Mitigating the impact of chemical transport model biases on top-down CO and NO_X emission estimates using multi-species chemical data assimilation

by

Xuesong Zhang

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy
Graduate Department of Physics
University of Toronto

Abstract

Mitigating the impact of chemical transport model biases on top-down CO and NO_x emission estimates using multi-species chemical data assimilation

Xuesong Zhang
Doctor of Philosophy
Graduate Department of Physics
University of Toronto
2019

Chemical transport models such as GEOS-Chem simulate tropospheric CO-HO_x-NO_x-O_3 chemistry, which determines the oxidizing power of the atmosphere as well as the abundance of tropospheric CO, NO_2, and O_3. Inverse modelling analyses to estimate CO and NO_x emissions traditionally rely only on observations of the species of interest, neglecting chemical biases in the models employed in the analyses. However, top-down estimates of CO and NO_x emissions are sensitive to tropospheric OH and O_3, respectively. In this thesis, two approaches were used to mitigate the model biases to estimate CO and NO_x emission. First, we developed a multi-species data assimilation framework using the GEOS-Chem 4D-Var data assimilation system. Satellite observations of CO, NO_2, O_3, HCHO, and HNO_3 were assimilated to optimize CO and NO_x emissions, as well as O_3, HCHO and HNO_3 initial conditions. After the assimilation, improved chemical states were produced with CO, NO_2, and O_3 fields that were in closer agreement with independent observations. The multi-species DA also provided a means of indirectly constraining the tropospheric OH. The resulting global mean tropospheric OH was more consistent with independent empirically derived OH estimates. We also examined the a posteriori CO and NO_x emissions. During the study period, my thesis showed the a priori CO emissions were underestimated and the a priori NO_x emissions were overestimated in the northern extratropics. The emission estimates in Europe, the Amazon, and equatorial Africa were most sensitive to the multi-species information assimilated

into GEOS-Chem. Second, we assessed the utility of a new OMI-minus-OSIRIS (OmO) NO_2 retrieval product for improving North American NO_x emission estimates. We assimilated the OmO data using the nested version of GEOS-Chem with assimilated O_3 boundary conditions. The results suggested that both the anthropogenic and biomass burning sources were significantly overestimated in the a priori. The changes in the a posteriori NO_x emissions inferred from the OmO data between 2008 and 2011 were found to be in better agreement with the changes in independent NO_2 observations than those estimated from OMI data.

Acknowledgements

This thesis project was funded by the Canadian Space Agency (CSA), the Natural Sciences and Engineering Research Council of Canada (NSERC), and Environment and Climate Change Canada (ECCC).

I would like to express my appreciation to Dylan Jones for supervising my thesis research. I am very thankful to receive countless thoughtful ideas from him not only in research but also in everyday life. My Ph.D. experience would not be as enjoyable without Dylan supporting me throughout this thesis project, encouraging me to solve challenging scientific problems independently, helping me improve my oral and writting skills, and funding me to attend various academic conferences to present the thesis project to a broader audience. I would also like to thank Kazuyuki Miyazaki, Cristen Adams, Jennifer Kaiser and Chris Sioris for offering me opportunity to collaborate with them on research topics not limited to this thesis project. I obtained key knowledge via many insightful discussions with them. Taking this opportunity, I would also like to show my appreciation to my committee member Paul Kushner, and Kaley Walker. I received my first research internship with Paul, who later on convinced me to work in the field of atmospheric science. I also benefited from Kaley on understanding key features of the employed observations in this research.

This thesis project is also heavily reliant on the GEOS-Chem model and various satellite and in-situ observations provided by many atmospheric physics and chemistry scientists. I would like to thank Martin Keller, Feng Deng, Jiang Zhe and Thomas Walker for helping me with setting up the early phase data assimilation framework and designing the observation operators. I would also like to express my gratitude to my co-authors as well as code and data providers: Daven Henze, Adam Bourassa, Doug Degenstein, Helen Worden, Cathy Clerbaux, Yves Rochon, Lok Lamsal, Eloise Marais, Steve Wofsy, Daniel Jacob, Anne Thompson, the Earth Observing System Data and Information System (EOS DIS), Tropospheric Emission Monitoring Internet Service (TEMIS), the United States Environmental Protection Agency Air Quality System (EPA AQS) program, and the National Air Pollution Surveillance (NAPS) program. Prensentations of the thesis material were awarded by the 36th International Technical Meeting on Air Pollution modelling and its application, and the joint 14th International Commission on Atmospheric Chemistry and Global Pollution (iCACGP) Symposium and 15th Internatial Global Atmospheric Chemistry (IGAC) Science Conference. I would like to thank the committees from the ITM, the joint iCACGP and IGAC for their awards.

I had a grateful eight years with the Earth, Atmospheric and Planetary Physics group since I was a second year undergraduate. I want to sincerely thank everyone in the group who have helped me pursue my academic career up to this point. Many enjoyable interactions with the group will long last in my memories. I would also like to thank Ana Sousa, Krystyna Biel, April Seely, Peter Hurley and everyone from the Physics Undergraduate Teaching Group for helping with my administrative and teaching assistant duties, with a special shoutout to Ruxandra Sebanescu and Larry Avramidis, who have spent every Tuesday and Thursday mornings with me teaching Physics labs over the last four years. I would like to thank all my roommates, friends and everyone in Tiebreakers and U of T St George D-League basketball team for accepting nothing less than excellence from me.

Finally, I would like to dedicate this thesis to my parents: Mingli Wang and Yonggen Zhang, and to all my great family members. The love, encouragement and support I receive from them give me the passion, inspiration and dedication in my life.

Contents

List of Tables

List of Figures

List of Acronyms and Abbreviations

CH_3	Methyl Radical
CH_3CCl_3	Methyl Chloroform
CH_3O	Methoxyl Radical
CH_3O_2	Methylperoxy Radical
CH_4	Methane
CO	Carbon Monoxide
CO_2	Carbon Dioxide
HNO_3	Nitric Acid
$HCHO/H_2CO/CH_2O$	Formaldehyde
HO_2	Hydroperoxyl Radical
HO_x	Hydrogen Oxide Radical, $HO_x \equiv H+OH+HO_2$
LNO_x	Lightning produced Nitrogen Oxides
MCF	Methyl Chloroform
NH_3	Ammonia
NO	Nitrogen Monoxide
NO_2	Nitrogen Dioxide
NO_3	Nitrate Radical
NO_x	Reactive Nitrogen Oxides, $NO_x \equiv NO + NO_2$
NO_y	Reactive Nitrogen Compounds, $NO_y \equiv NO + NO_2 + NO_z$
NO_z	Other Nitrogen Oxides than NO_x
OH	Hydroxyl Radical

O	Atomic Oxygen
O_2	Oxygen
O_3	Ozone
RO_2	Organic Peroxy Radical
3D-Var	Three-dimensional variational data assimilation
4D-Var	Four-dimensional variational data assimilation
ACCMIP	Atmospheric Chemistry and Climate Model Intercomparison Project
ACE-FTS	Atmospheric Chemistry Experiment-Fourier Transform Spectrometer
AIRS	Atmospheric Infrared Sounder
AK	Averaging Kernel
AMF	Air Mass Factor
AQS	Air Quality System
ARCTAS	Arctic Research of the Composition of the Troposphere from Aircraft and Satellites
ATHOS	Airborne Tropospheric Hydrogen Oxides Sensor
ATom	Atmospheric Tomography Mission
BC	Boundary condition
BOAS	Basic Optical Absorption Spectroscopy
BRAVO	Big Bend Regional Aerosol and Visibility Observational Study Emissions Inventory
CAC	Criteria Air Contaminants (Canadian emission inventory)
CCM	Chemistry and Climate Model
CHASER	CHemical Atmospheric general circulation model for Study of atmospheric Environment and Radiative forcing
CrIS	Cross-track Infrared Sounder
CSA	Canadian Space Agency

CTM	Chemical transport model
CMDL	Climate Monitoring and Diagnostics Laboratory
DAS	Data Assimilation System
DIS	Data and Information System
DJF	December-January-February
DMF	Column-averaged dry-air mole fractions
DOAS	Differential Optical Absorption Spectroscopy
DOFS	Degrees of freedom for signal
DOMINO/DP	Dutch OMI NO_2 retrievals
DU	Dobson Unit
ECCC	Environment and Climate Change Canada
ECMWF	European Centre for Medium-Range Weather Forecasts
ENSO	El Niño Southern Oscillation
EDGAR	Emission Database for Global Atmospheric Research
EMEP	Cooperative Programme for Monitoring and Evaluation of Long-Range Transmission of Air Pollutants in Europe
EnKF	Ensemble Kalman Filter
EOS	Earth Observing System
EPA	US Environmental Protection Agency
EUMETSAT	European Organisation for the Exploitation of Meteorological Satellites
FLEXPART	FLEXible PARTicle dispersion model
FORLI	Fast Optimal Retrievals on Layers for IASI
FOV	Field-of-view
FTIR	Fourier transform infrared
FWHM	Full Width at Half Maximum
GCM	General Circulation Model

GEO	Geostationary satellite
GEOS	Goddard Earth Observing System
GEMS	Geostationary Environmental Monitoring Spectrometer
GFED	Global Fire Emissions Database
GGG/GFIT	Software suite for TCCON retrievals
GHG	Greenhouse gases
GMAO	Global Modeling and data Assimilation Office
GOME	Global Ozone Monitoring Experiment
HIAPER	High-Performance Instrumented Airborne Platform for Environmental Research
HIPPO	HIAPER Pole-to-Pole Observations
HTAP	Hemispheric Transport of Air Pollution
IR	Infrared
IASI	Infrared Atmospheric Sounding Interferometer
IMAGES	Intermediate Model of Global Evolution of Atmospheric Species
IPCC	Intergovernmental Panel on Climate Change
ISAF	In-Situ Airborne Formaldehyde
ISOP	Isoprene
ITCZ	Inter-tropical convergence zone
JAMSTEC	Japan Agency for Marine-Earth Science and Technology
JAS	July-August-September
JJA	June-July-August
KF	Kalman filter
KPP	Kinetic PreProcessor
L-BFGS-B	Limited-memory Broyden-Fletcher-Goldfarb-Shanno algorithm for Bound-constrained optimization
LETKF	Local Ensemble Transform Kalman Filter

LINOZ	Linearization of stratospheric ozone
LMR	Length-Modulated Radiometer
MAM	March-April-May
MAP	Maximum A Posteriori
MART	Multiplicative Algebraic Reconstruction Technique
MEGAN	Model of Emissions of Gases and Aerosols from Nature
MetOp	Meteorological Operational Satellite
MLS	Microwave Limb Sounder
MOCAGE	Modéle de Chimie Atmosphérique de Grande Echelle (CTM developed by Meteo-France)
MOPITT	Measurements of Pollution in The Troposphere
MODIS	Moderate-resolution imaging spectroradiometer
MOZART	Model for Ozone And Related Tracers
MSA	Multi-Species chemical data Assimilation
NAPS	National Air Pollution Surveillance
NASA	National Aeronautics and Space Administration
NEI05/08	National Emission Inventory for 2005/2008
NH	Northern Hemisphere
NIR	Near Infrared
NMHC	Non-methane hydrocarbons
NMVOC	Non-Methane Volatile Organic Compound
NOAA	National Oceanic and Atmospheric Administration
NO_yO_3	NOAA Nitrogen Oxides and Ozone
NSERC	Natural Sciences and Engineering Research Council of Canada
OI	Optimal Interpolation
OMI	Ozone Monitoring Instrument
OmO	OMI-minus-OSIRIS

OMPS	Ozone Mapping and Profiler Suite
OSIRIS	Optical, Spectroscopic, and Infrared Remote Imaging System
OSSE	Observation System Simulation Experiment
OTD/LIS	Optical Transient Detector and the Lightning Imaging Sensor
PBL	Planetary boundary layer
pdf	Probability Density Function
PMR	Pressure-Modulated Radiometer
ppb	parts per billion
ppbv	parts per billion volume
pptv	parts per trillion volume
QCLS	Harvard Quantum Cascade Laser System
RMSE	Root-Mean-Square Error
RTM	Radiative Transfer Model
SAGE-II	Stratospheric Aerosol and Gas Experiment II
SAO	Smithsonian Astrophysical Observatory retrieval algorithm of OMI observations
SaskTRAN	A radiative transfer model developed by University of Saskatchewan
SCD	Slant Column Density
SCIAMACHY	SCanning Imaging Absorption Spectrometer for Atmospheric CHartographY
SF	Scaling Factor
SH	Southern Hemisphere
SHADOZ	Southern Hemisphere ADditional OZonesondes
SON	September-October-November
SP	OMI standard product
SPCZ	South Pacific Convergence Zone
STE	Stratosphere-troposphere exchange

STS	Stratosphere-tropospheric separation
SWIR	Shortwave infrared
SZA	Solar Zenith Angle
TCCON	Total Carbon Column Observing Network
TCR-2	Tropospheric Chemical Reanalysis version 2
TEMIS	Tropospheric Emission Monitoring Internet Service
TEMPO	Tropospheric Emissions: Monitoring POllution
TES	Tropospheric Emission Spectrometer
TIR	Thermal Infrared
TOMCAD	Toulouse Off-line Model of Chemistry And Transport
TOMRAD	Total Ozone Mapping spectrometer RADiative transfer model
TROPOMI	TROPOspheric Monitoring Instrument
UTLS	Upper troposphere lower stratosphere
UCATS	Unmanned Chromatograph for Atmospheric Trace Species
VCD	Vertical Column Density
VCSEL	Vertical Cavity Surface Emitting Lasers
VLIDORT	Vector-based Linearized Discrete Ordinate Radiative Transfer
VMR	Volume Mixing Ratio
VOC	Volatile Organic Compound
VZA	Viewing Zenith Angle
WC	Weak constraint

Chapter 1

Introduction

Carbon monoxide (CO), non-methane volatile organic compounds (NMVOCs), nitrogen oxides (NO_x=NO_2+NO), and ozone (O_3) are environmentally important trace gases. High concentrations of NO_x and O_3 lead to increases in premature mortality, reduction in photosynthesis and other adverse effects on physiological functions in vegetation. O_3 is also a greenhouse gas that affects the radiative forcing of the atmosphere. Many of these chemical compounds that are directly emitted to the atmosphere, such as CO and NMVOCs, are chemically removed by reaction with the hydroxyl radical (OH), which is the key oxidant in the atmosphere. In turn, the abundance of these species also affect OH, and thus the oxidative capacity of the troposphere. Background atmospheric CO concentrations have increased to more than two times pre-industrial values (*Haan et al.*, 1996). During this period, the NO_2 concentration has increased by a factor of 3-6 above the pre-industrial level (*Prather and Ehhalt*, 2001). As result of the increase of CO and NO_x, tropospheric ozone has also exhibited a positive trend over the last 30 years (*Cooper et al.*, 2014). Since tropospheric chemistry involving CO and NO_x is affected significantly by human activities, good estimates of CO and NO_x sources are essential to develop reliable air quality control strategies and to predict future changes to global tropospheric chemistry. In this context, there has been much effort during the past two decades on inverse modelling emissions of CO and NO_x.

Traditionally, inverse modelling studies have relied only on observations of the species of interest, such as CO, and used an atmospheric model as a transfer function to relate changes in the surface emissions of the species to changes in its the atmospheric abundance. Implicit in this approach is the assumption that the models are free of chemical and transport-related biases, and that any discrepancies between the model simulation of the tracer and the observations are due to errors in the specified emissions of the tracer in the model. In this way, discrepancies between the atmospheric observations

and the model simulation can be used to obtain an improved estimate of the emissions. However, the model biases can be large. In the case of CO, for example, discrepancies in the modelled OH as well as the source of CO from the oxidation of methane (CH_4) and NMVOCs can significantly impact the modelled distribution of CO. Shown in Fig. 1.1 are the results from *Shindell et al.* (2006) of a comparison of the CO simulation in 26 different models with CO observations from MOPITT at 500 hPa. Although the specified anthropogenic and biomass burning emissions were the same in all the models, there is a large spread in the CO abundances between the models, mostly due to the modelled OH fields and the differences in the treatment of the chemical oxidation of NMVOCs in the models. In Fig. 1.1, the multi-model mean (solid black line in the left column) suggested that the models significantly underestimated CO in the northern extratropics throughout the year compared to MOPITT observations. In an inverse modelling context, this would suggest that the a priori anthropogenic emissions in the models were biased low. However, the GEOS-Chem model, which was one of the participating models, was one of a small number of model biased high relative to MOPITT, due to low OH abundances. Use of that version of GEOS-Chem for inverse modelling of CO emissions would have suggested that the a priori emissions in the model were biased high. The results of *Shindell et al.* (2006) highlight the importance of accounting for discrepancies in OH and in the NMVOC source of CO when using models for inverse modelling.

