion Studies.” *Geoscientific Model Development* 14 (8): 5331–54. https://doi.org/10.5194/gmd-14-5331-2021.

Bergthörsson, Páll, and Bo R. DÖÖs. 1955. “Numerical Weather Map Analysis.” *Tellus* 7 (3): 329–40. https://doi.org/10.3402/tellusa.v7i3.8902.

Berthet, Gwenaël, Fabrice Jégou, Valéry Catoire, Gisèle Krysztofiak, Jean-Baptiste Renard, Adam E. Bourassa, Doug A. Degenstein, et al. 2017. “Impact of a Moderate Volcanic Eruption on Chemistry in the Lower Stratosphere: Balloon-Borne Observations and Model Calculations.” *Atmospheric Chemistry and Physics* 17 (3): 2229–53. https://doi.org/10.5194/acp-17-2229-2017.

Bjerknes, Vilhelm. 1904. “Das Problem Der Wettervorhersage, Betrachtet Vom Standpunkte Der Mechanik Und Der Physik (The Problem of Weather Prediction, Considered from the Viewpoints of Mechanics and Physics).” *Meteorol. Z.* 21 (1–7): (translated and edited by VOLKEN E. and S. BRONNIMANN ¨ .-Meteorol. Z. 18 (2009), 663-667). https://doi.org/10.1127/0941-2948/2009/416.

Blanco-Ward, D., A. Rocha, C. Viceto, A.C. Ribeiro, M. Feliciano, E. Paoletti, and A.I. Miranda. 2021. “Validation of Meteorological and Ground-Level Ozone WRF-CHIMERE Simulations in a Mountainous Grapevine Growing Area for Phytotoxic Risk Assessment.” *Atmospheric Environment* 259 (August): 118507. https://doi.org/10.1016/j.atmosenv.2021.118507.

Bocquet, M., H. Elbern, H. Eskes, M. Hirtl, R. Žabkar, G. R. Carmichael, J. Flemming, et al. 2015. “Data Assimilation in Atmospheric Chemistry Models: Current Status and Future Prospects for

Coupled Chemistry Meteorology Models." *Atmospheric Chemistry and Physics* 15 (10): 5325–58. https://doi.org/10.5194/acp-15-5325-2015.

Boersma, K. F., H. J. Eskes, R. J. Dirksen, R. J. van der A, J. P. Veefkind, P. Stammes, V. Huijnen, et al. 2011. "An Improved Tropospheric NO2 Column Retrieval Algorithm for the Ozone Monitoring Instrument." *Atmospheric Measurement Techniques* 4 (9): 1905–28. https://doi.org/10.5194/amt-4-1905-2011.

Boersma, K. F., H. J. Eskes, J. P. Veefkind, E. J. Brinksma, R. J. van der A, M. Sneep, G. H. J. van den Oord, et al. 2007. "Near-Real Time Retrieval of Tropospheric NO2 from OMI." *Atmospheric Chemistry and Physics* 7 (8): 2103–18. https://doi.org/10.5194/acp-7-2103-2007.

Boersma, K. F., G. C. M. Vinken, and H. J. Eskes. 2016. "Representativeness Errors in Comparing Chemistry Transport and Chemistry Climate Models with Satellite UV–Vis Tropospheric Column Retrievals." *Geoscientific Model Development* 9 (2): 875–98. https://doi.org/10.5194/gmd-9-875-2016.

Boersma, K. Folkert, Henk J. Eskes, Andreas Richter, Isabelle De Smedt, Alba Lorente, Steffen Beirle, Jos H. G. M. van Geffen, et al. 2018. "Improving Algorithms and Uncertainty Estimates for Satellite NO2 Retrievals: Results from the Quality Assurance for the Essential Climate Variables (QA4ECV) Project." *Atmospheric Measurement Techniques* 11 (12): 6651–78. https://doi.org/10.5194/amt-11-6651-2018.

Boersma, K. Folkert, Henk Eskes, Andreas Richter, Isabelle De Smedt, Alba Lorente, Steffen Beirle, Jos Van Geffen, Enno Peters, Michel Van Roozendael, and Thomas Wagner. 2017. "QA4ECV NO2 Tropospheric and Stratospheric Column Data from OMI." Royal Netherlands Meteorological Institute (KNMI). http://www.qa4ecv.eu/ecv/no2-pre/data.

Boisgontier, H., V. Mallet, J.P. Berroir, M. Bocquet, I. Herlin, and B. Sportisse. 2008. "Satellite Data Assimilation for Air Quality Forecast." *Simulation Modelling Practice and Theory* 16 (10): 1541–45. https://doi.org/10.1016/j.simpat.2008.01.008.

Borger, Christian, Steffen Beirle, Steffen Dörner, Holger Sihler, and Thomas Wagner. 2020. "Total Column Water Vapour Retrieval from S-5P/TROPOMI in the Visible Blue Spectral Range." *Atmospheric Measurement Techniques* 13 (5): 2751–83. https://doi.org/10.5194/amt-13-2751-2020.

Bouttier, B, and P. Courtier. 1999. "Data Assimilation Concepts and Methods." Meteorological Training Course Lecture Series. ECMWF.

Bovensmann, H., J. P. Burrows, M. Buchwitz, J. Frerick, S. Noël, V. V. Rozanov, K. V. Chance, and A. P. H. Goede. 1999. "SCIAMACHY: Mission Objectives and Measurement Modes." *Journal of the Atmospheric Sciences* 56 (2): 127–50. https://doi.org/10.1175/1520-0469(1999)056<0127:SMOAMM>2.0.CO;2.

Burgers, Gerrit, Peter Jan van Leeuwen, and Geir Evensen. 1997. “Analysis Scheme in the Ensemble Kalman Filter.” *Monthly Weather Review* 126 (6): 1719–24. https://doi.org/10.1175/1520-0493(1998)126<1719:ASITEK>2.0.CO;2.

Burrows, John P., Mark Weber, Michael Buchwitz, Vladimir Rozanov, Annette Ladstätter-Weißenmayer, Andreas Richter, Rüdiger DeBeek, et al. 1999. “The Global Ozone Monitoring Experiment (GOME): Mission Concept and First Scientific Results.” *Journal of the Atmospheric Sciences* 56 (2): 151–75. https://doi.org/10.1175/1520-0469(1999)056<0151:TGOMEG>2.0.CO;2.

CAA (Clean Air Asia). 2016. “China Air 2016—Air Pollution Prevention and Control Progress in Chinese Cities.” http://www.allaboutair.cn/a/cbw/bg/2016/0822/472.html.

Chai, Tianfeng, Gregory R. Carmichael, Youhua Tang, Adrian Sandu, Andreas Heckel, Andreas Richter, and John P. Burrows. 2009. “Regional NOx Emission Inversion through a Four-Dimensional Variational Approach Using SCIAMACHY Tropospheric NO2 Column Observations.” *Atmospheric Environment* 43 (32): 5046–55. https://doi.org/10.1016/j.atmosenv.2009.06.052.

Citepa. 2021. “Inventaire Des Émissions de Polluants Atmosphériques En France Métropolitaine.” Format CEE-NU.

Colette, Augustin, Claire Granier, Ø. Hodnebrog, H. Jakobs, A. Maurizi, A. Nyiri, S. Rao, et al. 2012. “Future Air Quality in Europe: A Multi-Model Assessment of Projected Exposure to Ozone.” *Atmospheric Chemistry and Physics* 12 (21): 10613–30. https://doi.org/10.5194/acp-12-10613-2012.

Compernolle, Steven, Tijl Verhoelst, Gaia Pinardi, José Granville, Daan Hubert, Arno Keppens, Sander Niemeijer, et al. 2020. “Validation of Aura-OMI QA4ECV NO2 Climate Data Records with Ground-Based DOAS Networks: The Role of Measurement and Comparison Uncertainties.” *Atmospheric Chemistry and Physics* 20 (13): 8017–45. https://doi.org/10.5194/acp-20-8017-2020.