The biases in the model chemistry can be mitigated by constraining modelled O_3, CO, NO_x and OH simultaneously, which leads to improved top down CO and NO_x emission estimates. In a recent inverse modelling study, *Jiang et al.* (2015a) tried to mitigate the impact of OH when estimating top down CO emissions. However, their approach for correcting the potential impact of biases in modelled OH was fairly ad-hoc. With more tropospheric composition monitoring instruments becoming available, multi-species chemical data assimilation is able to adjust emissions and concentrations simultaneously, and provide improved modelled states better characterizing the tropospheric chemistry. *Eibern and Schmidt* (1999) found that four-dimensional variational (4D-Var) approach is an effective smoother for obtaining chemically consistent modelled states that best match the observations. However, before proposing this thesis project, little work has been done to design an to design a multi-species chemical data assimilation (MSA) framework using the 4D-Var scheme for constraining the tropospheric chemistry. Considering all the issues and gaps addressed above, implementation and application of MSA using 4D-Var approach will have practical benefits to improve the modelled CO and NO_x emission estimates via mitigating the modelled biases on OH and O_3. The focus of this thesis is to integrate the growing wealth of space-based observations, in an

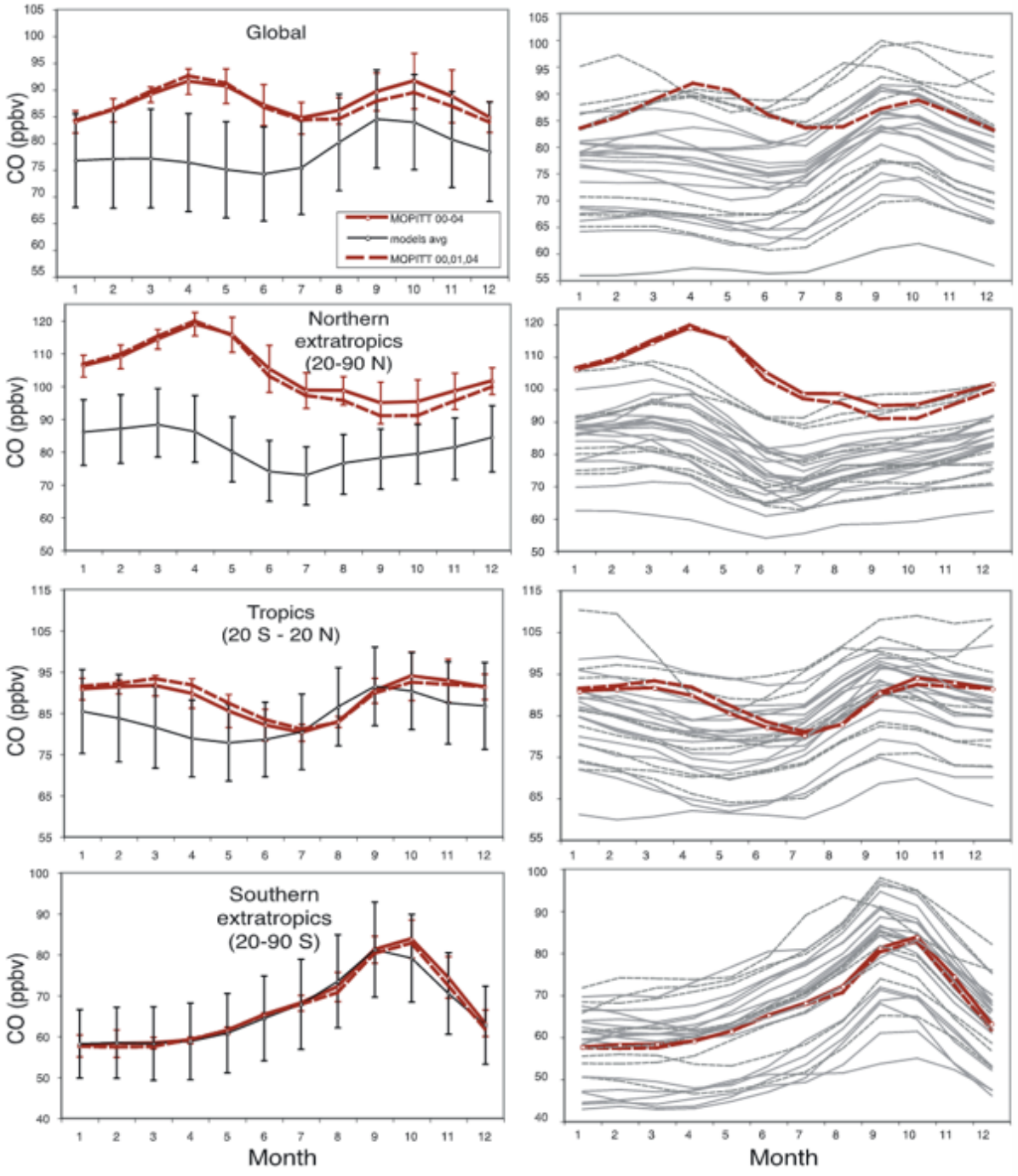

Figure 1.1: Annual cycle of CO (in ppbv) at 500 hPa in MOPITT observations and in the models over 2000-2004. The left panel shows the multi-model mean (shown in solid black) with its standard deviation and the MOPITT CO observations (shown in red). The full list of all the participated models were available in Table 1 from *Shindell et al.* (2006). The right panel shows the CO simulations from each individual model. Dashed red line shows MOPITT CO observations in 2002 and 2003. Dashed black line shows the model whose methane lifetime are outside of the IPCC Third Assessment Report range. (*Shindell et al.*, 2006)

MSA framework to mitigate the chemical biases in the GEOS-Chem model. MSA is able to adjust emissions and concentrations simultaneously, and provide improved modelled states that better capture the tropospheric chemistry.

1.1 Tropospheric CO-OH-NO_x-O_3 chemistry

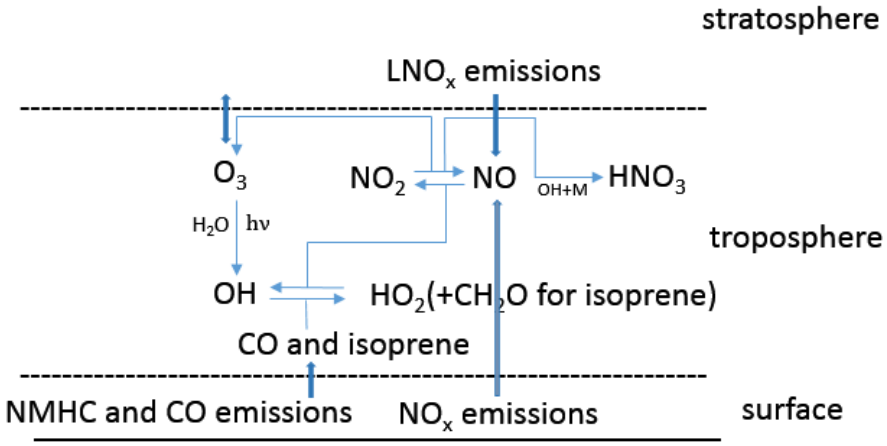

Figure 1.2: Mechanism for CO-OH-NO_x-O_3 chemistry cycle.

A general schematic of the key chemical processes in the CO-OH-NO_x-O_3 tropospheric chemistry is shown in Fig. 1.2. Tropospheric OH is generated as a result of the photolysis of O_3 (at wavelengths less than about 325 nm), which produces excited atomic oxygen that reacts with water vapour as follows:

$$O_3 + h\nu \rightarrow O_2 + O(^1D), \tag{1.1}$$

$$O(^1D) + H_2O \rightarrow 2OH. \tag{1.2}$$

O_3 can also produce OH via reaction with HO_2

$$O_3 + HO_2 \rightarrow OH + 2O_2. \tag{1.3}$$

It should be noted that this reaction converts HO_2 to OH. Because of the rapid interconversion of OH and HO_2, they are typically grouped together as the HO_x (= OH + HO_2) family. Thus, Reaction 1.3 is not a net source of HO_x, unlike Reaction 1.2. OH can also be recycled via reaction with NO

$$HO_2 + NO \rightarrow NO_2 + OH. \tag{1.4}$$

Under low NO_x conditions, recent atmospheric chemistry laboratory or modelling studies (*Rohrer et al.*, 2014; *Lelieveld et al.*, 2016; *Archibald et al.*, 2009) have identified secondary OH production mechanisms through OH recycling via radical reaction chains. For instance, *Kang et al.* (2004) proposed that HO_2 is able to react with organic peroxy radical (RO_2) under low-NO_x, high-NMVOC conditions:

$$RO_2 + HO_2 \rightarrow RO + OH + O_2. \tag{1.5}$$

OH is primarily removed by reaction with CO, which accounts for 40% of the total OH sink.

$$CO + OH \rightarrow CO_2 + H, \tag{1.6}$$

The H atom produced will then rapidly react with O_2 to form HO_2,

$$H + O_2 + M \rightarrow HO_2 + M. \tag{1.7}$$

Quantifying OH abundances has been a long-term objective of the community since it is important for understanding the changing oxidizing power of the atmosphere. Since the production of OH relies on sunlight and water vapour, tropical regions have the highest OH concentration (2×10^7 molec/cm^3) in the free troposphere (*Jacob*, 1999) as shown in Fig. 1.3. However, in regions under low NO_x conditions such as the Amazon and equatorial Africa, where OH is strongly titrated due to the presence of NMVOCs, the GEOS-Chem simulation suggests much lower OH concentrations (1×10^5 molec/cm^3). CTMs such as GEOS-Chem have become standard tools for improving knowledge of atmospheric chemical tracers and the physical and chemical processes that control them. Near the surface in the model, high OH concentrations are distributed over northern mid-latitudes in summer, which is associated with high levels of CO, NO_x and NMVOC emissions. Much lower OH concentrations are distributed over high latitudes in both hemispheres, as well as regions with low NO_x conditions such as the Amazon, equatorial Africa, and Southeast Asia. It is noted that the lifetime of tropospheric OH is extremely short, on the order of seconds. OH concentrations also exhibit an extreme diurnal cycle, with the highest concentrations during the day and values close to zero at night.

As shown in Table 1.1, atmospheric CO is generated by surface emissions from incomplete combustion as well as from the oxidation of atmospheric hydrocarbons. The lifetime of tropospheric CO ranges from one month (in the tropics and in the summer hemisphere) to one year (in the winter hemisphere). The abundance of atmospheric CO will impact the oxidative power of the troposphere (*Shindell et al.*, 2003). A month-long lifetime sug-

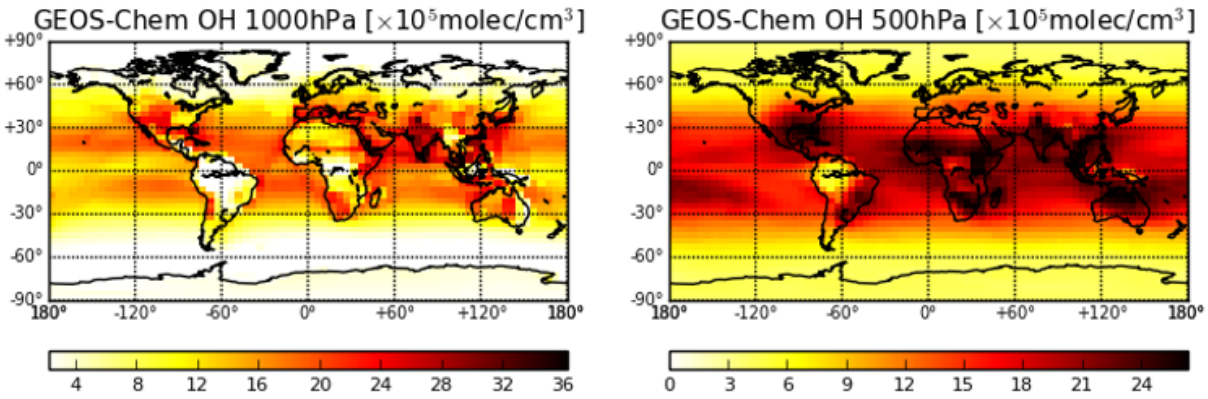

Figure 1.3: GEOS-Chem (version 9) simulated mass weighted mean OH concentrations (in 10^5 molec/cm^3) at the surface and 500hPa in 2016.

gests that the distribution of CO would be strongly influenced by atmospheric transport. Hence, CO is also an ideal tracer to characterize atmospheric transport in models (*Jiang et al.*, 2015a). Oxidation of methane is the dominant hydrocarbon source of CO. Its chemical oxidation starts with reaction with OH as follows:

$$CH_4 + OH \rightarrow CH_3 + H_2O. \tag{1.8}$$

The methyl radical CH_3 then immediately reacts with O_2,

$$CH_3 + O_2 + M \rightarrow CH_3O_2 + M. \tag{1.9}$$

The methylperoxy radical (CH_3O_2) has similar chemical features as HO_2 and reacts with NO as follows:

$$CH_3O_2 + NO \rightarrow CH_3O + NO_2. \tag{1.10}$$

The methoxyl radical (CH_3O) then would react with O_2,

$$CH_3O + O_2 \rightarrow CH_2O + HO_2. \tag{1.11}$$

Combining Reactions 1.8-1.11, formaldehyde (CH_2O) is produced. In fact, NMVOCs follow similar chemical reaction pathways as methane in producing CO and CH_2O, but with large uncertainties in the chemical schemes implemented in current CTMs (*Shindell et al.*, 2006). The oxidation of NMVOCs and CH_4 accounts for 30% and 15% of the OH sink, respectively. The remaining 15% of the sink is associated with the reaction of OH with O_3, HO_2 and hydrogen gas (H_2).

Ozone, as a precursor of tropospheric OH, is produced in the troposphere by the

Table 1.1: Global budget of atmospheric CO (Tg CO/year). Sources are based on CO emission estimates from the GEOS-Chem model (*Duncan et al.*, 2007) and sinks are based on the estimates from *Jacob* (1999).

Type	Flux estimates
Sources: 2256-2689	
Direct Emissions	
Anthropogenic	391-411
Biomass burning	406-516
Biofuels	159
Photochemical oxidation of VOCs	
Anthropogenic	72-76
Biomass burning	45-57
Biofuels	30
Biogenic NMVOCs	
Isoprene	170-184
Monoterpenes	68-71
Methanol	95-103
Acetone	21
Oxidation of CH_4	778-861
Oceans	20-200
Sinks: 2100-3000	
OH oxidation	1400-2600
Soil uptake	250-640
Stratosphere transport	100

oxidation of CO and hydrocarbons in the presence of NO_x (*Jacob*, 1999). When HO_2, formed in Reactions 1.7 and 1.11, reacts with NO in Reaction 1.4 to produce NO_2, this NO_2 rapidly leads to O_3 production as follows:

$$NO_2 + h\nu \rightarrow NO + O, \tag{1.12}$$

and

$$O + O_2 + M \rightarrow O_3 + M. \tag{1.13}$$

where M is a third body. The $O(^1D)$ produced in Reaction 1.1 can be quenched through

$$O(^1D) + M \rightarrow O + M, \tag{1.14}$$

producing atomic oxygen in its ground state, which can rapidly return to O_3 through Reaction 1.13.

It should be noted that the abundance of NO_x and hydrocarbons can limit O_3 produc-

tion. When NO_x concentrations are low, O_3 production is limited by the abundance of NO_x so that O_3 production increases linearly with NO concentration regardless of abundance of hydrocarbons. This is called the NO_x-limited regime. The NO_x-limited regime is typical in rural areas as well as urban areas without severe NO_x pollution. When NO_x concentrations are very high, O_3 production increases linearly with hydrocarbon concentrations but varies inversely with NO_x concentrations. That is to say, for regions with severe NO_x pollution, local O_3 production could be low due to low abundances of hydrocarbons. This refers to the hydrocarbon-limited regime.

Table 1.2: Global source estimates for tropospheric NO_x (Tg N/year) (*Jacob*, 1999)

Type	Flux estimates
Anthropogenic	21
Biomass burning	12
Lightning	3
Aircraft	0.5
Soil	6
NH_3 oxidation	3
Stratosphere	0.1

As shown in Table 1.2, the dominant source of atmospheric NO_x is from combustion (including anthropogenic and biomass burning), which produce NO. The cycling between NO and NO_2 is fast since NO emissions are rapidly converted to NO_2 through Reaction 1.4, and the NO_2 produced by this reaction will photolyze and generate NO and O_3 via Reaction 1.12 and 1.13 (*Jacob*, 1999). NO_x is lost by either conversion into its reservoir species temporarily or being removed permanently from the atmosphere. Permanent removal of tropospheric NO_x involves formation of nitric acid (HNO_3) during daytime,

$$OH + NO_2 + M \rightarrow HNO_3 + M, \tag{1.15}$$

followed by removal of HNO_3 from the atmosphere. Due to increasing NO_x in the atmosphere, HNO_3 became one of the most harmful acidic air pollutants with concentrations of up to 1 ppbv in polluted regions (*Goldan et al.*, 1983). HNO_3 is removed within a few days by dry and wet deposition. In the upper troposphere, where the photolysis of HNO_3 occurs, its abundance is strongly affected by NO_2 and O_3.