Cooper, Matthew, Randall V. Martin, Akhila Padmanabhan, and Daven K. Henze. 2017. “Comparing Mass Balance and Adjoint Methods for Inverse Modeling of Nitrogen Dioxide Columns for Global Nitrogen Oxide Emissions.” *Journal of Geophysical Research: Atmospheres* 122 (8): 4718–34. https://doi.org/10.1002/2016JD025985.

Core Writing Team, Rajendra K. Pashauri, and A. (eds.) Reisinger. 2008. *Climate Change 2007: Synthesis Report. Contribution of Working Groups I, II and III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Geneva, Switzerland: IPCC.

Courtier, P., E. Andersson, W. Heckley, D. Vasiljevic, M. Hamrud, A. Hollingsworth, F. Rabier, M. Fisher, and J. Pailleux. 1998. “The ECMWF Implementation of Three-Dimensional

Variational Assimilation (3D-Var). I: Formulation.” *Quarterly Journal of the Royal Meteorological Society* 124 (550): 1783–1807. https://doi.org/10.1002/qj.49712455002.

Courtier, P., J.-N. Thépaut, and A. Hollingsworth. 1994. “A Strategy for Operational Implementation of 4D-Var, Using an Incremental Approach.” *Quarterly Journal of the Royal Meteorological Society* 120 (519): 1367–87. https://doi.org/10.1002/qj.49712051912.

Cressman, George P. 1959. “AN OPERATIONAL OBJECTIVE ANALYSIS SYSTEM.” *Monthly Weather Review* 87 (10): 367–74. https://doi.org/10.1175/1520-0493(1959)087<0367:AOOAS>2.0.CO;2.

Crippa, Monica, Diego Guizzardi, Marilena Muntean, Edwin Schaaf, Frank Dentener, John A. van Aardenne, Suvi Monni, et al. 2018. “Gridded Emissions of Air Pollutants for the Period 1970–2012 within EDGAR v4.3.2.” *Earth System Science Data* 10 (4): 1987–2013. https://doi.org/10.5194/essd-10-1987-2018.

Crippa, Monica, Greet Janssens-Maenhout, Frank Dentener, Diego Guizzardi, Katerina Sindelarova, Marilena Muntean, Rita Van Dingenen, and Claire Granier. 2016. “Forty Years of Improvements in European Air Quality: Regional Policy-Industry Interactions with Global Impacts.” *Atmospheric Chemistry and Physics* 16 (6): 3825–41. https://doi.org/10.5194/acp-16-3825-2016.

Daley, Roger. 1991. *Atmospheric Data Analysis*. Cambridge Atmospheric and Space Science Series 2. Cambridge ; New York: Cambridge University Press.

Derognat, C. 2003. “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* 108 (D17): 8560. https://doi.org/10.1029/2001JD001421.

Dimet, François-Xavier Le, and Olivier Talagrand. 1986. “Variational Algorithms for Analysis and Assimilation of Meteorological Observations: Theoretical Aspects.” *Tellus A* 38A (2): 97–110. https://doi.org/10.1111/j.1600-0870.1986.tb00459.x.

Ding, J., R. J. van der A, B. Mijling, P. F. Levelt, and N. Hao. 2015. “NOx Emission Estimates during the 2014 Youth Olympic Games in Nanjing.” *Atmospheric Chemistry and Physics* 15 (16): 9399–9412. https://doi.org/10.5194/acp-15-9399-2015.

Ding, Jieying, Kazuyuki Miyazaki, Ronald Johannes van der A, Bas Mijling, Jun-ichi Kurokawa, SeogYeon Cho, Greet Janssens-Maenhout, Qiang Zhang, Fei Liu, and Pieternel Felicitas Levelt. 2017. “Intercomparison of NOx Emission Inventories over East Asia.” *Atmospheric Chemistry and Physics* 17 (16): 10125–41. https://doi.org/10.5194/acp-17-10125-2017.

Elbern, H., H. Schmidt, O. Talagrand, and A. Ebel. 2000. “4D-Variational Data Assimilation with an Adjoint Air Quality Model for Emission Analysis.” *Environmental Modelling & Software*, Air pollution modelling and simulation, 15 (6): 539–48. https://doi.org/10.1016/S1364-8152(00)00049-9.

Elbern, H., A. Strunk, H. Schmidt, and O. Talagrand. 2007. “Emission Rate and Chemical State Estimation by 4-Dimensional Variational Inversion.” *Atmospheric Chemistry and Physics* 7 (14): 3749–69. https://doi.org/10.5194/acp-7-3749-2007.

Elbern, Hendrik, and Hauke Schmidt. 1999. “A Four-Dimensional Variational Chemistry Data Assimilation Scheme for Eulerian Chemistry Transport Modeling.” *Journal of Geophysical Research: Atmospheres* 104 (D15): 18583–98. https://doi.org/10.1029/1999JD900280.

Elbern, Hendrik, and Hauke Schmidt. 2001. “Ozone Episode Analysis by Four-Dimensional Variational Chemistry Data Assimilation.” *Journal of Geophysical Research: Atmospheres* 106 (D4): 3569–90. https://doi.org/10.1029/2000JD900448.

Elbern, Hendrik, Hauke Schmidt, and Adolf Ebel. 1997. “Variational Data Assimilation for Tropospheric Chemistry Modeling.” *Journal of Geophysical Research: Atmospheres* 102 (D13): 15967–85. https://doi.org/10.1029/97JD01213.

Elguindi, N., C. Granier, T. Stavrakou, S. Darras, M. Bauwens, H. Cao, C. Chen, et al. 2020. “Intercomparison of Magnitudes and Trends in Anthropogenic Surface Emissions From Bottom-Up Inventories, Top-Down Estimates, and Emission Scenarios.” *Earth’s Future* 8 (8): e2020EF001520. https://doi.org/10.1029/2020EF001520.

Elguindi, Nellie, Claire Granier, Trissevgeni Stavrakou, Sabine Darras, Maite Bauwens, Hansen Cao, C. Chen, et al. 2020. “Intercomparison of Magnitudes and Trends in Anthropogenic Surface Emissions From Bottom-Up Inventories, Top-Down Estimates, and Emission Scenarios.” *Earths Future* 8 (8). https://doi.org/10.1029/2020EF001520.

EPA United States Environmental Protection Agency. n.d. “Setting and Reviewing Standards to Control NO2 Pollution, National Ambient Air Quality Standards (NAAQS).” https://www.epa.gov/no2-pollution/setting-and-reviewing-standards-control-no2-pollution.

Eskes, H. J., A. J. M. Piters, P. F. Levelt, M. A. F. Allaart, and H. M. Kelder. 1999. “Variational Assimilation of GOME Total-Column Ozone Satellite Data in a 2D Latitude–Longitude Tracer-Transport Model.” *Journal of the Atmospheric Sciences* 56 (20): 3560–72. https://doi.org/10.1175/1520-0469(1999)056<3560:VAOGTC>2.0.CO;2.

Eskes, Henk, Jos van Geffen, Maarten Sneep, J. Pepijn Veefkind, Sander Niemeijer, and Claus Zehner. 2021. *S5P Nitrogen Dioxide V02.03.01 Intermediate Reprocessing on the S5P-PAL System: Readme File*. 1.0. https://data-portal.s5p-pal.com/product-docs/no2/PAL_reprocessing_NO2_v02.03.01_20211215.pdf.

Eskes, Henk, Jos Van Geffen, Folkert Boersma, Kai-Uwe Eichmann, Arnoud Apituley, Mattia Pedergnana, Maarten Sneep, J. Pepijn Veefkind, and Diego Loyola. 2019. *Sentinel-5*

Precursor/TROPOMI Level 2 Product User Manual Nitrogendioxide. Vol. 3.0.0. Royal Netherlands Meteorological Institute (KNMI).

European Environment Agency. 2022. *Europe’s Air Quality Status 2021*. EEA Report (Online). LU: Publications Office. https://data.europa.eu/doi/10.2800/488115.

European Space Agency, ed. 1995. *Global Ozone Monitoring Experiment: GOME: Users Manual*. SP 1182. Noordwijk, the Netherlands: ESA Publications Division : European Space Research and Technology Centre.