Since the surface emissions of NO_x are mostly due to anthropogenic and biomass burning, regions with high emissions of NO_x coincide with regions exhibiting high CO emissions (*Stavrakou and Müller*, 2006) (see Fig. 1.4), suggesting strong spatial correlation between CO and NO_x emissions. Other than combustion sources, soil emissions also contribute 13% of the total NO_x emission budget (Table 1.2). Emissions from lightning

(LNO_x) refer to nitric oxide production during thunderstorms as a result of a lightning. They are most abundant within the tropical and subtropical regions, such as over the Amazon, equatorial Africa and Indonesia, where deep convection may occur frequently. The NO_x concentration in some NO_x emission hotspots (e.g. East Asia) can be as high as 50 ppbv (*Lamsal et al.*, 2008). Since the lifetime of surface NO_x only lasts for hours, tropospheric NO_x is not well mixed globally. The distribution of mid- and lower-tropospheric NO_x suggests high NO_x concentrations are strongly confined to its source regions. Moreover, tropospheric NO_x is unable to be transported from surface to free troposphere effectively. In the upper troposphere, NO_x is mostly contributed by lightning sources distributed in the intertropical convergence zone.

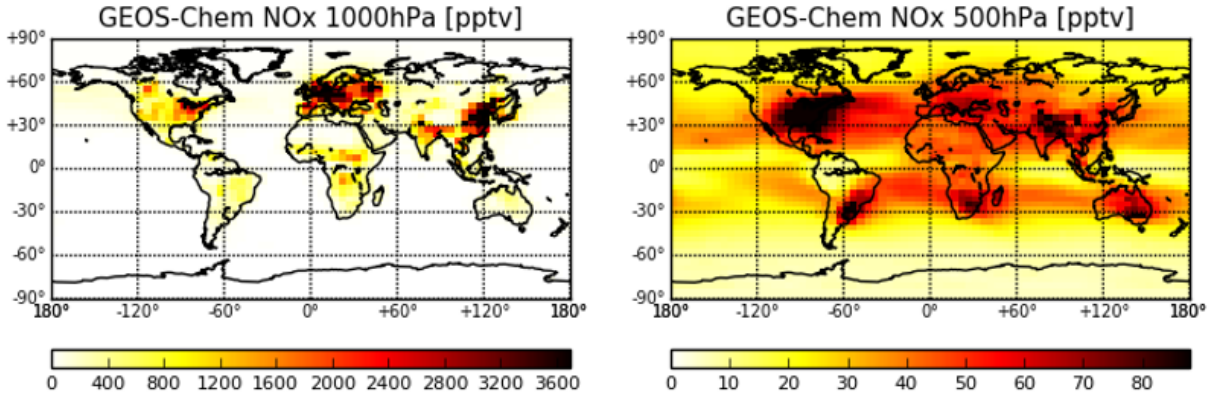

Figure 1.4: Mean NO_x concentrations (in pptv) at the surface and at 500 hPa in 2016 from a simulation of the GEOS-Chem model (version 9).

1.2 Uncertainties on modelled OH

Tropospheric OH is extremely difficult to estimate due to its short chemical lifetime. Steady state equilibrium is considered when computing the global tropospheric OH mean concentration, $[\overline{\mathrm{OH}}]$ (*Jacob*, 1999). The mathematical expression for $[\overline{\mathrm{OH}}]$ is given by

$$[\overline{\mathrm{OH}}] = \frac{\int_{\mathrm{trop}} n_a[\mathrm{OH}]dV}{\int_{\mathrm{trop}} n_a dV}, \tag{1.16}$$

where n_a is the number density of air, serving as the weighting factor in computation of the mean, [OH] is the local OH concentration, and the volume integral is calculated over the troposphere. The interannual trend in $[\overline{\mathrm{OH}}]$ over the past two decades is largely unknown compared to other trace gases (e.g. O_3) because both OH sources (e.g. O_3) and sinks (e.g. CO) are changing (*Montzka et al.*, 2011). Recent studies such as *Naik*

et al. (2013) compared the OH benchmark from multiple chemistry and climate models (CCMs), and suggested that there was a slight positive trend since 2000 in tropospheric OH despite the large uncertainty in OH. In addition to the global mean OH, the inter-hemispheric ratio (NH/SH) of OH serves as a key metric to investigate whether these OH changes over the last two decades were caused by anthropogenic HO_x precursors (which are concentrated in the NH), hemisphere-specific processes, or global processes (such as well-mixed, long-lived species).

Global OH abundances can also be characterized by the lifetime of OH-dependent long-lived gases (*Prinn et al.*, 1987). For instance, the main sink for methylchloroform (CH_3CCl_3, informally referred to as MCF) is through reaction with OH in the troposphere. Since CH_3CCl_3 sources are only from anthropogenic activities, its lifetime inferred from surface measurement of CH_3CCl_3 (obtained from the late 1970s to the early 2000s) can be related to OH concentrations through the following equations (*Jacob*, 1999):

$$[\overline{\mathrm{OH}}] \approx \frac{P - L}{[\mathrm{CH_3CCl_3}] \int_{\mathrm{trop}} n_a dV}, \tag{1.17}$$

$$\tau_{\mathrm{CH_3CCl_3}} = \frac{1}{k[\mathrm{OH}]}, \tag{1.18}$$

where P and L are the production and loss rate of CH_3CCl_3 respectively, $[CH_3CCl_3]$ is the mixing ratio of MCF, $k(T)$ is the rate constant (temperature dependent) for MCF oxidation by OH. L and $[CH_3CCl_3]$ near the surface are obtained from in-situ observations. It is noted that the abundance of MCF has been decreasing since the 1990s as a consequence of prohibition on MCF production under the legislation of the Montreal protocol (*Jacob*, 1999). As a result, P is assumed to be negligible with $[CH_3CCl_3]$ assumed to be uniform within the troposphere. In this way, $[\overline{\mathrm{OH}}]$ can be computed with the measurement of MCF. In turn, the lifetime of MCF can be presented given temperature and OH concentration profiles. Similar to MCF, the lifetime of CH_4 can also be inferred given temperature and OH information. According to *Zhang et al.* (2018), the lifetime of well-mixed tropospheric methane due to OH oxidation can be expressed as

$$\tau_{\mathrm{CH_4}} = \frac{\int_{\mathrm{trop}} n_a[\mathrm{OH}]dV}{\int_{\mathrm{trop}} k(T) n_a dV}, \tag{1.19}$$

where n_a, $k(T)$, V, [OH] are defined the same as Eq. 1.16 and Eq. 1.17. The only difference between MCF and CH_4 is the reaction rate constant with respect to OH oxidation, which is based on various laboratory studies (e.g., *Sander et al.*, 2003; *Burkholder et al.*, 2015).

Before the 2000s, estimates of the global mean tropospheric OH inferred from MCF

observations exhibited an uncertainty of 10% (*Nicely et al.*, 2017). But since the 2010s, there have been large uncertainties on MCF-based OH constraints because the MCF abundance in the troposphere is now too low (*Montzka et al.*, 2011). Recently, combinations of hydrofluorocarbons (HFCs) have been used as an alternative to MCF to constrain OH. However, the uncertainty of the inferred OH abundance was still large. *Zhang et al.* (2018) proposed joint optimization of surface CH_4 fluxes and global OH concentrations using shortwave infrared (SWIR) and thermal infrared (TIR) CH_4 observations as an alternative method for characterizing global OH concentrations. The inferred NH/SH OH ratio was 0.92, with an accuracy of less than 10%.

Alternatively, tropospheric OH can be estimated empirically by combining observations and models. *Spivakovsky et al.* (2000) estimated tropospheric OH based on the observed distribution of a suite of OH-related chemical species (O_3, NO_x, NO_y, CO, hydrocarbons) as well as water vapour, temperature and cloud optical depth. The estimated annual global mean OH was 11.6×10^5 molec/cm^3 with the hemispheric ratio close to unity. *Spivakovsky et al.* (2000) conducted a sensitivity analysis and concluded that in the absence of large changes in the tropospheric chemical mechanism, "it is difficult to contemplate errors in excess of $\pm$15% in global and hemispheric climatological averages for the abundance of OH, given our present knowledge of the distributions of precursors and the limited sensitivity of computed concentrations of OH to changes in their specification." Since the estimates of both the global mean OH and the NH/SH ratio from *Spivakovsky et al.* (2000) were well-matched with observation-based constraints at the time, the 3D monthly mean climatology from *Spivakovsky et al.* (2000) is still considered a good standard when evaluating modelled OH concentrations (*Müller et al.*, 2018).

Since the early 2000s there have been updated OH chemistry schemes implemented in various CTMs and CCMs. *Yantosca* (2018) has tracked the evolution of GEOS-Chem model one-year benchmark simulations of tropospheric OH abundances shown in Fig. 1.5. As more chemical tracers and more reactions were implemented in the model, the global mean OH benchmark varied significantly. Also, *Naik et al.* (2013) analyzed 17 global models in the Atmospheric Chemistry and Climate Model Intercomparison Project (ACCMIP) and found a slight increase of 3.5$\pm$2.2% in the multi-model mean for present day (2000) OH compared to OH in the 1980s. The multi-model mean, global mean OH was $11.1\pm1.6\times10^5$molec/cm^3 with a NH/SH ratio ranging from 1.13 to 1.42.

Although the differences in $[\overline{OH}]$ among different CTMs and CCMs were smaller than 15% (*Nicely et al.*, 2017), such disagreements among model would lead to large discrepancies on simulating key atmospheric substances such as methane, CO and O_3.

For instance, *Voulgarakis et al.* (2013) calculated the CH_4 lifetime based on the OH profile estimated from ACCMIP. The resulting multi-model mean of global mean CH_4 lifetime was 9.8±1.6 years, with even larger uncertainties associated with regional OH.

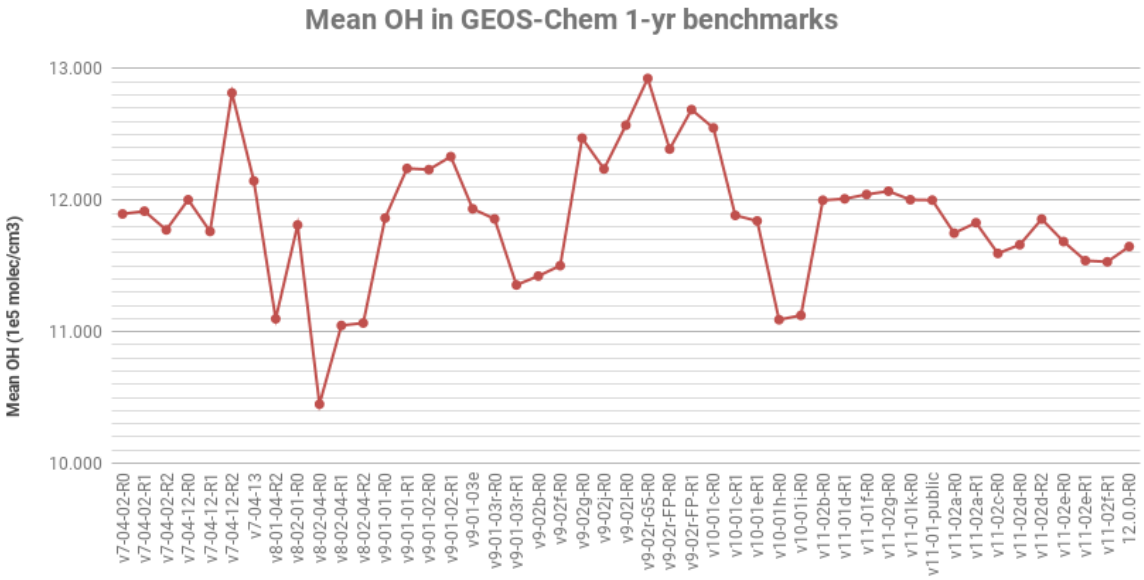

Figure 1.5: GEOS-Chem simulated mass weighted mean OH concentrations (in 10^5 molec/cm^3) from 1-year benchmark simulations among various model versions (*Yantosca*, 2018). GEOS-Chem v08-02-01 were more consistent with the MCF-based OH constraints. The mean OH concentrations from simulations retrieved from v08 to v09 of the model were also within the estimates (with slight overestimation) compared to *Spivakovsky et al.* (2000).

The large regional OH uncertainties are caused by not only different CTM setups (transport schemes and parameterizations), but also by discrepancies in the chemistry scheme on regional scales. *Strode et al.* (2015) found that the low biases of CO in the northern extratropics seen in Fig. 1.1 were related to high biases on the NH/SH ratio of OH. For the NH, *Patra et al.* (2014) further explained that the OH could be overestimated due to the overestimation of HO_x-related species such as NO_x. For the tropics and SH, *Rohrer et al.* (2014) argued that regions with low-NO_x, high-NMVOCs (e.g., South America and Indonesia) had large OH uncertainties due to unrecognized OH recycling mechanisms. Some missing mechanisms in the models are related to the peroxyl radical reactions such as Reaction 1.5 (*Rohrer et al.*, 2014). In fact, these unrecognized OH recycling mechanisms would result in up to ten times more OH being produced than current CTMs estimate in certain regions. They are dependent on the photochemical production rate as well as the regional NMVOC and NO_x abundances, which are difficult to simu-

late using laboratory approaches (*Rohrer et al.*, 2014). To improve the OH concentration estimation, tropospheric chemistry mechanisms as well as OH measurement techniques require further development. Since neither MCF-based OH constraints (e.g., *Montzka et al.*, 2000; *Krol et al.*, 1998; *Bousquet et al.*, 2005) nor CTMs (e.g., *Naik et al.*, 2013) suggest consistent information on global mean OH abundances or the NH/SH ratio, more studies on providing tropospheric OH constraints such as assimilating CO, NO_x and O_3 observations are required.

1.3 Estimating CO emissions

Since CO plays a key role in determining the oxidative capacity of the atmosphere, quantifying the CO budget within CTMs has become a new point of interest in recent atmospheric carbon cycle studies. To simulate CO as a trace gas, most chemical transport models rely on "bottom up" emission estimates that are based on regional and international energy consumption statistics and CO emission factors for different energy sectors. For instance, anthropogenic emission inventories such as the Emissions Database for Global Atmospheric Research (EDGAR) (*Olivier and Berdowski*, 2001) use statistics of international activity data, emission factors per activity and scaling factors accounting for local meteorological and time variables to estimate gaseous and particulate air pollutant emissions that are related to human activities. Biomass burning emission inventories such as the Global Fire Emissions Database, version 3 (GFED-3) (*Giglio et al.*, 2013) use burned area, fuel loads, and combustion efficiency, together with scaling factors to account for emissions that are related to wildfires. However, these emission estimates may exhibit large uncertainties. Some emission types are also highly variable in time and cannot be accurately represented using the heavily-averaged emission inventories. In comparison, "top down" emission estimates relying on inverse modelling techniques would produce emissions with improved spatial and temporal coverage. There are different types of observations with different precision and spatiotemporal resolution, as well as with different sensitivity to the vertical distribution of CO in the troposphere. In general, surface and aircraft observations will have higher observation precision. However, the surface observing network and aircraft campaigns have limited spatiotemporal coverage. They are mostly distributed in the regions with high social-economic activity, which leaves large gaps in observational coverage in the tropics and polar regions (*Rodgers*, 2000). Satellite measurements, in contrast, provide observations with higher spatial and temporal coverage. For example, the current CO satellite observations from Measurements Of Pollution In The Troposphere (MOPITT) and Infrared Atmospheric Sounding

Interferometer (IASI) have measured atmospheric CO abundances globally over 20 years, providing global CO coverage in 3 days for MOPITT, and twice daily for IASI.

CO inverse modelling studies have applied surface (e.g., *Pétron et al.*, 2002; *Pison et al.*, 2009; *Hooghiemstra et al.*, 2012), aircraft (e.g., *Palmer et al.*, 2003a; *Heald et al.*, 2004) and satellite observations (e.g., *Arellano et al.*, 2004; *Stavrakou and Müller*, 2006; *Jones et al.*, 2009; *Kopacz et al.*, 2010; *Jiang et al.*, 2013; *Fortems-Cheiney et al.*, 2011) to optimize global and regional CO sources. There are also long-term CO inversions available (e.g., *Miyazaki et al.*, 2015; *Jiang et al.*, 2017) to study CO emission trends in the past two decades. Many of these studies have suggested the existence of significant biases in bottom up inventories. For instance, *Kopacz et al.* (2010) used GEOS-Chem with CO observations from MOPITT, the Tropospheric Emission Spectrometer (TES), and the Atmospheric InfraRed Sounder (AIRS). For anthropogenic sources, a significant underestimation (up to 60%) was found in Europe, North America, and Asia. Such underestimation was also confirmed in *Jiang et al.* (2013) and *Fortems-Cheiney et al.* (2011). Moreover, most of the bottom up emission inventories did not consider strong seasonal variations in northern mid-latitudes compared to the a posteriori (*Kopacz et al.*, 2010). For biomass burning sources, *Jones et al.* (2009) used GEOS-Chem with CO observations from TES and MOPITT to constrain tropical CO sources. They showed an underestimation in equatorial Africa and Indonesia (up to 90%) in November 2004. The biomass burning emission underestimation was also confirmed in *Kopacz et al.* (2010) throughout 2004. To validate the CO emission estimates from inversions, *Hooghiemstra et al.* (2012) compared the global and regional top down CO emission estimates among many of the aforementioned studies. Despite the fact that the global CO emission estimates were within 10%, the regional CO emission estimates for North America, Europe and East Asia differed by 30-50%. This showed that the chemical sources and sinks of atmospheric CO on regional scales were still poorly characterized.

Among constituents that are chemically related to tropospheric CO, NMVOCs and OH both have large uncertainties that affect CO inversions. The chemical production of CO by NMVOC oxidation accounts for 10-30% of the total atmospheric CO source and is responsible for 50% of background CO (*Duncan et al.*, 2007). *Miyazaki et al.* (2012a) argued that modelled CO biases were partially due to poor characterization of hydrocarbons. As the oxidation of NMVOCs generates formaldehyde (HCHO), there were recent inverse modelling studies using satellite HCHO observations (e.g., *Millet et al.*, 2006; *Hudman et al.*, 2008a; *Marais et al.*, 2012; *Barkley et al.*, 2013) to characterize biogenic isoprene emissions. For instance, *Marais et al.* (2012) used GEOS-Chem with OMI HCHO observations in western equatorial Africa. Overestimation of isoprene emis-

sions by 24-48% was found in the Model of Emissions of Gases and Aerosols from Nature (MEGAN) isoprene inventory used in GEOS-Chem. Moreover, 40-90% uncertainties of biogenic CO emissions were found in the chemistry scheme of NMVOCs in low NO_x conditions (*Marais et al.*, 2012). In fact, these HCHO inversions would help reduce the uncertainty of NMVOCs, which would help to better constrain CO emission estimates.

In addition to biases in the chemistry, inverse modelling of CO emissions can also be affected by model transport biases (e.g. in convection) and observation biases. If the transport scheme used in a CTM exhibits large systematic biases, then the differences between the modelled CO fields and observations are not totally due to biases in the emission estimates. Sometimes model biases could lead to unrealistic emission adjustment when assimilating column datasets (*Qu et al.*, 2017). *Jiang et al.* (2015a) used GEOS-Chem with Near Infrared (NIR) and TIR/NIR profiles of MOPITT version 5 retrievals. Their CO emissions inferred from TIR/NIR profiles were larger than those inferred from NIR profiles in India and Southeast Asia, suggesting that the strength of convection in the GEOS-Chem model is too weak. In contrast, the profile inversion provided lower CO emissions for North America and Europe, indicating that the boundary layer mixing in these two regions tends to be too sluggish (*Jiang et al.*, 2015a). *Yan et al.* (2014) used the GEOS-Chem global model (at $2^\circ \times 2.5^\circ$) with three nested domains (at $0.5^\circ \times 0.667^\circ$) covering Asia, North America, and Europe to interpret data from the High-Performance Instrumented Airborne Platform for Environmental Research (HIAPER) Pole-to-Pole Observations (HIPPO) campaign. They found less bias in the nested-grid models than in the global model, suggesting that model performance could also be influenced by model resolution. Improved CO emission estimates as a result of increased model resolution was also confirmed in the North American regional inversion study by *Jiang et al.* (2015b).

1.4 Estimating NO_x emissions

Estimating the modelled NO_x budget is helpful for quantifying the modelled O_3 burden, which then indirectly determines the oxidative capacity of the modelled atmosphere. Inverse modelling studies have used NO_x in-situ aircraft measurements (e.g., *Heland et al.*, 2002) and satellite observations (e.g., *Martin et al.*, 2003; *van der A et al.*, 2008; *Boersma et al.*, 2008; *Müller and Stavrakou*, 2005; *Miyazaki et al.*, 2012b) to estimate NO_x emissions. These emission estimates are important for both air quality and atmospheric chemistry studies. Satellite observation of tropospheric NO_2 began with Global Ozone Monitoring Experiment (GOME), then continued with the SCanning Imaging Absorption spectroMeter for Atmospheric CHartographY (SCIAMACHY), the Ozone

Monitoring Instrument (OMI), GOME-2 and TROPOMI. To examine bottom-up inventories, *Martin et al.* (2003) used GEOS-Chem with GOME NO_2 column abundances. They found that anthropogenic NO_x emissions were underestimated in winter over most regions in the Northern Hemisphere. Such findings were also confirmed by *Miyazaki et al.* (2012a). Biomass burning emissions were overestimated in India, central Africa and Brazil (*Miyazaki et al.*, 2012a). Moreover, *Jaeglé et al.* (2005) argued that NO_x biomass burning emission inventories did not capture the seasonal cycle effectively. LNO_x emissions were estimated by *Martin et al.* (2007) by assimilating both tropospheric NO_2 column from SCIAMACHY, O_3 columns from OMI and Microwave Limb Sounder (MLS), and upper tropospheric HNO_3 from Atmospheric Chemistry Experiment-Fourier Transform Spectrometer (ACE-FTS). Large enhancement (2-6$\times 10^{14}$ molec/cm^2) of NO_2 in the UTLS was found over the tropical Atlantic and Africa compared to the a priori profile. In contrast, LNO_x was 25-50% smaller over the tropical Pacific than the a priori (*Martin et al.*, 2007). They estimated the global LNO_x emissions to be 6 ± 2 Tg N/year, which dominated the total NO_x emissions in the free troposphere during summertime.

Top down NO_x emissions are affected by several factors. First, model resolution determines the representativeness of surface NO_x emissions. Increasing the model resolution could reduce the uncertainty of NO_x emission estimates. *Qu et al.* (2017) used the nested-grid GEOS-Chem model with NO_2 columns from OMI. Many localized emission changes were detected in the regional CTM, which could not be captured in global studies. Second, the emission estimates can be affected by the observation error. The observation error, depending on the retrieval algorithm, is large. The majority of the measurement biases are associated with the stratosphere-troposphere separation (STS) algorithm. Full details of the STS are further explained in Section 2.1.1. *Zheng et al.* (2014) compared the standard OMI NO_2 product (SP) and the Dutch OMI NO_2 product (informally as DOMINO or DP) and found that OMI-DP was 9% lower in summer but 13% higher in winter compared to OMI-SP. These differences were mostly due to the difference in relative stratosphere and tropospheric contributions and in the air mass factor calculation. The two products also exhibited large discrepancies for observations over heavily polluted regions. Furthermore, the traditional STS techniques used in OMI-DP (*Boersma et al.*, 2008) and OMI-SP (*Bucsela et al.*, 2006) assumed stratospheric NO_2 contribution interpolated from NO_2 abundance from remote and unpolluted regions. Recently, *Adams et al.* (2016) retrieved OMI-minus-OSIRIS (OmO) tropospheric NO_2 columns based on a limb-nadir matching technique, relying on the subtraction of the total nadir observed NO_2 column using non-coincident limb measurements. Details of the limb-nadir matching techniques are further discussed in Section 2.1.1. The OmO tropospheric vertical

column densities (VCDs) have shown consistency in background NO_2 with both OMI-SP and OMI-DP, with higher variability of stratospheric NO_2 contributions inferred from OSIRIS NO_2 observations. *Gu et al.* (2014) argued that the difference among the OMI-DP and OMI-SP retrieval products can lead to differences in the NO_x emission estimates.

1.5 Uncertainties on modelled O_3

According to the CO-HO_x-NO_x-O_3 chemistry discussed in Section 1.1, CO and NO_x concentrations will determine the abundance of tropospheric O_3. In turn, model biases in tropospheric O_3 could impact CO and NO_x inversion estimates. *Müller and Stavrakou* (2005) found that inverse modelling using both CO and NO_x observations could provide better CO estimates than assimilating CO observations alone. *Jones et al.* (2009) explained that such CO estimation improvements could be caused by changing tropospheric O_3 and OH. *Wild* (2007) examined the impact of precursor emissions, lightning, STE, wet and dry deposition, humidity, and model resolution on modelled O_3 concentrations. They found that increases in CO and NO_x emissions would lead to enhancement of lower tropospheric O_3. O_3 concentrations in the UTLS were sensitive to LNO_x emissions, which are poorly constrained in current CTMs. Some variability of O_3 abundance can be explained by the dry deposition and convection scheme used in the model. The relative importance among these factors depends on the analyzed regions. *Wild* (2007) also showed that coarse model resolution leads to overestimation of STE, but underestimation of dry deposition and convection.

To correct modelled O_3 biases, previous inverse modelling studies have relied on combining CTMs with satellite observations from GOME (e.g., *Tellmann et al.*, 2004; *Liu et al.*, 2007), TES (e.g., *Bowman et al.*, 2006; *Jourdain et al.*, 2007; *Nassar et al.*, 2009; *Parrington et al.*, 2008), IASI (e.g., *Emili et al.*, 2014; *Peiro et al.*, 2018) and MLS (e.g., *Barré et al.*, 2012; *Miyazaki et al.*, 2012a). For instance, *Parrington et al.* (2008) used GEOS-Chem with CO and O_3 observations from TES. They found that the mean difference between the model and the observed ozone reduced from about 30% to less than 5%. In order to characterize the uncertainty associated with STE, *Barré et al.* (2012) used Modéle de Chimie Atmosphérique de Grande Echelle (MOCAGE, a CTM developed by Meteo-France) to assimilate stratospheric O_3 observations from MLS. They improved the O_3 profile in the UTLS, with the posterior O_3 agreeing with independent ozonesonde data to within 5%. *Emili et al.* (2014) and *Peiro et al.* (2018) assimilated IASI and MLS observations, and found that the posterior O_3 state in the free troposphere and lower stratosphere followed the variability from IASI O_3 observations, despite their O_3 fields

being driven by a simplified chemistry model. So far, these studies have shown that the modelled chemical state of O_3 and NO_2 can be effectively improved via O_3 assimilation. However, some surface O_3 biases still remain partially due to lack of observation sensitivity in the boundary layer for the remote sensing instruments.

1.6 Multi-species chemical data assimilation

Considering the uncertainty on estimating modelled OH, CO, NO_x, and O_3, current inverse modelling studies have applied advanced data assimilation frameworks to provide constraints on the CO-HO_x-NO_x-O_3 chemistry. With the large suite of space-based measurements of tropospheric composition, MSA offers a powerful means for simultaneously optimizing different chemical species. For example, assimilation of O_3, CO, NO_2, and HCHO could provide constraints on both the sources and sinks of OH.

In the past, various inverse modelling studies (e.g., *Müller and Stavrakou*, 2005; *Chai et al.*, 2006; *Zhang et al.*, 2008a; *Miyazaki et al.*, 2012a, 2015, 2017) have utilized adjoint, Ensemble Kalman Filter (EnKF), and four-dimensional variational (4D-Var) data assimilation approaches to estimate CO and NO_x emissions, and have shown improvements on modelled CO, O_3 and NO_2 concentrations when comparing with independent measurements. These studies have used advanced data assimilation techniques to optimize multiple types of emissions and concentrations simultaneously. For example, *Müller and Stavrakou* (2005) used the Intermediate Model of Global Evolution of Atmospheric Species (IMAGES, a CTM) with an adjoint modelling technique optimizing CO and surface NO_x emissions. Observations of both CO observations provided by the National Oceanic and Atmospheric Administration/ Climate Monitoring and Diagnostics Laboratory (NOAA/CMDL) ground-based measurements and tropospheric NO_2 columns from OMI were used. The uncertainties of CO and NO_x emission estimates in the northern extratropics are significantly reduced when assimilating both observations. Moreover, they proposed that simultaneous inversions of the chemically active compounds would be promising for further inverse modelling studies.

Miyazaki et al. (2012a) used the CHemical Atmospheric general circulation model for Study of atmospheric Environment and Radiative forcing–Data Assimilation System (CHASER-DAS, a CTM) with a Local ensemble transform Kalman Filter (LETKF) data assimilation scheme to simultaneously optimize 35 tropospheric tracers as well as CO, NO_x and LNO_x emissions. The experiments used observations of CO from MOPITT, O_3 from TES and MLS, NO_2 from OMI, and HNO_3 from MLS (*Miyazaki et al.*, 2012a). In their experiments, optimizing emissions were effective to constrain tracers' concentrations

near the surface. Optimizing concentrations were effective to constrain the concentrations in free troposphere. The a posteriori provided the most significant corrections over the tropics for O_3, OH, and LNO_x, which showed large uncertainties in the bottom-up inventories and background concentrations in this region. Meanwhile, significant increase was obtained along northern midlatitudes for CO and NO_x emissions. Reductions of both residual bias (by 85%) and Root-Mean-Square Error (RMSE) (by 50%) were obtained after the inversion, compared with independent station, aircraft and ozonesonde data. This indicates that applying MSA is an effective method to reduce biases in assimilated tracers and improve CO and NO_x emission estimations (*Miyazaki et al.*, 2012a).

More recently, *Miyazaki et al.* (2015) conducted a long-term LETKF-based chemical reanalysis to reduce biases on modelled NO_2 and O_3, and to improved the seasonal and interannual variability of CO and NO_x emissions. In their tropospheric chemical reanalysis, in addition to improving the modelled O_3, NO_2 and CO, *Miyazaki et al.* (2015) also calculated the global mean OH and the NH/SH ratio of tropospheric OH, both of which approached the MCF-based ratio estimates and fell in the range of the ACCMIP multi-model mean estimates. This shows MSA as a decent solution to improve modelled OH concentrations. The 4D-Var based MSA has been studied by *Inness et al.* (2015). They obatined reductions in modelled CO and O_3 biases through chemical feedbacks of assimilating all species. Moreover, the chemical states assimilated by the 4D-Var scheme were able to produce smoothed 3D fields that were consistent with all simulated chemical species. The assimilation of multiple O_3 datasets significantly improved the modelled O_3 fields relative to the a priori, especially in the free troposphere and lower stratosphere. They, however, found that assimilating the initial conditions of NO_2 was not impactful due to the short lifetime of NO_2 compared to the 12-hour assimilation window. There were also regional CTM inversion studies (e.g., *Chai et al.*, 2006; *Zhang et al.*, 2008b) focused on O_3, NO_x and NO_y by assimilating multiple chemical species using the 4D-Var data assimilation scheme. *Chai et al.* (2006) found that the adjusted initial chemical tracers improved NO_y after simultaneously assimilating aircraft observations of O_3, NO, NO_2, HNO_3, PAN and lumped organic nitrates (RNO_3). In *Zhang et al.* (2008b), they also improved the chemical field of NO_2 by adjusting the initial conditions of chemical tracers after assimilating NO_2 data from SCIAMACHY and O_3 data from AirNow ground level observations.

1.7 The objectives and structure of the thesis

1.7.1 The objectives

The focus of this thesis is on implementing a 4D-Var based multi-species chemical data assimilation framework to mitigate modelled chemical biases in the inverse modelling of CO and NO_x emission. In particular, I would like to improve tropospheric OH and its precursor, O_3, through data assimilation of multiple OH- and O_3-related species, such as CO, O_3, NO_2, HCHO, and HNO_3. Several objectives were proposed at the beginning of the research:

1. In the first phase, I was interested in designing an MSA framework for the GEOS-Chem adjoint model. I used the GEOS-Chem adjoint model with the 4D-Var data assimilation scheme to optimize CO, NO_x emissions and O_3 initial conditions. Satellite observations of CO from MOPITT, O_3 from both TES and OSIRIS, NO_2 from OMI in November 2009 and July 2010 were assimilated. The two months were chosen as a case study for seasonal contrasts. By combining all the observations with the CTM, my objective was to improve tropospheric O_3, and thus OH, in order to provide better global estimates of CO and NO_x emissions.

2. In the second phase, I improved the MSA framework and then conducted a one-year chemical reanalysis to better constrain tropospheric OH based on the extended MSA approach. Extension of the MSA will focus on better constraining the CO sources due to the oxidation of NMVOCs as well as the LNO_x source. Hence, I assimilated CO profiles from MOPITT, HCHO and NO_2 tropospheric columns from OMI, O_3 partial columns from IASI and OSIRIS, and HNO_3 profiles from MLS. The objective was to estimate annual CO emissions as well as indirectly constrain tropospheric OH.

3. In the third (last) phase, I evaluated the utility of new satellite retrievals of tropospheric NO_2 for inverse modelling NO_x emissions. I performed a regional NO_x inversion with O_3 and NO_2 data. It is noted that different from the joint assimilations operated in phase one and two, observations of O_3 and NO_2 were assimilated separately. The top-down summertime NO_x emission estimates for North America for 2008-2011 were obtained by assimilating the OmO data using the GEOS-Chem nested grid model with the O_3 assimilated boundary conditions.

The first and second thesis objectives have been presented in two papers. The study in MSA phase one has been published in *Journal of Geophysical Research: Atmosphere* in

January, 2019. The study in MSA phase two will be submitted to *Journal of Geophysical Research: Atmosphere*. The title of the two papers are shown below:

1. Quantifying emissions of CO and NO_x using observations from MOPITT, OMI, TES,and OSIRIS.

2. Constraints on the oxidative capacity of the atmosphere: Implications for top-down CO emission estimates.

1.7.2 The significance of this thesis

The two MSA studies are the first attempt using the GEOS-Chem model with 4D-Var assimilation to conduct a chemical reanalysis to estimate CO, NO_x, O_3 and OH concentrations. Cost function scaling algorithm among different species is tested to account for correlation among different chemical species in the MSA framework. Traditional single instrument inversions can be compared with MSA to show the impact of assimilating each observation to the posterior modelled states. Chemical coupling and adjoint model sensitivity can also be examined. The posterior OH concentrations, CO emissions, and O_3 burden would reveal modelled biases on CO-OH-NO_x-O_3 chemistry. The assimilated global mean OH can be used as an improved OH profile for other inverse modelling studies on OH-related species such as CH_4. The regional NO_x analysis is the first study using OmO NO_2 observations in the context of NO_x inverse modelling. This regional study is also innovative since the boundary condition for the nested-grid model uses the assimilated concentrations, rather than forward model simulations only. Our posterior NO_x emission estimates can be compared with the top down estimates using OMI observations to examine whether stratospheric NO_2 observations would impact surface emission estimates.