European Union Emission Inventory Report. 2021. Luxembourg: Publications Office of the European Union.

Evensen, Geir. 1994. “Sequential Data Assimilation with a Nonlinear Quasi-Geostrophic Model Using Monte Carlo Methods to Forecast Error Statistics.” *Journal of Geophysical Research: Oceans* 99 (C5): 10143–62. https://doi.org/10.1029/94JC00572.

Farman, J. C., B. G. Gardiner, and J. D. Shanklin. 1985. “Large Losses of Total Ozone in Antarctica Reveal Seasonal ClOx/NOx Interaction.” *Nature* 315 (6016): 207–10. https://doi.org/10.1038/315207a0.

Fisher, M., and D. J. Lary. 1995. “Lagrangian Four-Dimensional Variational Data Assimilation of Chemical Species.” *Quarterly Journal of the Royal Meteorological Society* 121 (527): 1681–1704. https://doi.org/10.1002/qj.49712152709.

Flemming, J, R Stern, and R Yamartino. 2005. “A New Air Quality Regime Classification Scheme for O, NO, SO and PM10 Observations Sites.” *Atmospheric Environment* 39 (33): 6121–29. https://doi.org/10.1016/j.atmosenv.2005.06.039.

Fortems-Cheiney, Audrey, Isabelle Pison, Grégoire Broquet, Gaëlle Dufour, Antoine Berchet, Elise Potier, Adriana Coman, Guillaume Siour, and Lorenzo Costantino. 2021. “Variational Regional Inverse Modeling of Reactive Species Emissions with PYVAR-CHIMERE-V2019.” *Geoscientific Model Development* 14 (5): 2939–57. https://doi.org/10.5194/gmd-14-2939-2021.

Foy, Benjamin de, Zifeng Lu, and David G. Streets. 2016. “Satellite NO2 Retrievals Suggest China Has Exceeded Its NOx Reduction Goals from the Twelfth Five-Year Plan.” *Scientific Reports* 6 (1): 35912. https://doi.org/10.1038/srep35912.

Foy, Benjamin de, Joseph L. Wilkins, Zifeng Lu, David G. Streets, and Bryan N. Duncan. 2014. “Model Evaluation of Methods for Estimating Surface Emissions and Chemical Lifetimes from Satellite Data.” *Atmospheric Environment* 98 (December): 66–77. https://doi.org/10.1016/j.atmosenv.2014.08.051.

Galloway, J. N., F. J. Dentener, D. G. Capone, E. W. Boyer, R. W. Howarth, S. P. Seitzinger, G. P. Asner, et al. 2004. “Nitrogen Cycles: Past, Present, and Future.” *Biogeochemistry* 70 (2): 153–226. https://doi.org/10.1007/s10533-004-0370-0.

Galmarini, Stefano, Brigitte Koffi, Efisio Solazzo, Terry Keating, Christian Hogrefe, Michael Schulz, Anna Benedictow, et al. 2017. "Technical Note: Coordination and Harmonization of the Multi-Scale, Multi-Model Activities HTAP2, AQMEII3, and MICS-Asia3: Simulations, Emission Inventories, Boundary Conditions, and Model Output Formats." *Atmospheric Chemistry and Physics* 17 (2): 1543–55. https://doi.org/10.5194/acp-17-1543-2017.

Geffen, Jos van, Henk Eskes, Steven Compernolle, Gaia Pinardi, Tijl Verhoelst, Jean-Christopher Lambert, Maarten Sneep, et al. 2022. "Sentinel-5P TROPOMI NO2 Retrieval: Impact of Version v2.2 Improvements and Comparisons with OMI and Ground-Based Data." *Atmospheric Measurement Techniques* 15 (7): 2037–60. https://doi.org/10.5194/amt-15-2037-2022.

Ghil, M. 1989. "Meteorological Data Assimilation for Oceanographers. Part I: Description and Theoretical Framework." *Dynamics of Atmospheres and Oceans* 13 (3–4): 171–218. https://doi.org/10.1016/0377-0265(89)90040-7.

Gilbert, Jean Charles, and Claude Lemaréchal. 1989. "Some Numerical Experiments with Variable-Storage Quasi-Newton Algorithms." *Mathematical Programming* 45 (1–3): 407–35. https://doi.org/10.1007/BF01589113.

Gilbert, Jean Charles, Claude Lemaréchal, and The Simulator. 2009. "The Module M1QN3 Version 3.3 (October 2009)."

Granier, C., S. Darras, H. Denier van der Gon, J. Doubalova, N. Elguindi, B. Galle, M. Gauss, et al. 2019. "The Copernicus Atmosphere Monitoring Service Global and Regional Emissions (April 2019 Version)." https://doi.org/10.24380/D0BN-KX16.

Granier, Claire, Bertrand Bessagnet, Tami Bond, Ariela D'Angiola, Hugo Denier van der Gon, Gregory J. Frost, Angelika Heil, et al. 2011. "Evolution of Anthropogenic and Biomass Burning Emissions of Air Pollutants at Global and Regional Scales during the 1980–2010 Period." *Climatic Change* 109 (1–2): 163–90. https://doi.org/10.1007/s10584-011-0154-1.

Guenther, A. B., X. Jiang, C. L. Heald, T. Sakulyanontvittaya, T. Duhl, L. K. Emmons, and X. Wang. 2012. "The Model of Emissions of Gases and Aerosols from Nature Version 2.1 (MEGAN2.1): An Extended and Updated Framework for Modeling Biogenic Emissions." *Geoscientific Model Development* 5 (6): 1471–92. https://doi.org/10.5194/gmd-5-1471-2012.

Guenther, A., T. Karl, P. Harley, C. Wiedinmyer, P. I. Palmer, and C. Geron. 2006. "Estimates of Global Terrestrial Isoprene Emissions Using MEGAN (Model of Emissions of Gases and Aerosols from Nature)." *Atmospheric Chemistry and Physics* 6 (11): 3181–3210. https://doi.org/10.5194/acp-6-3181-2006.

Guevara, Marc, Oriol Jorba, Albert Soret, Hervé Petetin, Dene Bowdalo, Kim Serradell, Carles Tena, et al. 2021. “Time-Resolved Emission Reductions for Atmospheric Chemistry Modelling in Europe during the COVID-19 Lockdowns.” *Atmospheric Chemistry and Physics* 21 (2): 773–97. https://doi.org/10.5194/acp-21-773-2021.

Haas-Laursen, Danielle E., Dana E. Hartley, and Ronald G. Prinn. 1996. “Optimizing an Inverse Method to Deduce Time-Varying Emissions of Trace Gases.” *Journal of Geophysical Research: Atmospheres* 101 (D17): 22823–31. https://doi.org/10.1029/96JD02018.

Hadamard, Jacques. 1923. *Lectures on Cauchy’s Problem in Linear Partial Differential Equations*. New Haven Yale University Press. http://archive.org/details/lecturesoncauchy00hadauoft.

Han, Kyung M., Hyun S. Kim, and Chul H. Song. 2020. “An Estimation of Top-Down NOx Emissions from OMI Sensor Over East Asia.” *Remote Sensing* 12 (12): 2004. https://doi.org/10.3390/rs12122004.

Hansen, Per Christian. 2010. *Discrete Inverse Problems: Insight and Algorithms*. Society for Industrial and Applied Mathematics. https://doi.org/10.1137/1.9780898718836.

Hartley, Dana, and Ronald Prinn. 1993. “Feasibility of Determining Surface Emissions of Trace Gases Using an Inverse Method in a Three-Dimensional Chemical Transport Model.” *Journal of Geophysical Research: Atmospheres* 98 (D3): 5183–97. https://doi.org/10.1029/92JD02594.

Hauglustaine, D. A., F. Hourdin, L. Jourdain, M.-A. Filiberti, S. Walters, J.-F. Lamarque, and E. A. Holland. 2004. “Interactive Chemistry in the Laboratoire de Météorologie Dynamique General Circulation Model: Description and Background Tropospheric Chemistry Evaluation: INTERACTIVE CHEMISTRY IN LMDZ.” *Journal of Geophysical Research: Atmospheres* 109 (D4): D04314. https://doi.org/10.1029/2003JD003957.