1.7.3 The structure of this thesis

The outline for remaining chapters in this thesis is shown as follows:

Chapter 2 introduces the datasets used for assimilation and validation. The instrumentation, retrieval and bias correction information are explained. I also depict why I use each observation for certain purposes within my study. After that, the GEOS-Chem model and the inversion methodologies are presented. I will go over the most fundamental properties of GEOS-Chem, including its basic driven meteorological fields, chemical schemes as well as the employed bottom up inventories. For inversion methodologies, optimal estimation for retrievals, strong-constraint (traditional) 4D-Var, weak-constraint 4D-Var, and LETKF will be described to illustrate the basic theory of all the inversions.

Chapter 3 investigates the results of the first phase of the MSA. Results from single instrument assimilation (referred to as the single instrument run) will be compared with those from MSA (referred to as the all instrument run). Validation on modelled CO, NO_2 and O_3 fields against independent observations are explained. The top down estimates using MSA are then investigated, followed by comparison with other CO and NO_x inverse modelling studies. In the end, I will include a discussion on the impact of observation error and the assimilation window on the assimilated results before a brief summary.

Chapter 4 looks at the results of the second phase of the MSA. I will also compare the resulting OH concentration and O_3 burden with other OH and O_3 benchmark studies (e.g., *Spivakovsky et al.*, 2000; *Gaudel et al.*, 2018). Validation of modelled CO, O_3 and OH fields against independent observations are explained. The top down CO emission estimates for a full year chemical reanalysis will be compared with other CO inversions. In the end, I will include a discussion on the difference among all the experiments employed in the study as well as the OH sensitivity to different assimilated species.

Chapter 5 analyzes results on the regional NO_x inversion studies. I will highlight the employed experiments, evaluate the optimized tracers against various independent observations, and discuss the top down North American NO_x emission estimates. I will also include an uncertainty analysis and a summary at the end.

Chapter 6 summarizes the main findings of this thesis and proposes some future pathways for further MSA studies.

1.8 Contributions

The analyses presented in the thesis were all performed by the author, but also assisted by the contributions of many co-authors. All the authors have communicated and provided feedback for all the projects. Their contributions for each project are as follows:

Chapter 3 was a collaboration with Dylan Jones, Martin Keller, Thomas Walker, Zhe Jiang, Daven Henze, Helen Worden, Adam Bourassa, Doug Degenstein, and Yves Rochon. The study was initially proposed by the author and Dylan Jones, motivated by the initial MSA framework designed by Martin Keller. The GEOS-Chem adjoint model was designed by Daven Henze. Observation operators for MOPITT CO, TES O_3, OMI NO_2, and OSIRIS O_3 were prepared by Zhe Jiang, Dylan Jones, Martin Keller, and Thomas Walker respectively. Helen Worden, Adam Bourassa, Doug Degenstein and Yves Rochon have provided some instructions on using MOPITT CO and OSIRIS O_3 observations. The implementation of all the observations in the MSA framework, design of the super-observation and cost function scaling algorithm as well as the research analyses were all performed by the author.

Chapter 4 was a collaboration with Dylan Jones, Martin Keller, Kazuyuki Miyazaki, Zhe Jiang, Daven Henze, Helen Worden, Adam Bourassa, Doug Degenstein, Cathy Clerbaux, Debra Wunch, Steven Wofsy, and Anne Thompson. The study was proposed by the author and Dylan Jones, motivated by the MSA phase one. Besides the ongoing support from the co-authors mentioned in Chapter 3, Cathy Clerbaux, Debra Wunch, Steven Wofsy and Anne Thompson have provided the data with its bias correction updates for IASI O_3, TCCON XCO, ATom measurements, and SHADOZ O_3 observations respectively. The CO and NO_x emission estimates as well as O_3 and OH fields from Tropospheric Chemical Reanalysis version 2 (TCR-2) dataset were provided by Kazuyuki Miyazaki. The top down CO emission estimates according to *Jiang et al.* (2017) were provided by Zhe Jiang. The re-design of the super-observation, further design of the observation operator for IASI O_3, MLS HNO_3 and OMI HCHO, implementation of the updated MSA scheme as well as the research analyses were all performed by the author.

Chapter 5 was a collaboration with Dylan Jones, Cristen Adams, Martin Keller, Adam Bourassa, Doug Degenstein, Cathy Clerbaux and Daniel Jacob. The project was proposed by the author and Dylan Jones, motivated by the successful retrieval on OmO NO_2 observations led by Cristen Adams. Martin Keller has designed the weak constraint 4D-Var scheme in GEOS-Chem model. OmO NO_2 observations

were prepared by Cristen Adams, with the algorithm instruction provided by Adam Bourassa and Doug Degenstein. The IASI data and its retrieval algorithms were provided by Cathy Clerbaux. The ARCTAS data was made available from the measurement team led by Daniel Jacob. The design of the observation operator for OmO NO_2, implementation of the regional NO_x framework using the assimilated boundary conditions as well as the research analyses were all performed by the author.

Chapter 2

Methods

2.1 Observations

2.1.1 Assimilated satellite observations

MOPITT CO retrievals

MOPITT is a multi-spectral nadir instrument on the EOS-Terra satellite that was launched in December 1999. The satellite is in a sun-synchronous orbit with a local equatorial crossing time of 10:30/22:30. MOPITT measurement strategy consists of a 612 km cross-track scan with spatial resolution of 22 km x 22 km at nadir, which achieves global observational coverage every three days. Figure 2.1 shows the global coverage for one day of MOPITT observations. Tropospheric CO is retrieved from measurements at 4.7 μm in the thermal infrared (TIR) region of the spectrum and at 2.3 μm in the near infrared (NIR). The retrieved CO is estimated on 10 pressure levels (the surface, 900, 800, 700, 600, 500, 400, 300, 200 and 100 hPa) using the optimal estimation approach (*Rodgers*, 2000). The cost function representing differences between the simulated and observed radiances, and between the retrieved and a priori state vectors, is minimized by optimizing the state vector. The mathematical derivation of the optimal estimation approach will be discussed in Section 2.4.2. The MOPITT CO retrieval ($\mathbf{z}^{\mathrm{MOP}}$) can be written as:

$$\mathbf{z}^{\mathrm{MOP}} = \mathbf{z_a}^{\mathrm{MOP}} + \mathbf{A}^{\mathrm{MOP}}(\mathbf{z} - \mathbf{z_a}^{\mathrm{MOP}}) + \epsilon, \tag{2.1}$$

where $\mathbf{z}$ is the true CO state (expressed in based 10 logarithm: $log(CO_{true})$), $\mathbf{z_a}^{\mathrm{MOP}}$ (expressed in $log(CO_{ap})$) is the a priori MOPITT profile, which was obtained from simulations from the Model for Ozone And Related Tracers version 4 (MOZART-4, a CTM), ϵ is the retrieval error, and $\mathbf{A}^{\mathrm{MOP}}$ is the averaging kernel matrix, which represents the ver-

tical sensitivity of the retrieval to the true CO state in the atmosphere. Figure 2.2 shows the mean profiles for the normalized MOPITT averaging kernel observed over Toronto, Canada, for the multi-spectral NIR+TIR retrievals. The total column averaging kernel is indicated by the black solid line on the right, and peaks in the free troposphere. There is less sensitivity near the surface and in the UTLS, suggesting greater contributions of a prior information in the retrievals in these regions. Details of the MOPITT retrieval are described in *Worden et al.* (2010) and *Deeter et al.* (2011).

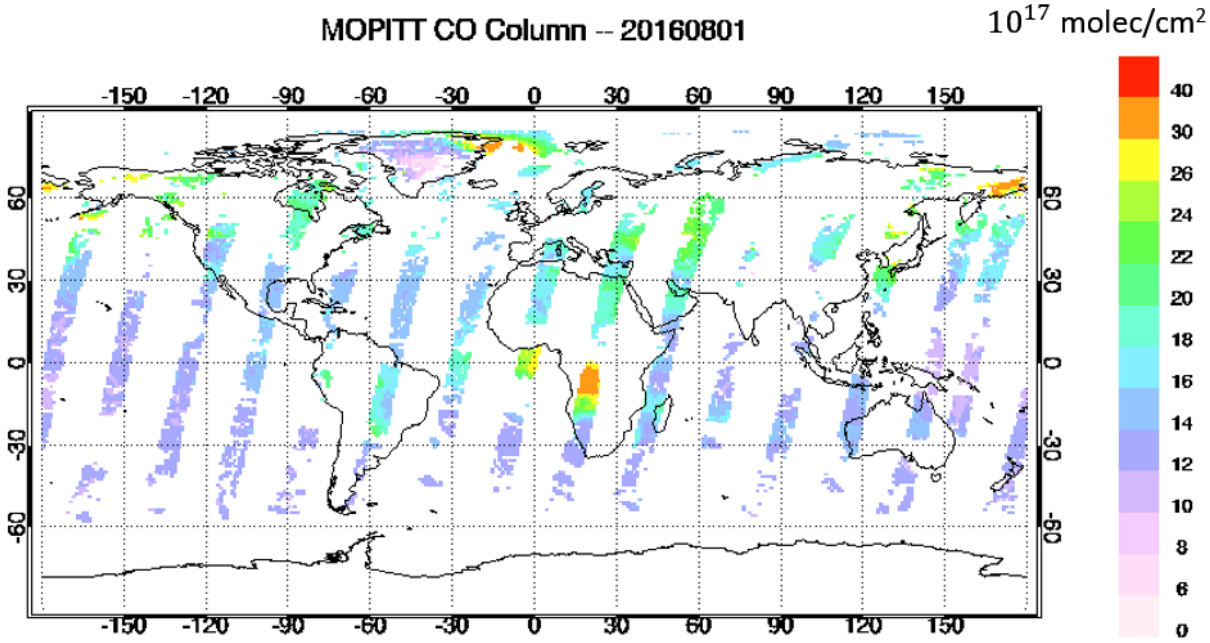

Figure 2.1: Example of the MOPITT observational coverage for a given day. Shown are the MOPITT V7J CO column densities (in 10^{17} molec/cm^2) on 1 August 2016. (*NCAR: ACOM Science Team*, 2016)

This study uses version 7 of the multi-spectral NIR+TIR retrievals (V7J), which were validated by *Deeter et al.* (2017) using NOAA flask and HIAPER Pole-to-Pole Observations (HIPPO) aircraft measurements. They found smaller biases in the V7J retrievals compared to V6 data. I use only daytime profile retrievals located at altitudes below 200 hPa and equatorward of 60°. CO profiles are assimilated instead of CO column abundances because the profiles provide information on CO vertical structure (*Kar et al.*, 2008). The degrees of freedom for signal (DOFS) for the retrieved profiles range from 0.5 to 2 within 60° (*Heald et al.*, 2004), which is conducive to constraining both surface emissions and free tropospheric CO.

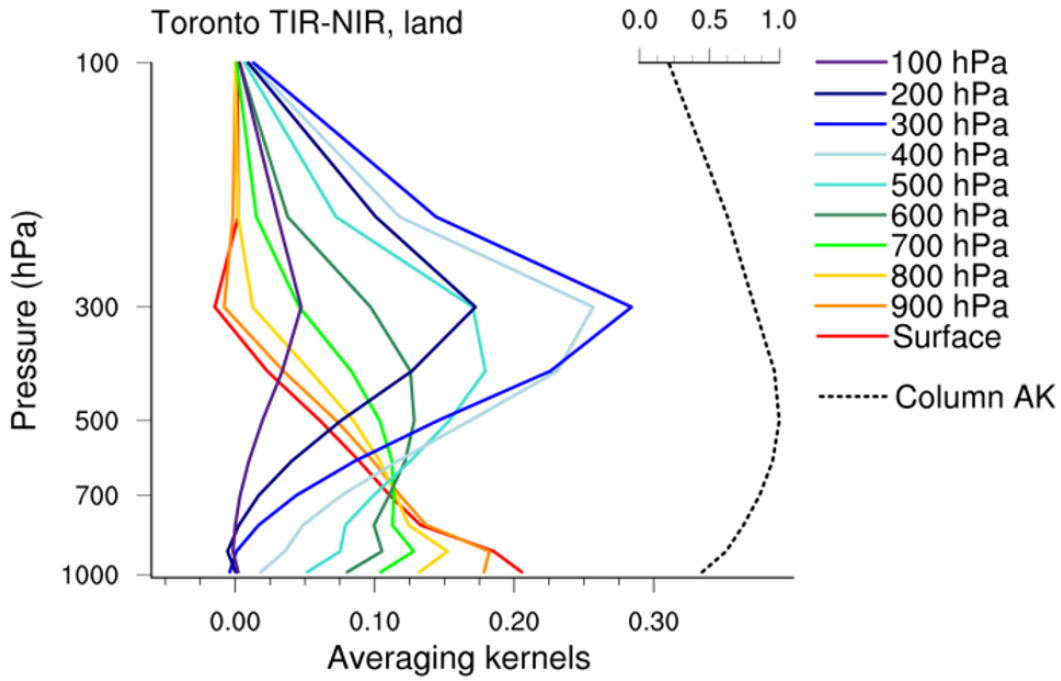

Figure 2.2: Normalized averaging kernel observed over Toronto, Canada, for MOPITT CO V7J retrievals. (*Buchholz et al.*, 2017)

TES O_3 retrievals

TES is a Fourier-transform infrared (FTIR) spectrometer that was launched on the EOS Aura satellite in July 2004 (*Beer et al.*, 2001). It is in a sun-synchronous orbit at an altitude of 705 km with an equatorial crossing time of 13:30 ± 15 minute for the ascending node. The two modes operated by the instrument include a global survey mode where observations are spaced about 2° along the orbit track, and a step and stare mode where each observation is spaced about 40 km along the orbit. The spatial resolution of the measurement is 8 km × 5 km. TES measures radiance spectra of Earth's atmosphere at wavelengths between 3.3 - 15.4 μm with a spectral resolution of 0.1 cm^{-1} (*Verstraeten et al.*, 2013). The instrument completes 14.5 orbits per day and achieves global coverage every 16 days (*Worden et al.*, 2007). An example of the observation coverage from one day of observations is shown in Fig. 2.3. Details of the instrumentation are explained in *Beer et al.* (2001). Tropospheric O_3 is retrieved from measurements around 9.6 μm (*Jourdain et al.*, 2007).

The retrievals use a similar optimal estimation method as the MOPITT CO retrievals (*Bowman et al.*, 2006). Details of the TES O_3 retrievals are described in *Bowman et al.* (2006). The retrievals are conducted with respect to the natural logarithm (ln) of the volume mixing ratio on a 67-level vertical grid with a grid spacing of about 1 km.

The TES O_3 profile retrievals can be expressed mathematically as:

$$\mathbf{z}^{\mathrm{TES}} = \mathbf{z_a}^{\mathrm{TES}} + \mathbf{A}^{\mathrm{TES}}(\mathbf{z} - \mathbf{z_a}^{\mathrm{TES}}) + \epsilon, \tag{2.2}$$

where $\mathbf{z_a}^{\mathrm{TES}}$ is the TES a priori O_3 profile, $\mathbf{z}$ is the true O_3 state, ϵ is the retrieval error, and $\mathbf{A}^{\mathrm{TES}}$ is the TES averaging kernel. Figure 2.4 shows the TES averaging kernel observed over De Bilt, the Netherlands, on 14 July 2005. The vertical sensitivity for retrieval below 400 hPa (shown in red solid curves) peaks around 600 hPa. The sensitivity for 400-150hPa retrieval levels (shown in green solid curves) peaks around 200 hPa. None of the retrieval levels are sensitive to the surface, suggesting the a priori information dominates the surface O_3 retrievals. In this thesis, TES O_3 retrievals (Level 2 data) from the version 4 product are used for the analysis in Chapter 3. This version of the data has good agreement in the troposphere with later versions of the retrievals (*Herman and Osterman*, 2014). TES O_3 retrievals were evaluated by *Nassar et al.* (2008) and *Boxe et al.* (2010). Both studies identified positive biases in TES O_3 retrievals from the surface to the upper stratosphere, with a strong dependency on altitude and latitude. In this study, these spatially dependent biases are removed to improve the quality of the TES retrievals. Typically, TES O_3 profiles have 1-2 DOFS in the troposphere. In this thesis, only retrievals equatorward of 60° with the Master quality flag, cloud fraction<15%, and solar zenith angle<84° are assimilated. This vertically-resolved information is vital for concentration (initial condition) optimization in this assimilation framework. However, the TES global survey observations were phased out after 2011, so for later stages of the work in this thesis (e.g. Chapter 4), O_3 retrievals from IASI will be assimilated.