He, Tai-Long, Dylan B. A. Jones, Kazuyuki Miyazaki, Kevin W. Bowman, Zhe Jiang, Xiaokang Chen, Rui Li, Yuxiang Zhang, and Kunna Li. 2022. “Inverse Modeling of Chinese NOx Emissions Using Deep Learning: Integrating in Situ Observations with a Satellite-Based Chemical Reanalysis.” Preprint. Gases/Atmospheric Modelling/Troposphere/Chemistry (chemical composition and reactions). https://doi.org/10.5194/acp-2022-251.

Henze, D K, A Hakami, and J H Seinfeld. 2007. “Development of the Adjoint of GEOS-Chem.” *Atmos. Chem. Phys.*, 21.

Houghton, J. T., Y. Ding, D.J. Griggs, M. Noguer, P.J. van der Linden, X. Dai, K. Maskell, and C.A. Johnson (eds.). 2001. *IPCC, 2001: Climate Change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press. https://www.jstor.org/stable/10.2307/20033020?origin=crossref.

Houtekamer, P. L., and Herschel L. Mitchell. 1998. “Data Assimilation Using an Ensemble Kalman Filter Technique.” *Monthly Weather Review* 126 (3): 796–811. https://doi.org/10.1175/1520-0493(1998)126<0796:DAUAEK>2.0.CO;2.

Houtekamer, P. L., and Herschel L. Mitchell. 2001. “A Sequential Ensemble Kalman Filter for Atmospheric Data Assimilation.” *Monthly Weather Review* 129 (1): 123–37. https://doi.org/10.1175/1520-0493(2001)129<0123:ASEKFF>2.0.CO;2.

Houtekamer, P. L., Herschel L. Mitchell, Gérard Pellerin, Mark Buehner, Martin Charron, Lubos Spacek, and Bjarne Hansen. 2005. “Atmospheric Data Assimilation with an Ensemble Kalman Filter: Results with Real Observations.” *Monthly Weather Review* 133 (3): 604–20. https://doi.org/10.1175/MWR-2864.1.

Huijnen, V., H. J. Eskes, A. Poupkou, H. Elbern, K. F. Boersma, G. Foret, M. Sofiev, et al. 2010. “Comparison of OMI NO2 Tropospheric Columns with an Ensemble of Global and European Regional Air Quality Models.” *Atmospheric Chemistry and Physics* 10 (7): 3273–96. https://doi.org/10.5194/acp-10-3273-2010.

Huszar, Peter, Jan Karlický, Jana Marková, Tereza Nováková, Marina Liaskoni, and Lukáš Bartík. 2021. “The Regional Impact of Urban Emissions on Air Quality in Europe: The Role of the Urban Canopy Effects.” *Atmospheric Chemistry and Physics* 21 (18): 14309–32. https://doi.org/10.5194/acp-21-14309-2021.

Ide, Kayo, Philippe Courtier, Michael Ghil, and Andrew C. Lorenc. 1997. “Unified Notation for Data Assimilation : Operational, Sequential and Variational (GtSpecial IssueltData Assimilation in Meteology and Oceanography: Theory and Practice).” *Journal of the Meteorological Society of Japan. Ser. II* 75 (1B): 181–89. https://doi.org/10.2151/jmsj1965.75.1B_181.

Itahashi, Syuichi, Keiya Yumimoto, Jun-ichi Kurokawa, Yu Morino, Tatsuya Nagashima, Kazuyuki Miyazaki, Takashi Maki, and Toshimasa Ohara. 2019. “Inverse Estimation of NOx Emissions over China and India 2005–2016: Contrasting Recent Trends and Future Perspectives.” *Environmental Research Letters* 14 (12): 124020. https://doi.org/10.1088/1748-9326/ab4d7f.

Jamali, Sadegh, Daniel Klingmyr, and Torbern Tagesson. 2020. “Global-Scale Patterns and Trends in Tropospheric NO2 Concentrations, 2005–2018.” *Remote Sensing* 12 (21): 3526. https://doi.org/10.3390/rs12213526.

Janssens-Maenhout, G., M. Crippa, D. Guizzardi, F. Dentener, M. Muntean, G. Pouliot, T. Keating, et al. 2015. “HTAP_v2.2: A Mosaic of Regional and Global Emission Grid Maps for 2008 and 2010 to Study Hemispheric Transport of Air Pollution.” *Atmospheric Chemistry and Physics* 15 (19): 11411–32. https://doi.org/10.5194/acp-15-11411-2015.

Jazwinski, Andrew H. 1970. *Stochastic Processes and Filtering Theory*. Burlington: Elsevier Science.

Jin, Yana, Henrik Andersson, and Shiqiu Zhang. 2016. “Air Pollution Control Policies in China: A Retrospective and Prospects.” *International Journal of Environmental Research and Public Health* 13 (12): 1219. https://doi.org/10.3390/ijerph13121219.

Judd, Laura M., Jassim A. Al-Saadi, James J. Szykman, Lukas C. Valin, Scott J. Janz, Matthew G. Kowalewski, Henk J. Eskes, et al. 2020. “Evaluating Sentinel-5P TROPOMI Tropospheric NOxcolumn Densities with Airborne and Pandora Spectrometers near New York City and Long Island Sound.” *Atmospheric Measurement Techniques* 13 (11): 6113–40. https://doi.org/10.5194/amt-13-6113-2020.

Kabanikhin, S. I. 2008. “Definitions and Examples of Inverse and Ill-Posed Problems.” *Journal of Inverse and Ill-Posed Problems* 16 (4). https://doi.org/10.1515/JIIP.2008.019.

Kalman, R. E. 1960. “A New Approach to Linear Filtering and Prediction Problems.” *Journal of Basic Engineering* 82 (1): 35–45. https://doi.org/10.1115/1.3662552.

Kampa, Marilena, and Elias Castanas. 2008. “Human Health Effects of Air Pollution.” *Environmental Pollution* 151 (2): 362–67. https://doi.org/10.1016/j.envpol.2007.06.012.

Keller, Christoph A., and Mat J. Evans. 2019. “Application of Random Forest Regression to the Calculation of Gas-Phase Chemistry within the GEOS-Chem Chemistry Model V10.” *Geoscientific Model Development* 12 (3): 1209–25. https://doi.org/10.5194/gmd-12-1209-2019.

Keller, Joseph B. 1976. “Inverse Problems.” *The American Mathematical Monthly* 83 (2): 107–18. https://doi.org/10.1080/00029890.1976.11994053.

Kenagy, Hannah S., Tamara L. Sparks, Carlena J. Ebben, Paul J. Wooldrige, Felipe D. Lopez-Hilfiker, Ben H. Lee, Joel A. Thornton, et al. 2018. “NOx Lifetime and NOy Partitioning During WINTER.” *Journal of Geophysical Research: Atmospheres* 123 (17): 9813–27. https://doi.org/10.1029/2018JD028736.

Konovalov, I B, M Beekmann, A Richter, and J P Burrows. 2006. “Inverse Modelling of the Spatial Distribution of NOx Emissions on a Continental Scale Using Satellite Data.” *Atmos. Chem. Phys.*, 24.

Konovalov, I. B., M. Beekmann, A. Richter, J. P. Burrows, and A. Hilboll. 2010. “Multi-Annual Changes of NOx Emissions in Megacity Regions: Nonlinear Trend Analysis of Satellite Measurement Based Estimates.” *Atmospheric Chemistry and Physics* 10 (17): 8481–98. https://doi.org/10.5194/acp-10-8481-2010.

Konovalov, I. B., N. F. Elanskii, A. M. Zvyagintsev, I. B. Belikov, and M. Beekmann. 2009. “Validation of Chemistry Transport Model of the Lower Atmosphere in the Central European Region of Russia Using Ground-Based and Satellite Measurement Data.” *Russian Meteorology and Hydrology* 34 (4): 236–42. https://doi.org/10.3103/S1068373909040062.