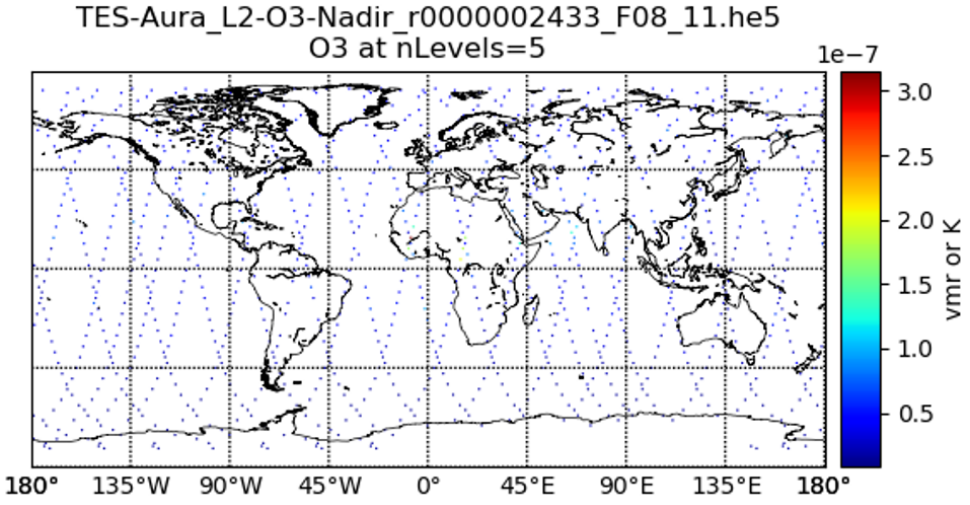

Figure 2.3: TES O_3 on TES retrieval level 5 (approximately 5 km) on 2 November 2009. (*TES Science Team*, 2008)

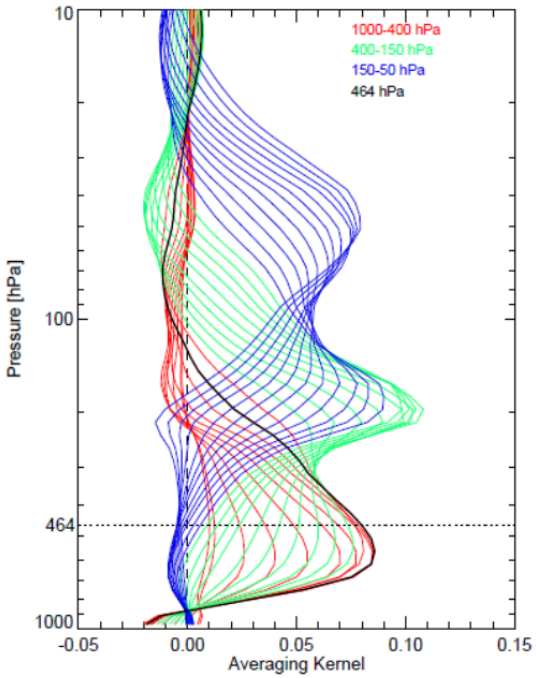

Figure 2.4: TES averaging kernels for retrievals over at De Bilt, the Netherlands, on 14 July 2005. (*Verstraeten et al.*, 2013)

IASI O_3 partial columns

The IASI instrument is a FTIR spectrometer onboard the Meteorological Operational (Metop) Satellites, which include Metop-A (IASI-A), Metop-B (IASI-B), and Metop-C (IASI-C). (Note that data from IASI-C are not used in this work.) Metop-A was launched in October 2006. Metop-B was launched in September 2012. The instruments onboard both satellites are still in operation. The Metop satellites are in a sun-synchronous orbit at an altitude of approximately 817 km, with an equatorial crossing time of around 09:30 local time in the morning (and 21:30 in the evening). IASI is nadir viewing and measures radiation emitted by the Earth's atmosphere between 3.62 - 15.5 μm. Similar to TES, the IASI instrument also completes around 14 orbits a day. But because the measurement consists of a wide 2200 km swath, the instrument reaches global observational coverage twice a day (see Fig. 2.5). Full details of the instrumentation are described in *Clerbaux et al.* (2009). The O_3 retrieval is carried out in the spectral range of 9.3 - 9.8 μm. The full description of the Fast Optimal Retrievals on Layers for IASI (FORLI), which is based on an optimal estimation approach, can be found in *Hurtmans et al.* (2012). As with MOPITT and TES, the IASI O_3 retrievals can be expressed as:

$$\mathbf{z}^{\mathrm{IASI}} = \mathbf{z_a}^{\mathrm{IASI}} + \mathbf{A}^{\mathrm{IASI}}(\mathbf{z} - \mathbf{z_a}^{\mathrm{IASI}}) + \epsilon, \tag{2.3}$$

where $\mathbf{z}$ is the true O_3 partial columns, $\mathbf{z_a}^{\text{IASI}}$ represents a single IASI priori profile retrieved from long-term satellite limb observations and ozonesonde measurements (*McPeters et al.*, 2007), ϵ is the retrieval error, and $\mathbf{A}^{\text{IASI}}$ is the averaging kernel matrix. Fig. 2.6 shows the IASI-A averaging kernels for Nassriya, Iraq, on 15 July 2014. Similar to TES, the vertical sensitivity to tropospheric O_3 peaks in the mid-troposphere (between 6 and 8 km). There is greater sensitivity to O_3 near the surface in this retrieval than in the TES retrieval shown in Fig. 2.4 because of the high thermal contrast at the surface for this scene in Iraq.

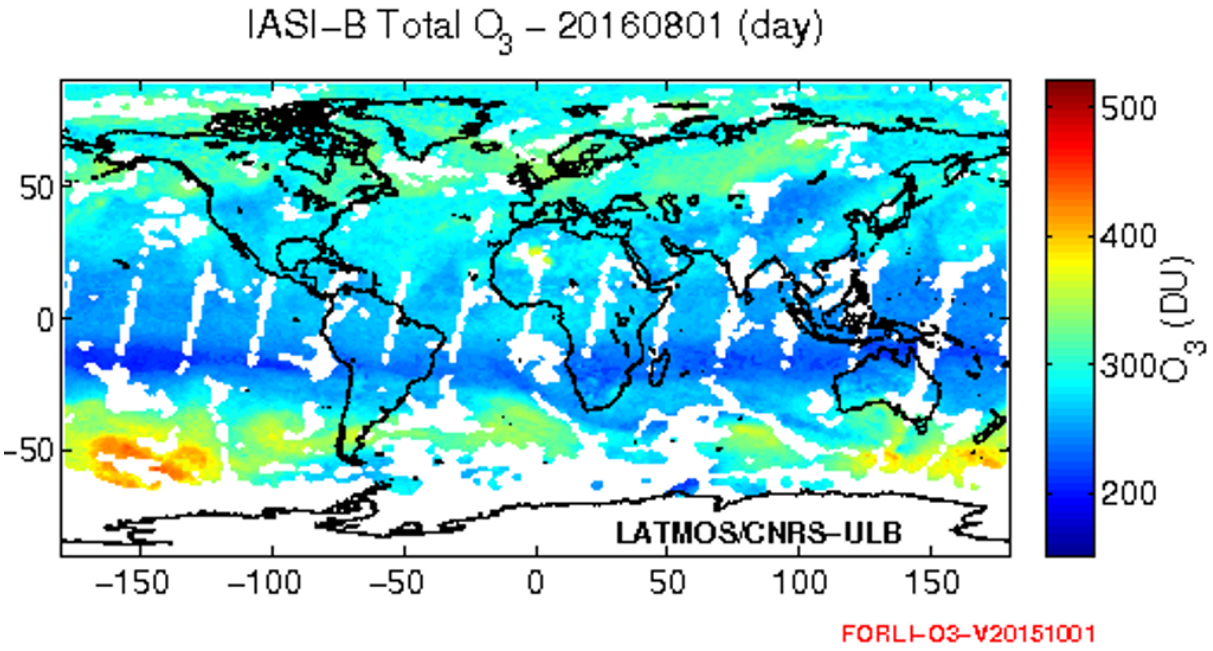

Figure 2.5: IASI-B FORLI (v20151001) total O_3 columns (in DU) on 1 August 2016. (*IASI Science Team*, 2017)

IASI-B O_3 partial columns retrievals with averaging kernel information from the FORLI v20151001 product are used in Chapter 4. In Chapter 5, since the regional NO_x study period covers 2008-2011 (before the launch of IASI-B), I use the FORLI v20151001 O_3 retrievals from only the IASI-A instrument. The retrievals as well as the validation of IASI-FORLI O_3 data are described in *Boynard et al.* (2016). According to their study, the differences between IASI-A and IASI-B O_3 partial columns were within 2% over all altitudes. The daytime data and the nighttime data are both useful. Based on their validation against aircraft data and ozonesondes, the bias correction discussed in *Boynard et al.* (2016) is applied to the data used in this thesis. Moreover, all observations either with cloud fraction larger than 10% or with root mean square error (RMSE) of the spectral fit larger than 3.4×10^{-8} W cm^{-2} sr^{-1} cm^{-1} are rejected in the analysis. Since the

focus here is on optimizing tropospheric O_3, I only assimilate IASI O_3 partial columns at latitude between 60 °S and 75 °N, altitudes below 300 hPa. Additional quality control includes quality flag=0, solar zenith angle<80°, cloud fraction<0.13, and DOFS>3.0. It is an advantage using IASI data because the instrument provides consistent high data density for each assimilation period; throughout 2016 there are more than 10^5 observations available daily. The O_3 profile has total DOFS of 3.0-4.2 with two maxima located within the troposphere (at 2 and 8 km), which is useful for constraining the modelled O_3 concentrations in the lower- and mid-troposphere.

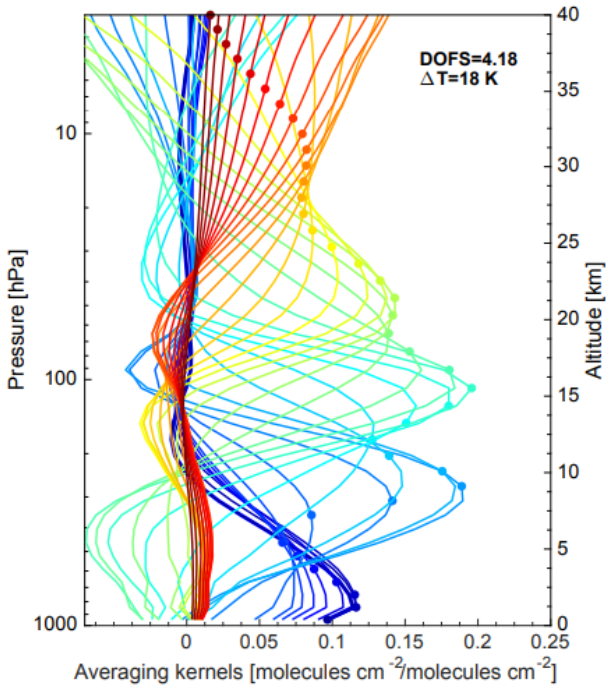

Figure 2.6: Averaging kernels for IASI(A)-FORLI O_3 retrievals on 15 July 2014 over Nassriya, Iraq. (*Boynard et al.*, 2016)

OSIRIS O_3 retrievals

The Canadian-made OSIRIS is a limb instrument onboard the Swedish Odin satellite that was launched in February 2001. Odin flies in a polar, sun-synchronous orbit with equator crossing time at 18:00 local time on the ascending track and 06:00 local time on the descending track. The satellite provides year-round coverage in the tropics and spring/fall hemisphere, but no coverage in the winter hemisphere. The satellite completes 15 orbits daily with an orbit period of 96 min. The OSIRIS observational coverage on 2 November 2009 is shown in Fig. 2.7. The OSIRIS spectrograph measures scattered sunlight between 280 nm to 810 nm with a spectral resolution of 1 nm. Observations are

made with cross track resolution of between 6 and 60 km and an along track resolution from 500 to 1000 km, with a vertical resolution of 2 km (*Llewellyn et al.*, 2004). In normal operational mode, OSIRIS scans the limb from 7 to 70 km in tangent altitude. The retrieval method of OSIRIS is the Multiplicative Algebraic Reconstruction Technique (MART), which combines ozone absorption information in both the ultraviolet (UV) and visible parts of the spectrum to retrieve number density profiles from 10 km (clear sky) or cloud top to 60 km. The radiative transfer model is the SASKTRAN model designed by *Bourassa et al.* (2012). Aerosol and NO_2 are also retrieved along with O_3. The full retrieval algorithm is explained in *Roth et al.* (2007) and *Degenstein et al.* (2009)

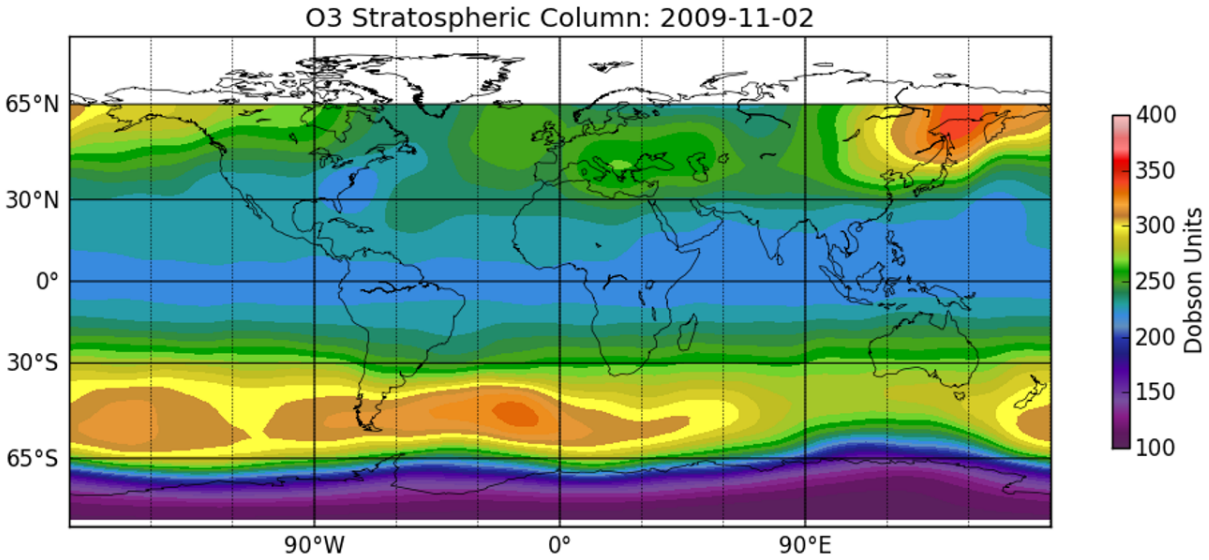

Figure 2.7: Gridded O_3 stratospheric columns on 2 November 2009 retrieved from measurements from the OSIRIS instrument (*Degenstein et al.*, 2009).

This thesis uses the O_3 number density vertical profile retrieved from SaskMART O_3 version 5 product. *Bourassa et al.* (2012) performed an error analysis on the retrieved OSIRIS O_3 concentrations and found that there was less than a 2% difference between OSIRIS and Stratospheric Aerosol and Gas Experiment II (SAGE II) data between 18-53 km. However, much larger biases were found at levels below 18 km as well as at levels above 53 km. As a result, only the data located between 18 and 53 km, with the observation error smaller than 1% of the observed O_3 abundance, are used in my assimilation. Without averaging kernels given in the retrievals, I use the mean O_3 concentrations on each model level interpolated from the OSIRIS retrieval levels. It is noted that the instrument's performance has degraded recently. Despite that, there are sufficient (more than 100) OSIRIS observations available daily to provide stratospheric O_3 constraints for my assimilation.

OMI tropospheric NO_2 columns

The OMI instrument shown in Fig. 2.8 is a push broom UV-visable (UV-vis) spectrograph onboard the Aura satellite. The OMI measurement consists of cross track swath of 2600 km comprised of 13 km $\times$ 24 km pixels. This viewing strategy allows OMI to achieve global coverage daily. However, since 2007, there has been a physical blockage of the entrance optics, which has affected many measurements (the "row anomaly"). As a result of this loss of data, OMI data now achieves the global coverage about every two days. The instrument measures direct and scattered sunlight in the wavelength range between 270 nm to 550 nm. The retrieval process for the standard tropospheric NO_2 product (OMI-SP) is described in *Bucsela et al.* (2006). It should be noted that this retrieval algorithm is different from that used for the other OMI retrieval products, such as the Dutch OMI NO_2 product (OMI-DOMINO) and OMI-minus-OSIRIS. The OMI-SP is based on a two-step approach (*Krotkov et al.*, 2017). First, an NO_2 slant column (S) is estimated by fitting earth-shine radiances in the wavelength range between 405 to 465 nm using differential optical absorption spectroscopy (DOAS) technique (*Sigrist et al.*, 1994). Second, for each latitude, a reference sector correction is applied by choosing remote areas where the total slant columns are dominated by their stratospheric contributions, providing an estimate of the stratospheric NO_2 slant columns at all longitudes for the corresponding latitude. This inferred stratospheric slant column is then subtracted from the total slant column to retrieve the tropospheric NO_2 slant column. The tropospheric NO_2 vertical column is then computed using the tropospheric slant NO_2 columns with the air mass factor (AMF) computed by the Total Ozone Mapping spectrometer RADiative transfer model (TOMRAD) (*Bucsela et al.*, 2006).