Kreher, Karin, Michel Van Roozendael, Francois Hendrick, Arnoud Apituley, Ermioni Dimitropoulou, Udo Frieß, Andreas Richter, et al. 2020. “Intercomparison of NO2, O4, O3 and HCHO Slant Column Measurements by MAX-DOAS and Zenith-Sky UV–Visible Spectrometers during CINDI-2.” *Atmospheric Measurement Techniques* 13 (5): 2169–2208. https://doi.org/10.5194/amt-13-2169-2020.

Kuenen, J. J. P., A. J. H. Visschedijk, M. Jozwicka, and H. A. C. Denier van der Gon. 2014. “TNO-MACC_II Emission Inventory; a Multi-Year (2003–2009) Consistent High-Resolution European Emission Inventory for Air Quality Modelling.” *Atmospheric Chemistry and Physics* 14 (20): 10963–76. https://doi.org/10.5194/acp-14-10963-2014.

Kuenen, Jeroen, Stijn Dellaert, Antoon Visschedijk, Jukka-Pekka Jalkanen, Ingrid Super, and Hugo Denier van der Gon. 2022. “CAMS-REG-v4: A State-of-the-Art High-Resolution European Emission Inventory for Air Quality Modelling.” *Earth System Science Data* 14 (2): 491–515. https://doi.org/10.5194/essd-14-491-2022.

Kurokawa, Jun-ichi, Keiya Yumimoto, Itushi Uno, and Toshimasa Ohara. 2009. “Adjoint Inverse Modeling of NOx Emissions over Eastern China Using Satellite Observations of NO2 Vertical Column Densities.” *Atmospheric Environment* 43 (11): 1878–87. https://doi.org/10.1016/j.atmosenv.2008.12.030.

Lachatre, Mathieu, Gilles Foret, Benoit Laurent, Guillaume Siour, Juan Cuesta, Gaëlle Dufour, Fan Meng, Wei Tang, Qijie Zhang, and Matthias Beekmann. 2020. “Air Quality Degradation by Mineral Dust over Beijing, Chengdu and Shanghai Chinese Megacities.” *Atmosphere* 11 (7): 708. https://doi.org/10.3390/atmos11070708.

Lahoz, William A., and Philipp Schneider. 2014. “Data Assimilation: Making Sense of Earth Observation.” *Frontiers in Environmental Science* 2 (May). https://doi.org/10.3389/fenvs.2014.00016.

Lamsal, L. N., R. V. Martin, A. van Donkelaar, M. Steinbacher, E. A. Celarier, E. Bucsela, E. J. Dunlea, and J. P. Pinto. 2008. “Ground-Level Nitrogen Dioxide Concentrations Inferred from the Satellite-Borne Ozone Monitoring Instrument.” *Journal of Geophysical Research: Atmospheres* 113 (D16). https://doi.org/10.1029/2007JD009235.

Lasek, Janusz Andrzej, and Radosław Lajnert. 2022. “On the Issues of NOx as Greenhouse Gases: An Ongoing Discussion….” *Applied Sciences* 12 (20): 10429. https://doi.org/10.3390/app122010429.

Lattuati, M. 1997. “Contribution a l’etude Du Bilan de l’ozone Tropospherique a l’interface de l’Europe et de l’Atlantique Nord: Modelisation Lagrangienne et Mesures En Altitude.” Ph.D. Thesis, Paris: Universite de Paris 6.

Lee, D.S., I. Köhler, E. Grobler, F. Rohrer, R. Sausen, L. Gallardo-Klenner, J.G.J. Olivier, F.J. Dentener, and A.F. Bouwman. 1997. “Estimations of Global NOx Emissions and Their Uncertainties.”

Atmospheric Environment 31 (12): 1735–49. https://doi.org/10.1016/S1352-2310(96)00327-5.

Leue, C., M. Wenig, T. Wagner, Oliver Klimm, U. Platt, and B. Jähne. 2001. "Quantitative Analysis of NOx Emissions from Global Ozone Monitoring Experiment Satellite Image Sequences." *Journal of Geophysical Research: Atmospheres* 106 (D6): 5493–5505. https://doi.org/10.1029/2000JD900572.

Levelt, P.F., G.H.J. van den Oord, M.R. Dobber, A. Malkki, Huib Visser, Johan de Vries, P. Stammes, J.O.V. Lundell, and H. Saari. 2006. "The Ozone Monitoring Instrument." *IEEE Transactions on Geoscience and Remote Sensing* 44 (5): 1093–1101. https://doi.org/10.1109/TGRS.2006.872333.

Lewis, John M., and John C. Derber. 1985. "The Use of Adjoint Equations to Solve a Variational Adjustment Problem with Advective Constraints." *Tellus A* 37A (4): 309–22. https://doi.org/10.1111/j.1600-0870.1985.tb00430.x.

Li, Meng, Huan Liu, Guannan Geng, Chaopeng Hong, Fei Liu, Yu Song, Dan Tong, et al. 2017. "Anthropogenic Emission Inventories in China: A Review." *National Science Review* 4 (6): 834–66. https://doi.org/10.1093/nsr/nwx150.

Li, Meng, Qiang Zhang, Jun-ichi Kurokawa, Jung-Hun Woo, Kebin He, Zifeng Lu, Toshimasa Ohara, et al. 2017. "MIX: A Mosaic Asian Anthropogenic Emission Inventory under the International Collaboration Framework of the MICS-Asia and HTAP." *Atmospheric Chemistry and Physics* 17 (2): 935–63. https://doi.org/10.5194/acp-17-935-2017.

Liao, Hong. 2005. "Global Impacts of Gas-Phase Chemistry-Aerosol Interactions on Direct Radiative Forcing by Anthropogenic Aerosols and Ozone." *Journal of Geophysical Research* 110 (D18): D18208. https://doi.org/10.1029/2005JD005907.

Lin, Jintai, Da Pan, Steven J. Davis, Qiang Zhang, Kebin He, Can Wang, David G. Streets, Donald J. Wuebbles, and Dabo Guan. 2014. "China's International Trade and Air Pollution in the United States." *Proceedings of the National Academy of Sciences* 111 (5): 1736–41. https://doi.org/10.1073/pnas.1312860111.

Liu, Fei, Ronald J. van der A, Henk Eskes, Jieying Ding, and Bas Mijling. 2018. "Evaluation of Modeling NO2 Concentrations Driven by Satellite-Derived and Bottom-up Emission Inventories Using in Situ Measurements over China." *Atmospheric Chemistry and Physics* 18 (6): 4171–86. https://doi.org/10.5194/acp-18-4171-2018.

Liu, Fei, Steffen Beirle, Qiang Zhang, Ronald J. van der A, Bo Zheng, Dan Tong, and Kebin He. 2017. "NOx Emission Trends over Chinese Cities Estimated from OMI Observations during 2005 to 2015." *Atmospheric Chemistry and Physics* 17 (15): 9261–75. https://doi.org/10.5194/acp-17-9261-2017.

Liu, Fei, Qiang Zhang, Ronald J van der A, Bo Zheng, Dan Tong, Liu Yan, Yixuan Zheng, and Kebin He. 2016. "Recent Reduction in NOx Emissions over China: Synthesis of Satellite Observations and Emission Inventories." *Environmental Research Letters* 11 (11): 114002. https://doi.org/10.1088/1748-9326/11/11/114002.

Lorenc, A. C. 1986. "Analysis Methods for Numerical Weather Prediction." *Quarterly Journal of the Royal Meteorological Society* 112 (474): 1177–94. https://doi.org/10.1002/qj.49711247414.

Lorente, Alba, K. Folkert Boersma, Huan Yu, Steffen Dörner, Andreas Hilboll, Andreas Richter, Mengyao Liu, et al. 2017. "Structural Uncertainty in Air Mass Factor Calculation for NO2 and HCHO Satellite Retrievals." *Atmospheric Measurement Techniques* 10 (3): 759–82. https://doi.org/10.5194/amt-10-759-2017.

Lu, Xi, Shaojun Zhang, Jia Xing, Yunjie Wang, Wenhui Chen, Dian Ding, Ye Wu, Shuxiao Wang, Lei Duan, and Jiming Hao. 2020. "Progress of Air Pollution Control in China and Its Challenges and Opportunities in the Ecological Civilization Era." *Engineering* 6 (12): 1423–31. https://doi.org/10.1016/j.eng.2020.03.014.

Lu, Xiao, Xingpei Ye, Mi Zhou, Yuanhong Zhao, Hongjian Weng, Hao Kong, Ke Li, et al. 2021. "The Underappreciated Role of Agricultural Soil Nitrogen Oxide Emissions in Ozone Pollution Regulation in North China." *Nature Communications* 12 (1): 5021. https://doi.org/10.1038/s41467-021-25147-9.

Lynch, Peter. 2008. "The Origins of Computer Weather Prediction and Climate Modeling." *Journal of Computational Physics* 227 (7): 3431–44. https://doi.org/10.1016/j.jcp.2007.02.034.

Marenco, Alain, Hervé Gouget, Philippe Nédélec, Jean-Pierre Pagés, and Fernand Karcher. 1994. "Evidence of a Long-Term Increase in Tropospheric Ozone from Pic Du Midi Data Series: Consequences: Positive Radiative Forcing." *Journal of Geophysical Research* 99 (D8): 16617. https://doi.org/10.1029/94JD00021.

Martin, Randall V. 2003. "Global Inventory of Nitrogen Oxide Emissions Constrained by Space-Based Observations of NO2 Columns." *Journal of Geophysical Research* 108 (D17): 4537. https://doi.org/10.1029/2003JD003453.

Maybeck, P. S., and George M. Siouris. 1980. "Stochastic Models, Estimation, and Control, Volume I." *IEEE Transactions on Systems, Man, and Cybernetics* 10 (5): 282–282. https://doi.org/10.1109/TSMC.1980.4308494.

Menut, L., B. Bessagnet, D. Khvorostyanov, M. Beekmann, N. Blond, A. Colette, I. Coll, et al. 2013. "CHIMERE 2013: A Model for Regional Atmospheric Composition Modelling." *Geoscientific Model Development* 6 (4): 981–1028. https://doi.org/10.5194/gmd-6-981-2013.

Menut, L., R. Vautard, M. Beekmann, and C. Honoré. 2000. “Sensitivity of Photochemical Pollution Using the Adjoint of a Simplified Chemistry-Transport Model.” *Journal of Geophysical Research: Atmospheres* 105 (D12): 15379–402. https://doi.org/10.1029/1999JD900953.

Menut, Laurent. 2003. “Adjoint Modeling for Atmospheric Pollution Process Sensitivity at Regional Scale.” *Journal of Geophysical Research* 108 (D17): 8562. https://doi.org/10.1029/2002JD002549.

Menut, Laurent, Bertrand Bessagnet, Guillaume Siour, Sylvain Mailler, Romain Pennel, and Arineh Cholakian. 2020. “Impact of Lockdown Measures to Combat Covid-19 on Air Quality over Western Europe.” *Science of The Total Environment* 741 (November): 140426. https://doi.org/10.1016/j.scitotenv.2020.140426.

Merlaud, A., M. Van Roozendael, N. Theys, C. Fayt, C. Hermans, B. Quennehen, A. Schwarzenboeck, et al. 2011. “Airborne DOAS Measurements in Arctic: Vertical Distributions of Aerosol Extinction Coefficient and NO2 Concentration.” *Atmospheric Chemistry and Physics* 11 (17): 9219–36. https://doi.org/10.5194/acp-11-9219-2011.

Middleton, P, W.R. Stockwell, and W.P.L Carter. 1990. “Aggregation and Analysis of Volatile Organic Compound Emissions for Regional Modeling.” *Atmospheric Environment* Part A, General Topics (24): 1107–33.

Mijling, B., and R. J. van der A. 2012. “Using Daily Satellite Observations to Estimate Emissions of Short-Lived Air Pollutants on a Mesoscopic Scale.” *Journal of Geophysical Research: Atmospheres* 117 (D17): n/a-n/a. https://doi.org/10.1029/2012JD017817.

Mitchell, John F. B. 1989. “The ‘Greenhouse’ Effect and Climate Change.” *Reviews of Geophysics* 27 (1): 115. https://doi.org/10.1029/RG027i001p00115.

Miyazaki, Kazuyuki, Henk Eskes, Kengo Sudo, K. Folkert Boersma, Kevin Bowman, and Yugo Kanaya. 2017. “Decadal Changes in Global Surface NOx Emissions from Multi-Constituent Satellite Data Assimilation.” *Atmospheric Chemistry and Physics* 17 (2): 807–37. https://doi.org/10.5194/acp-17-807-2017.

Molteni, Ugo, Federico Bianchi, Felix Klein, Imad El Haddad, Carla Frege, Michel J. Rossi, Josef Dommen, and Urs Baltensperger. 2018. “Formation of Highly Oxygenated Organic Molecules from Aromatic Compounds.” *Atmospheric Chemistry and Physics* 18 (3): 1909–21. https://doi.org/10.5194/acp-18-1909-2018.

Müller, J.-F., and T. Stavrakou. 2005. “Inversion of CO and NOx Emissions Using the Adjoint of the IMAGES Model.” *Atmospheric Chemistry and Physics* 5 (5): 1157–86. https://doi.org/10.5194/acp-5-1157-2005.

Napelenok, S. L., R. W. Pinder, A. B. Gilliland, and R. V. Martin. 2008. “A Method for Evaluating Spatially-Resolved NOx Emissions Using Kalman Filter Inversion, Direct Sensitivities, and

Space-Based NO2 Observations." *Atmospheric Chemistry and Physics* 8 (18): 5603–14. https://doi.org/10.5194/acp-8-5603-2008.

NASA/Goddard Space Flight Center. 2009. "Lightning's 'NOx-Ious' Impact On Pollution, Climate." ScienceDaily. November 6, 2009. www.sciencedaily.com/releases/2009/10/091030100022.htm.

Ng, N L, P S Chhabra, A W H Chan, J D Surratt, J H Kroll, A J Kwan, D C McCabe, et al. 2007. "Effect of NOx Level on Secondary Organic Aerosol (SOA) Formation from the Photooxidation of Terpenes." *Atmos. Chem. Phys.* 7: 5159–74. https://doi.org/www.atmos-chem-phys.net/7/5159/2007/.

"Notice of the State Council on Printing and Distributing the '13th Five-Year' Ecological and Environmental Protection Plan_Government Information Disclosure Column." 2016. 2016. http://www.gov.cn/zhengce/content/2016-12/05/content_5143290.htm.

Orlando, John J., and Geoffrey S. Tyndall. 2012. "Laboratory Studies of Organic Peroxy Radical Chemistry: An Overview with Emphasis on Recent Issues of Atmospheric Significance." *Chemical Society Reviews* 41 (19): 6294. https://doi.org/10.1039/c2cs35166h.

Osani, Moses, and Stephanie Haysmith. 2023. "Ozone Layer Recovery Is on Track, Helping Avoid Global Warming by 0.5°C." *UNEP*, September 1, 2023. https://www.unep.org/news-and-stories/press-release/ozone-layer-recovery-track-helping-avoid-global-warming-05degc#:~:text=On%20track%20to%20full%20recovery&text=If%20current%20policies%20remain%20in,the%20rest%20of%20the%20world.

Owens, R, and Tim Hewson. 2018. "ECMWF Forecast User Guide." https://doi.org/10.21957/M1CS7H.

Palmer, Paul I. 2017. *The Atmosphere: A Very Short Introduction*. First edition. Very Short Introductions 518. Oxford, United Kingdom: Oxford University Press.

Paraschiv, Spiru, Daniel-Eduard Constantin, Simona-Lizica Paraschiv, and Mirela Voiculescu. 2017. "OMI and Ground-Based In-Situ Tropospheric Nitrogen Dioxide Observations over Several Important European Cities during 2005–2014." *International Journal of Environmental Research and Public Health* 14 (11): 1415. https://doi.org/10.3390/ijerph14111415.