In this thesis, I use the OMI-SP V003 tropospheric NO_2 product, which was validated by *Krotkov et al.* (2017). An example of the OMI NO_2 data are shown in Fig. 2.9. High concentrations can be observed over the eastern US, Europe, East Asia and equatorial Africa. Uncertainties of the NO_2 column abundance depend on air mass factors (AMFs) and measurement locations. To reduce the potential adverse impact of poor quality retrievals, I assimilate only retrievals equatorward of 60° with the following quality flag controls: cloud fraction < 0.2, quality flag $= 0$, viewing zenith angle $< 65°$, and solar zenith angle $< 75°$. *Krotkov et al.* (2017) validated the OMI-SP V003 data and found that the slant column in V003 of the standard product is 10-40% lower than V002 and has better agreement with independent satellite and in-situ measurements over unpolluted areas (*Krotkov et al.*, 2017). The column data, despite lack of vertical sensitivity, are still used to constrain both surface NO_x emissions as well as the NO_2 concentration in the free troposphere.

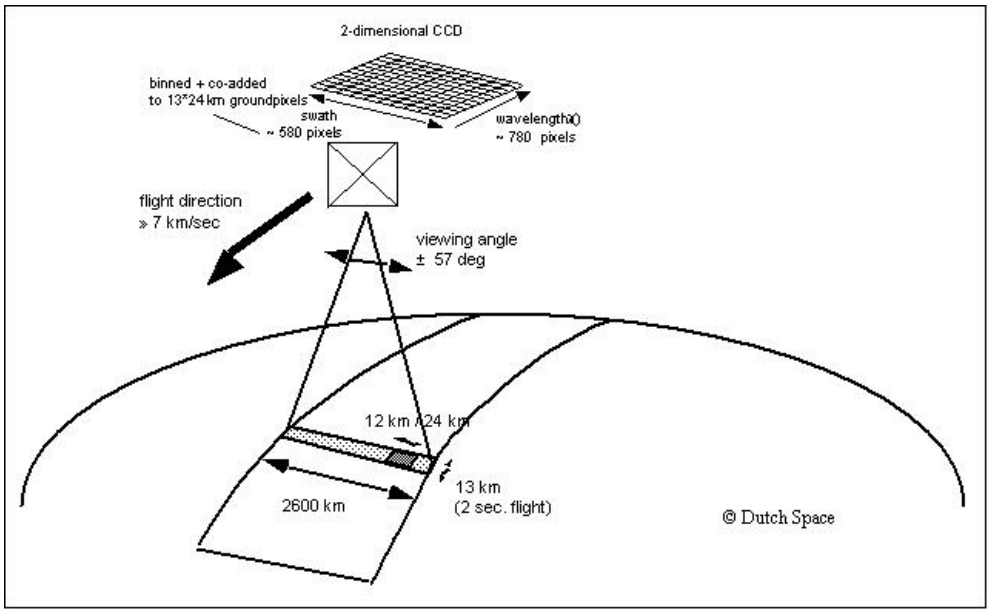

Figure 2.8: The nadir field of view with retrieval pixel information for the OMI instrument (*Dobber et al.*, 2006).

OMI tropospheric HCHO columns

The HCHO (also noted as CH_2O or H_2CO) retrievals assimilated in this study are from version 3 of the Smithsonian Astrophysical Observatory retrieval algorithm of OMI observations (OMI-SAO) product, which is described by *González Abad et al.* (2015). The retrieval process involves a similar two step approach as for NO_2. The first step is the radiance fitting to retrieve slant column densities (SCDs) using the basic optical absorption spectroscopy (BOAS) method (*Chance*, 1998). The second step is the vertical column density (VCD) calculation based on tropospheric SCDs and AMF information retrieved from Vector-based Linearized Discrete Ordinate Radiative Transfer (VLIDORT, an RTM) version 2.4. Tropospheric slant columns were retrieved using the reference sector correction algorithm described by *Barkley et al.* (2013), which involves calculation on air mass factor (AMF) and scattering weight. The long term record for OMI-SAO V003 has been validated by *Zhu et al.* (2017). An example of the observational coverage of HCHO from the OMI-SAO V003 product is shown in Fig. 2.9. High column abundances are distributed over the southeastern US, the Amazon, Europe and southeastern Asia. However, the retrievals are noisy, compared to the OMI NO_2 retrievals. The uncertainties of the HCHO SCDs for each single measurement range from 40% over HCHO hotspots to 100% in remote regions. The OMI HCHO data are comparable to HCHO retrievals from other satellite instruments, such as GOME-2A, GOME-2B, and Ozone Mapping and Profiler Suite (OMPS) (*Zhu et al.*, 2017). However, comparisons have shown that

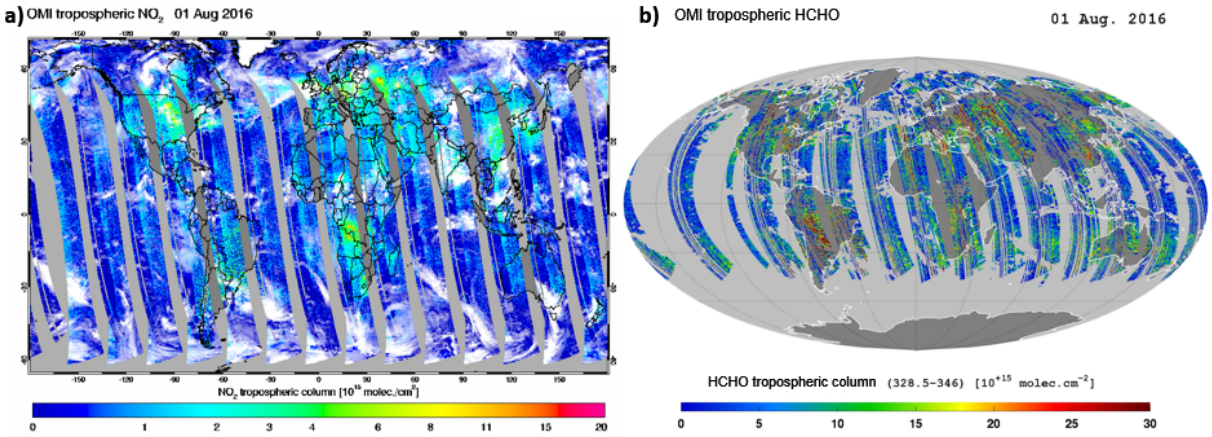

Figure 2.9: Sample OMI tropospheric (left) NO_2 columns from *Krotkov et al.* (2018) and (right) HCHO columns from *Chance* (2007) on 1 August 2016.

OMI HCHO retrievals over hotspots could be 30% lower than those from GOME and SCIAMACHY. *Zhu et al.* (2017) also found that all the retrievals within the southeastern US could be biased low by up to 50%. In this study, a uniform bias correction, with a correction factor of 1.4 is applied globally, following the bias correction in recent OMI HCHO assimilation studies (*Kaiser et al.*, 2018). The quality flags similar to OMI NO_2 are also used: cloud fraction<0.2, main quality flag=0, cross track quality flag=0, viewing zenith angle<65°, and solar zenith angle<75°. OMI HCHO retrievals have been assimilated to provide "top-down" constraints on surface isoprene emissions because tropospheric HCHO is a signature chemical product of isoprene oxidation (*Palmer et al.*, 2003b). However, the inferred isoprene emissions can have large uncertainties due to large retrieval errors and to chemical biases in the atmospheric models used in the inversion analyses.

MLS HNO_3 retrievals

The MLS instrument is also onboard the Aura satellite. It is a limb instrument measuring millimetre and sub-millimetre thermal emission. The instrument has seven radiometers covering the spectral range between 118 GHz and 2.5 THz. An example of the observational coverage for a given day is shown in Fig. 2.10. The cross track and along track resolution are 10 km and 165 km, respectively. There are 240 limb scans per orbit which provide more than 3000 profiles per day. The retrieval method of MLS HNO_3 uses the optimal estimation method described by *Livesey et al.* (2011). Fig. 2.11 shows the full width at half maximum of the rows of the averaging kernel matrix for MLS HNO_3 data. The vertical resolution is 3-4 km in the upper troposphere and lower stratosphere. It

should be noted that the peak of the averaging kernel for each pressure level suggests the retrieval at each vertical level is dominated by the HNO_3 abundance observed at the corresponding level. Moreover, according to the integrated vertical sensitivity (shown as solid black), retrievals at altitudes below 200 hPa are dominated by the a priori information. In Chapter 4, I use version 4.2 of the MLS HNO_3 product. I only use HNO_3 located at altitudes above 150 hPa and below 15 hPa with the following data filtering suggested by *Livesey et al.* (2011): precision > 0, Status flag = an even number, Quality > 0.5, Convergence < 1.4, and Cloud fraction $< 10\%$. *Santee et al.* (2017) validated the HNO_3 retrievals against aircraft measurements and found that the precision throughout the lower and middle stratosphere is about 0.7 ppbv. There are systematic biases that vary between ± 0.5 and ± 2 ppbv (or 5 to 15%) throughout the stratosphere. In this thesis, MLS HNO_3 retrievals are used to constrain modelled HNO_3 as well as lightning NO_x emissions. As explained in *Miyazaki et al.* (2015), assimilation of MLS HNO_3 data could correct the modelled NO_2 biases through chemical feedbacks between NO_x, O_3, and HNO_3.

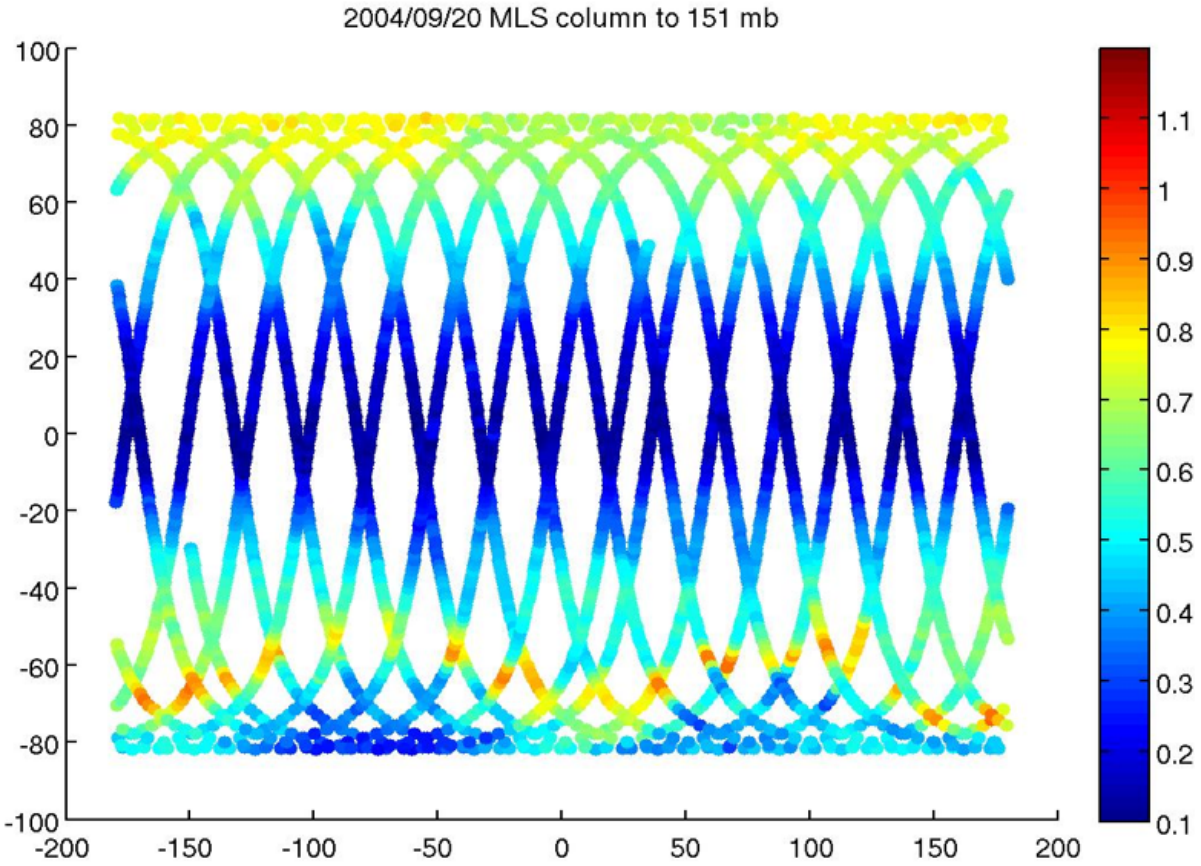

Figure 2.10: MLS HNO_3 columns (integrated down to 150 hPa) observed on 20 September 2004. (*EOS MLS Science Team*, 2011)

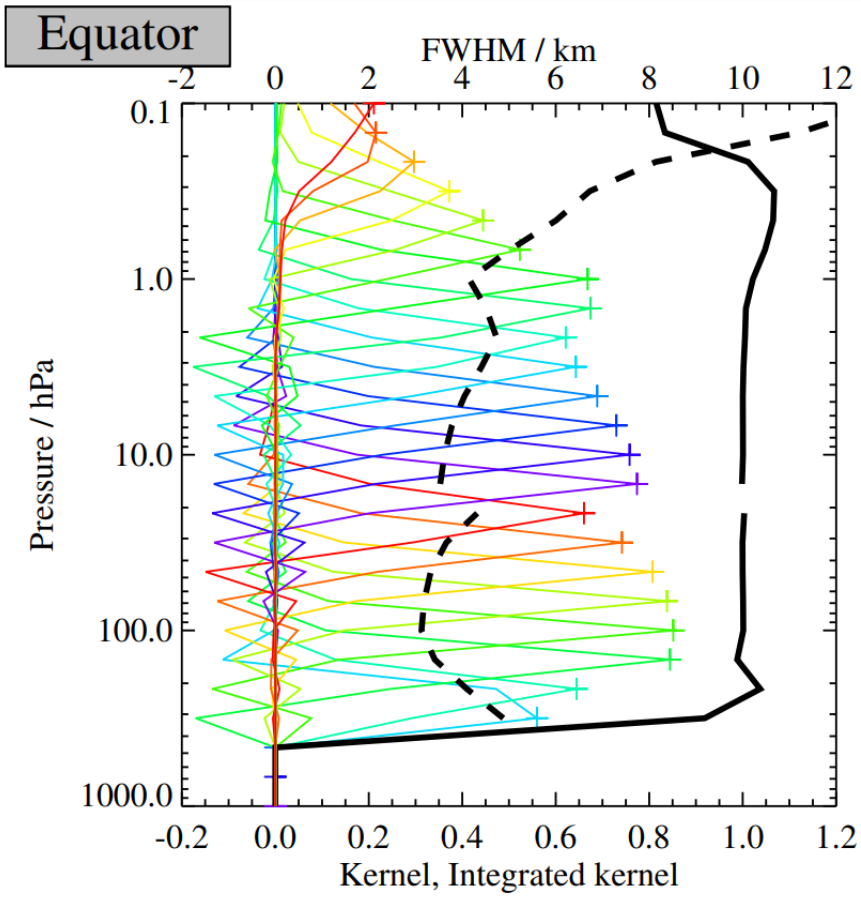

Figure 2.11: Colored lines show the averaging kernels as a function of MLS retrieval level. The dashed black line shows the full width at half maximum (FWHM) of the averaging kernels, approximately scaled into kilometers (top axis).The discontinuity shows the transition between the 240-GHz retrievals and the 190-GHz retrievals. The solid black line shows the integrated area under each kernel. The value approaching unity shows the retrieved information comes from the measurements. (*Santee et al.*, 2017)

OMI-minus-OSIRIS (OmO) tropospheric NO_2 columns

The OmO NO_2 dataset was produced by the limb-nadir matching technique using NO_2 total columns retrieved from OMI-SP V003 NO_2 and stratospheric SCD retrieved from the OSIRIS NO_2 V5.07, which have been described above. The limb-nadir matching technique was described in *Adams et al.* (2016). The total columns of NO_2 from OMI and OMO be written as

$$\mathrm{SCD}_{\mathrm{tot}}^{\mathrm{OMI}} = \mathrm{SCD}_{\mathrm{strat}}^{\mathrm{OMI}} + \mathrm{SCD}_{\mathrm{tropo}}^{\mathrm{OMI}} = \mathrm{VCD}_{\mathrm{strat}}^{\mathrm{OMI}} \cdot \mathrm{AMF}_{\mathrm{strat}}^{\mathrm{OMI}} + \mathrm{VCD}_{\mathrm{tropo}}^{\mathrm{OMI}} \cdot \mathrm{AMF}_{\mathrm{tropo}}^{\mathrm{OMI}} \quad (2.4)$$

$$\mathrm{SCD}_{\mathrm{tot}}^{\mathrm{OmO}} = \mathrm{SCD}_{\mathrm{strat}}^{\mathrm{OSI}} + \mathrm{SCD}_{\mathrm{tropo}}^{\mathrm{OmO}} = \mathrm{VCD}_{\mathrm{strat}}^{\mathrm{OSI}} \cdot \mathrm{AMF}_{\mathrm{strat}}^{\mathrm{OSI}} + \mathrm{VCD}_{\mathrm{tropo}}^{\mathrm{OmO}} \cdot \mathrm{AMF}_{\mathrm{tropo}}^{\mathrm{OmO}} \quad (2.5)$$