Peng, Weihan, Chen Le, William C. Porter, and David R. Cocker. 2022. "Variability in Aromatic Aerosol Yields under Very Low NOx Conditions at Different HO2/RO2 Regimes." *Environmental Science & Technology* 56 (2): 750–60. https://doi.org/10.1021/acs.est.1c04392.

Permadi, Didin Agustian, Nguyen Thi Kim Oanh, and Robert Vautard. 2018. "Assessment of Emission Scenarios for 2030 and Impacts of Black Carbon Emission Reduction Measures on Air Quality and Radiative Forcing in Southeast Asia." *Atmospheric Chemistry and Physics* 18 (5): 3321–34. https://doi.org/10.5194/acp-18-3321-2018.

Petetin, Hervé, Dene Bowdalo, Albert Soret, Marc Guevara, Oriol Jorba, Kim Serradell, and Carlos Pérez García-Pando. 2020. "Meteorology-Normalized Impact of COVID-19 Lockdown upon NO2 Pollution in Spain." Preprint. Gases/Atmospheric Modelling/Troposphere/Chemistry (chemical composition and reactions). https://doi.org/10.5194/acp-2020-446.

Phillips, Norman A. 1982. "On the Completeness of Multi-Variate Optimum Interpolation for Large-Scale Meteorological Analysis." *Monthly Weather Review* 110 (10): 1329–34. https://doi.org/10.1175/1520-0493(1982)110<1329:OTCOMV>2.0.CO;2.

Pison, Isabelle, Laurent Menut, and Gilles Bergametti. 2007. "Inverse Modeling of Surface NOx Anthropogenic Emission Fluxes in the Paris Area during the Air Pollution Over Paris Region (ESQUIF) Campaign." *Journal of Geophysical Research* 112 (D24): D24302. https://doi.org/10.1029/2007JD008871.

Platt, Ulrich, and Jochen Stutz. 2008. *Differential Optical Absorption Spectroscopy*. Physics of Earth and Space Environments. Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-75776-4.

Portmann, R. W., J. S. Daniel, and A. R. Ravishankara. 2012. "Stratospheric Ozone Depletion Due to Nitrous Oxide: Influences of Other Gases." *Philosophical Transactions of the Royal Society B: Biological Sciences* 367 (1593): 1256–64. https://doi.org/10.1098/rstb.2011.0377.

Pseftogkas, Andreas, Maria-Elissavet Koukouli, Arjo Segers, Astrid Manders, Jos van Geffen, Dimitris Balis, Charikleia Meleti, Trissevgeni Stavrakou, and Henk Eskes. 2022. "Comparison of S5P/TROPOMI Inferred NO2 Surface Concentrations with In Situ Measurements over Central Europe." *Remote Sensing* 14 (19): 4886. https://doi.org/10.3390/rs14194886.

Qu, Zhen, Daven K. Henze, Shannon L. Capps, Yi Wang, Xiaoguang Xu, Jun Wang, and Martin Keller. 2017. "Monthly Top-down NOx Emissions for China (2005–2012): A Hybrid Inversion Method and Trend Analysis." *Journal of Geophysical Research: Atmospheres* 122 (8): 4600–4625. https://doi.org/10.1002/2016JD025852.

Qu, Zhen, Daven K. Henze, Helen M. Worden, Zhe Jiang, Benjamin Gaubert, Nicolas Theys, and Wei Wang. 2022. "Sector-Based Top-Down Estimates of NOx, SO2, and CO Emissions in East Asia." *Geophysical Research Letters* 49 (2): e2021GL096009. https://doi.org/10.1029/2021GL096009.

Rabier, F., A. McNally, E. Andersson, P. Courtier, P. Undén, J. Eyre, A. Hollingsworth, and F. Bouttier. 1998. "The ECMWF Implementation of Three-Dimensional Variational Assimilation (3D-

Var). II: Structure Functions." *Quarterly Journal of the Royal Meteorological Society* 124 (550): 1809–29. https://doi.org/10.1002/qj.49712455003.

Rayner, Peter J., Anna M. Michalak, and Frédéric Chevallier. 2019. "Fundamentals of Data Assimilation Applied to Biogeochemistry." *Atmospheric Chemistry and Physics* 19 (22): 13911–32. https://doi.org/10.5194/acp-19-13911-2019.

Richardson, Lewis Fry. 1922. *Weather Prediction by Numerical Process*. Cambridge, The University press. http://archive.org/details/weatherpredictio00richrich.

Richter, Andreas, John P. Burrows, Hendrik Nüß, Claire Granier, and Ulrike Niemeier. 2005. "Increase in Tropospheric Nitrogen Dioxide over China Observed from Space." *Nature* 437 (7055): 129–32. https://doi.org/10.1038/nature04092.

Rodgers, Clive D. 2000. *Inverse Methods for Atmospheric Sounding - Theory and Practice*. Vol. 2. Atmospheric, Oceanic and Planetary Physics. University of Oxford: World Scientific Publishing Co. Pte. Ltd. https://library.wmo.int/doc_num.php?explnum_id=1654.

Schoeberl, M.R., A.R. Douglass, E. Hilsenrath, P.K. Bhartia, R. Beer, J.W. Waters, M.R. Gunson, et al. 2006. "Overview of the EOS Aura Mission." *IEEE Transactions on Geoscience and Remote Sensing* 44 (5): 1066–74. https://doi.org/10.1109/TGRS.2005.861950.

Schumann, U., and H. Huntrieser. 2007. "The Global Lightning-Induced Nitrogen Oxides Source." *Atmospheric Chemistry and Physics* 7 (14): 3823–3907. https://doi.org/10.5194/acp-7-3823-2007.

Seinfeld, John H., and Spyros N. Pandis. 2016. *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change*. John Wiley & Sons, Hoboken.

Shah, Viral, Daniel J. Jacob, Ke Li, Rachel F. Silvern, Shixian Zhai, Mengyao Liu, Jintai Lin, and Qiang Zhang. 2020. "Effect of Changing NOx Lifetime on the Seasonality and Long-Term Trends of Satellite-Observed Tropospheric NO2 Columns over China." *Atmospheric Chemistry and Physics* 20 (3): 1483–95. https://doi.org/10.5194/acp-20-1483-2020.

Shaw, Stephanie, and Bill Van Heyst. 2022. "Nitrogen Oxide (NOx) Emissions as an Indicator for Sustainability." *Environmental and Sustainability Indicators* 15 (September): 100–188. https://doi.org/10.1016/j.indic.2022.100188.

Shi, Yun, Yin-feng Xia, Bi-hong Lu, Nan Liu, Lei Zhang, Su-jing Li, and Wei Li. 2014. "Emission Inventory and Trends of NOx for China, 2000–2020." *Journal of Zhejiang University SCIENCE A* 15 (6): 454–64. https://doi.org/10.1631/jzus.A1300379.

Skiba, Ute, Sergiy Medinets, Laura M. Cardenas, Edward John Carnell, Nick Hutchings, and Barbara Amon. 2021. "Assessing the Contribution of Soil NOx Emissions to European Atmospheric Pollution." *Environmental Research Letters*, February. https://doi.org/10.1088/1748-9326/abd2f2.

Solomon, Susan, Rolando R. Garcia, F. Sherwood Rowland, and Donald J. Wuebbles. 1986. “On the Depletion of Antarctic Ozone.” *Nature* 321 (6072): 755–58. https://doi.org/10.1038/321755a0.

Souri, Amir H., Kelly Chance, Juseon Bak, Caroline R. Nowlan, Gonzalo González Abad, Yeonjin Jung, David C. Wong, Jingqiu Mao, and Xiong Liu. 2021. “Unraveling Pathways of Elevated Ozone Induced by the 2020 Lockdown in Europe by an Observationally Constrained Regional Model Using TROPOMI.” *Atmospheric Chemistry and Physics* 21 (24): 18227–45. https://doi.org/10.5194/acp-21-18227-2021.

Souri, Amir H., Caroline R. Nowlan, Gonzalo González Abad, Lei Zhu, Donald R. Blake, Alan Fried, Andrew J. Weinheimer, et al. 2020. “An Inversion of NOx and Non-Methane Volatile Organic Compound (NMVOC) Emissions Using Satellite Observations during the KORUS-AQ Campaign and Implications for Surface Ozone over East Asia.” *Atmospheric Chemistry and Physics* 20 (16): 9837–54. https://doi.org/10.5194/acp-20-9837-2020.

Stigler, Stephen M. 1981. “Gauss and the Invention of Least Squares.” *The Annals of Statistics* 9 (3). https://doi.org/10.1214/aos/1176345451.

Stordal, F., Slimane Bekki, D. Hauglustaine, M. Millan, R. Sausen, E. Schuepbach, D. Stevenson, R. van Dorland, and A. Volz-Thomas. 2003. “Climate Impact of Tropospheric Ozone Changes.” In *Air Pollution Report N°81: Ozone-Climate Interactions. EU 20623*, edited by Isaksen I.S. (Ed.), 73–95. chapter 4. European Communities. https://hal.archives-ouvertes.fr/hal-03327712.

Szopa, S., Vaishali Naik, B Adhikary, P. Artaxo, Terje Berntsen, W.D Collins, S. Fuzzi, et al. 2021. “2021: Short-Lived Climate Forcers. Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change.” [Masson-Delmotte, V., P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (Eds.)]. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press.

Talagrand, Olivier. 1997. “Assimilation of Observations, an Introduction (Special Issuel: Data Assimilation in Meteology and Oceanography: Theory and Practice).” *Journal of the Meteorological Society of Japan. Ser. II* 75 (1B): 191–209. https://doi.org/10.2151/jmsj1965.75.1B_191.

Talagrand, Olivier, and Philippe Courtier. 1987. “Variational Assimilation of Meteorological Observations With the Adjoint Vorticity Equation. I: Theory: VARIATIONAL ASSIMILATION. I: THEORY.” *Quarterly Journal of the Royal Meteorological Society* 113 (478): 1311–28. https://doi.org/10.1002/qj.49711347812.

Talbot, R. 2003. “Reactive Nitrogen in Asian Continental Outflow over the Western Pacific: Results from the NASA Transport and Chemical Evolution over the Pacific (TRACE-P) Airborne Mission.” *Journal of Geophysical Research* 108 (D20): 8803. https://doi.org/10.1029/2002JD003129.

Tarantola, Albert. 2005. *Inverse Problem Theory and Methods for Model Parameter Estimation*. Philadelphia, PA: Society for Industrial and Applied Mathematics.

Tikhonov, A. N. 1963. “Solution of Incorrectly Formulated Problems and the Regularization Method.” *Dokl. Akad. Nauk SSSR*, no. 151: 501–4.

United Nations Economic and Social Council. 2015. “Guidance Document on Control Techniques for Emissions of Sulphur, Nitrogen Oxides, Volatile Organic Compounds and Particulate Matter (Including PM10, PM2.5 and Black Carbon) from Stationary Sources*.” Executive Body for the Convention on Long-range Transboundary Air Pollution. https://unece.org/sites/default/files/2021-04/ENG_G1500924.pdf.

Van Geffen, Jos, Henk Eskes, K. Folkert Boersma, Joannes D. Maasakkers, and J. Pepijn Veefkind. 2019. *TROPOMI ATBD of the Total and Tropospheric NO2 Data Products*. 1.4.0. Royal Netherlands Meteorological Institute (KNMI).

Verhoelst, Tijl, Steven Compernolle, Gaia Pinardi, Jean-Christopher Lambert, Henk J. Eskes, Kai-Uwe Eichmann, Ann Mari Fjæraa, et al. 2021. “Ground-Based Validation of the Copernicus Sentinel-5P TROPOMI NO2 Measurements with the NDACC ZSL-DOAS, MAX-DOAS and Pandonia Global Networks.” *Atmospheric Measurement Techniques* 14 (1): 481–510. https://doi.org/10.5194/amt-14-481-2021.

Vestreng, V, L Ntziachristos, A Semb, S Reis, I S A Isaksen, and L Tarrason. 2009. “Evolution of NOx Emissions in Europe with Focus on Road Transport Control Measures.” *Atmos. Chem. Phys.*, 18.

Vinken, G. C. M., K. F. Boersma, J. D. Maasakkers, M. Adon, and R. V. Martin. 2014. “Worldwide Biogenic Soil NOx Emissions Inferred from OMI NO2 Observations.” *Atmospheric Chemistry and Physics* 14 (18): 10363–81. https://doi.org/10.5194/acp-14-10363-2014.

Vries, Wim de. 2021. “Impacts of Nitrogen Emissions on Ecosystems and Human Health: A Mini Review.” *Current Opinion in Environmental Science & Health* 21 (June): 100249. https://doi.org/10.1016/j.coesh.2021.100249.

Wellemeyer, C. G., S. L. Taylor, C. J. Seftor, R. D. McPeters, and P. K. Bhartia. 1997. “A Correction for Total Ozone Mapping Spectrometer Profile Shape Errors at High Latitude.” *Journal of Geophysical Research: Atmospheres* 102 (D7): 9029–38. https://doi.org/10.1029/96JD03965.

WHO Global Air Quality Guidelines. 2021. Geneva: WHO.

World Meteorological Organization (WMO). 2018. “Executive Summary: Scientific Assessment of Ozone Depletion: 2018.” Global Ozone Research and Monitoring Project–Report No. 58. Geneva, Switzerland. https://csl.noaa.gov/assessments/ozone/2018/executivesummary/.

Xia, Yinmin, Yu Zhao, and Chris P. Nielsen. 2016. “Benefits of China’s Efforts in Gaseous Pollutant Control Indicated by the Bottom-up Emissions and Satellite Observations 2000–2014.” *Atmospheric Environment* 136 (July): 43–53. https://doi.org/10.1016/j.atmosenv.2016.04.013.

Zara, Marina, K. Folkert Boersma, Isabelle De Smedt, Andreas Richter, Enno Peters, Jos H. G. M. van Geffen, Steffen Beirle, et al. 2018. “Improved Slant Column Density Retrieval of Nitrogen Dioxide and Formaldehyde for OMI and GOME-2A from QA4ECV: Intercomparison, Uncertainty Characterisation, and Trends.” *Atmospheric Measurement Techniques* 11 (7): 4033–58. https://doi.org/10.5194/amt-11-4033-2018.

Zhao, Y., C. P. Nielsen, Y. Lei, M. B. McElroy, and J. Hao. 2011. “Quantifying the Uncertainties of a Bottom-up Emission Inventory of Anthropogenic Atmospheric Pollutants in China.” *Atmospheric Chemistry and Physics* 11 (5): 2295–2308. https://doi.org/10.5194/acp-11-2295-2011.

Zhao, Y., J. Zhang, and C. P. Nielsen. 2013. “The Effects of Recent Control Policies on Trends in Emissions of Anthropogenic Atmospheric Pollutants and CO2 in China.” *Atmospheric Chemistry and Physics* 13 (2): 487–508. https://doi.org/10.5194/acp-13-487-2013.

Zhao, Y., J. Zhang, and C. P. Nielsen. 2014. “The Effects of Energy Paths and Emission Controls and Standards on Future Trends in China’s Emissions of Primary Air Pollutants.” *Atmospheric Chemistry and Physics* 14 (17): 8849–68. https://doi.org/10.5194/acp-14-8849-2014.

Zheng, Bo, Jing Cheng, Guannan Geng, Xin Wang, Meng Li, Qinren Shi, Ji Qi, Yu Lei, Qiang Zhang, and Kebin He. 2021. “Mapping Anthropogenic Emissions in China at 1 Km Spatial Resolution and Its Application in Air Quality Modeling.” *Science Bulletin* 66 (6): 612–20. https://doi.org/10.1016/j.scib.2020.12.008.

Zheng, Bo, Dan Tong, Meng Li, Fei Liu, Chaopeng Hong, Guannan Geng, Haiyan Li, et al. 2018. “Trends in China’s Anthropogenic Emissions since 2010 as the Consequence of Clean Air Actions.” *Atmospheric Chemistry and Physics* 18 (19): 14095–111. https://doi.org/10.5194/acp-18-14095-2018.

Table of Contents

Table of figures

List of Tables