The bias-corrected OMI SCDs with correction factor γ can be written as a combination of the OSIRIS stratospheric VCD and the OmO tropospheric VCD as the following mathematical expression:

$$\gamma \cdot \mathrm{SCD}_{\mathrm{tot}}^{\mathrm{OMI}} = \mathrm{VCD}_{\mathrm{strat}}^{\mathrm{OSI}} \cdot \mathrm{AMF}_{\mathrm{strat}}^{\mathrm{OMI}} + \mathrm{VCD}_{\mathrm{tropo}}^{\mathrm{OmO}} \cdot \mathrm{AMF}_{\mathrm{tropo}}^{\mathrm{OMI}} \tag{2.6}$$

In turn, the OmO tropospheric VCD can be written as:

$$\mathrm{VCD}_{\mathrm{tropo}}^{\mathrm{OmO}} = \frac{\gamma \cdot \mathrm{SCD}_{\mathrm{tot}}^{\mathrm{OMI}} - \mathrm{VCD}_{\mathrm{strat}}^{\mathrm{OSI}} \cdot \mathrm{AMF}_{\mathrm{strat}}^{\mathrm{OMI}}}{\mathrm{AMF}_{\mathrm{tropo}}^{\mathrm{OMI}}} \tag{2.7}$$

The OmO NO_2 VCDs are calculated by removing the stratospheric VCD of OMI NO_2 using the non-coincident NO_2 profiles retrieved from OSIRIS. The mismatch in the non-coincident profiles between OMI and OSIRIS measurements are corrected using a photochemical box model (*McLinden et al.*, 2000) and the correction factor γ. The box model is run for each altitude layer of the OSIRIS NO_2 profile and is based on the stratospheric chemistry scheme imposed by *Brohede et al.* (2008). For NO_x-related species in the model, O_3 is based OSIRIS retrievals, whereas N_2O, CH_4, NO_y, Cl_y and Br_y are climatological estimates from the Canadian Middle Atmosphere Model (CMAM) (*Jonsson et al.*, 2004). After reducing the mismatch, the STS is then applied to subtract the OSIRIS interpolated NO_2 stratospheric columns from the OMI total columns based on the date, location, and local time of each OMI measurement. Full details of the limb-nadir matching technique are explained in *Adams et al.* (2016). The OmO NO_2 VCD is then retrieved using the same AMF and scattering weight information used in the OMI-SP product described by *Krotkov et al.* (2017). Fig. 2.12 compares the mean and standard deviation of the OMI and OSIRIS stratospheric VCDs during June-July-August (JJA) for 2008-2011. Between 20°N-60°N, the OSIRIS stratospheric VCDs are 0.4-0.6$\times 10^{15}$ molec/cm^2 smaller than the OMI stratospheric VCDs. The spread of the JJA mean stratospheric VCD between OMI and OSIRIS over North America are comparable. Fig. 2.13 compares the mean and standard deviation of tropospheric VCDs during JJA. Tropospheric VCDs north of 45°N are significantly larger in the OMI-SP product than in the OmO product for all three months. In fact, the OmO tropospheric VCDs in northern high latitudes are close to 0 molec/cm^2. Between 5° and 45°N, the OmO mean tropospheric VCD range is 2.5$\times 10^{14}$ molec/cm^2, slightly larger than the OMI-SP mean tropospheric VCDs (approximately 1.9$\times 10^{14}$ molec/cm^2). It is also noticeable that the month-to-month variation for OmO is slightly larger than for OMI-SP. The standard deviation for the OmO NO_2 product is also higher than that for OMI-SP, suggesting a

higher variability for OmO tropospheric VCDs.

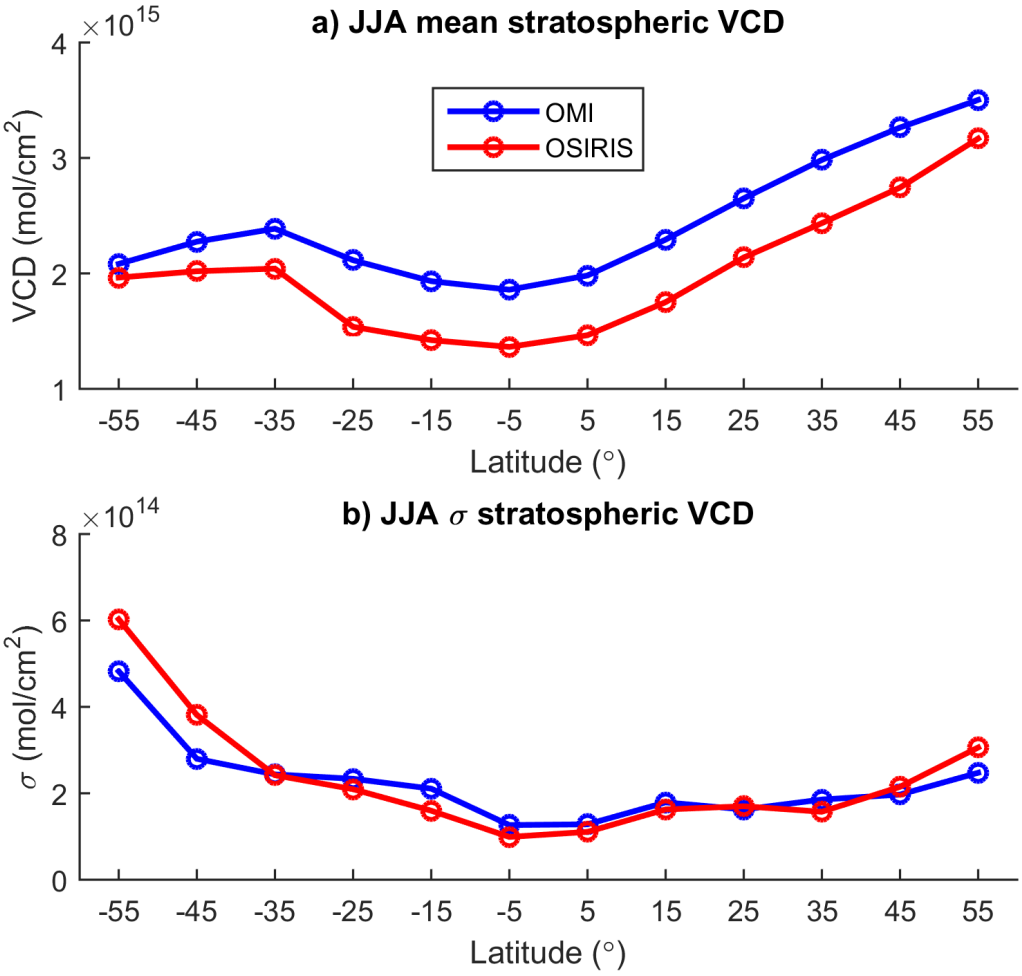

Figure 2.12: The mean and standard deviation of global stratospheric VCDs from OMI-SP and OmO retrievals. They are calculated from the VCDs in each latitude band during JJA for 2008-2011.

In this study, I assimilate OmO tropospheric NO_2 slant columns to quantify North American NO_x emissions. Following the suggestion of *Adams et al.* (2016), only measurements that met the following criteria were assimilated: cloud fraction<0.3, solar zenith angle<75°, and vertical azimuth angle<65°. Since the observation coverage of OSIRIS is limited to the summer hemisphere, only data for the summer season (JJA) over North America were used. Our analysis focus on the period of 2008-2011 since the data density and retrieval quality during that period is the best for all three of the instruments (IASI, OMI, OSIRIS) involved in my study. The differences in tropospheric VCD between OMI SP and OmO as well as their corresponding assimilated surface NO_x emissions are

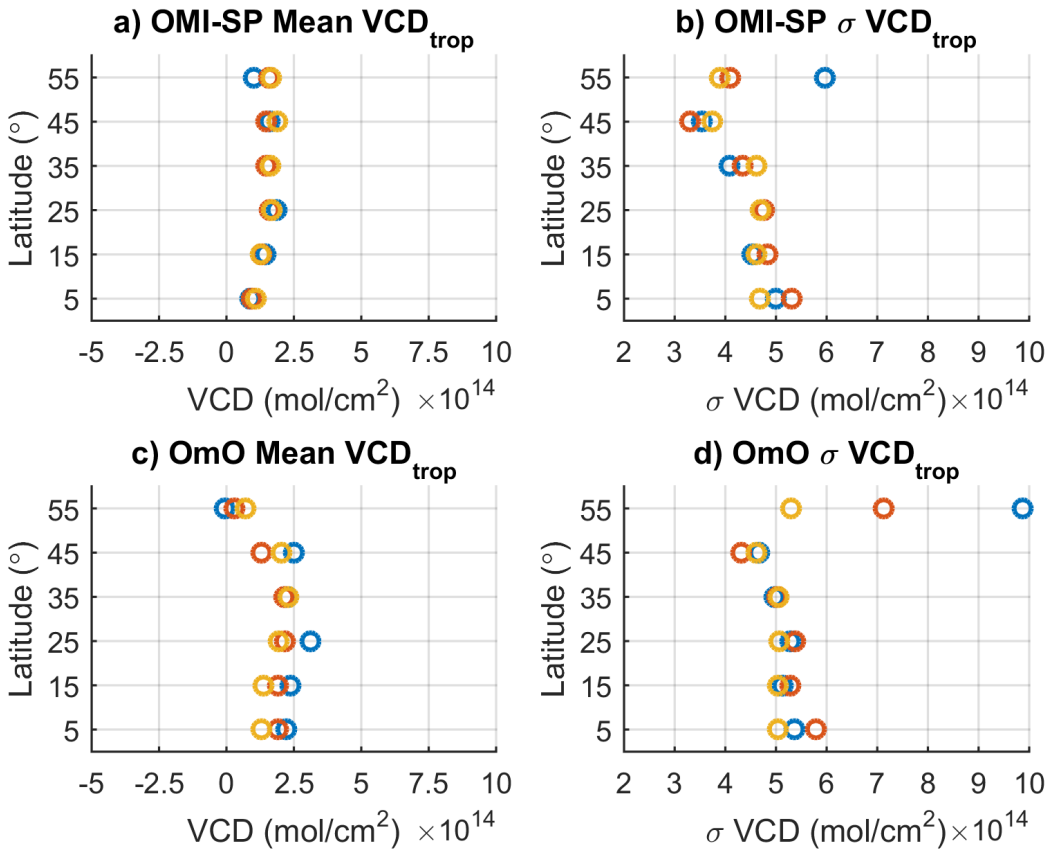

Figure 2.13: The mean and standard deviation of tropospheric VCDs from OMI-SP and OmO retrievals over North America during June (blue circles), July (orange circles), and August (yellow circles) for 2008-2011.

discussed in Chapter 5.

2.1.2 Observations used for Validations

HIPPO Aircraft data

The HIPPO programme consists of five aircraft campaigns flying across the Pacific Ocean. The observations extend from 85°N to 67°S, and from the ground to 14 km in altitude (*Wofsy et al.*, 2011). This study uses data from the HIPPO-2 campaign, which was carried out in November 2009. The flight path of HIPPO-2 is shown in Fig. 2.14. The HIPPO datasets consist of high frequency (up to 1 Hz) observations of multiple trace gases by different instruments onboard the Gulfstream V aircraft. Ten second averages of O_3 observations measured by Vertical Cavity Surface Emitting Lasers (VCSEL, a open path NIR multipass spectrometer) are utilized as an independent dataset to compare to

the a priori and the a posteriori GEOS-Chem simulated O_3 fields in Chapter 3.

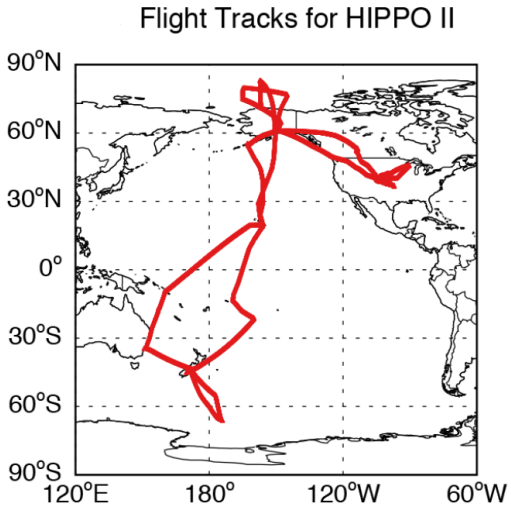

Figure 2.14: Flight path of HIPPO-2 aircraft campaign (*Alvarado et al.*, 2015).

ATom Aircraft data

The NASA Aircraft Tomography Mission (ATom) measures more than 200 air quality and greenhouse gas (GHG) species to study air pollution in the atmosphere. The aircraft DC-8 measures continuous air quality samples over an altitude range of 0.2-12 km altitude. The flight paths of the four deployments are shown in Fig. 2.15. They include coverage over the Pacific Ocean, the Atlantic Ocean, the Polar Caps, and inland North America. I use CO, O_3, and OH measurements made during ATom-1 (August 2016) and ATom-2 (January–February 2017) to evaluate the assimilation. The O_3 measurements were made by the Unmanned Chromatograph for Atmospheric Trace Species (UCATS) and the NOAA Nitrogen Oxides and Ozone (NO_yO_3) chemiluminescence instrument, the CO measurements were made by UCATS and the Harvard Quantum Cascade Laser System (QCLS), and the HO_x species were measured by the Airborne Tropospheric Hydrogen Oxides Sensor (ATHOS). Details of the instrumentation are available on the ATom Mission page (https://espo.nasa.gov/atom/instruments).

ARCTAS Aircraft data

ARCTAS was an aircraft campaign with the goal of better understanding atmospheric composition and climate in the Arctic. ARCTAS-CARB was based around the Central Valley with the focus of monitoring air quality off the coast of California during late June 2008. Immediately following ARCTAS-CARB, the ARCTAS-B summer campaigns were

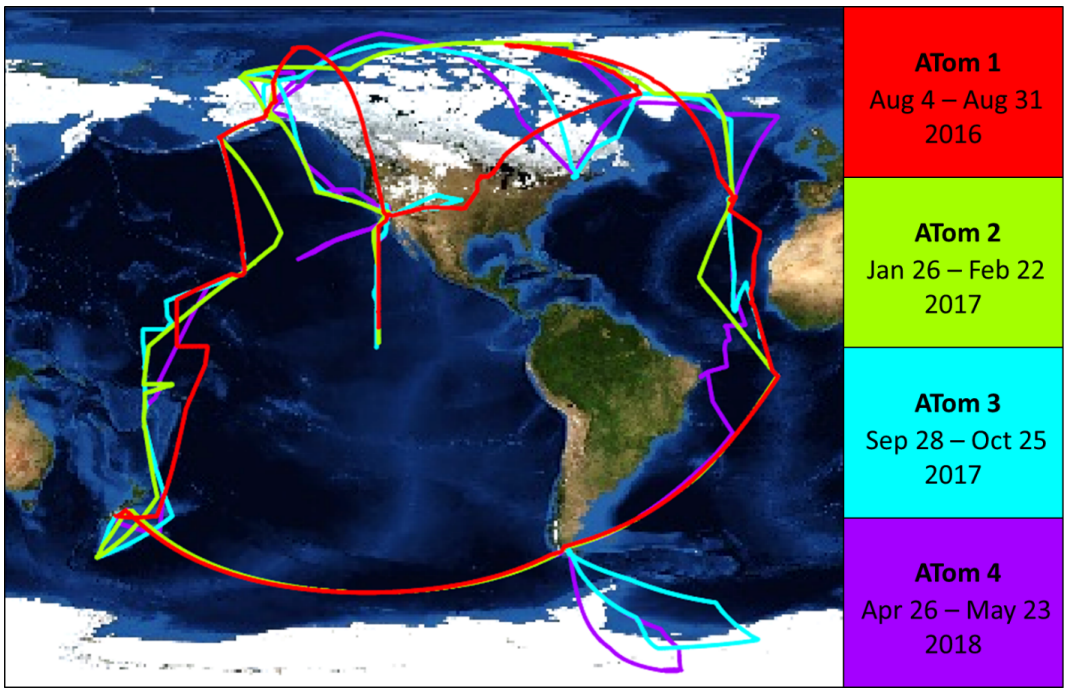

Figure 2.15: Flight path of the four ATom campaigns. Note that only the ATom-1 and ATom-2 campaigns are used in this thesis. (*ATom Science Team*, 2016)

conducted from 26 June to 14 July 2008 and were based around Cold Lake, Alberta, with flights extending across Canada (see Fig. 2.16). The instruments onboard the DC-8 aircraft sampled high-frequency measurements of multiple O_3 precursors (NO_2, HO_x, HCHO, PAN) over the northern US and the Canadian boreal forests. The NO_2 measurements from the ARCTAS campaign have been been used for evaluation in previous modelling studies, such as *Parrington et al.* (2013) and *Bousserez* (2014). In Chapter 5, data from both ARCTAS-CARB (noted as Summer Phase-California) and ARCTAS-B (noted as Summer Phase-Canada) are used to evaluate the modelled NO_2 fields.

TCCON data

The Total Carbon Column Observing Network (TCCON) is a ground-based network of FTIR instruments that provide retrievals of GHGs such as CO_2, CH_4, CO and N_2O. TCCON retrievals of CO columns are based on measurements of transmitted sunlight in the spectral range of 4233-4290 cm^{-1}. The retrieval algorithm uses a nonlinear least squares spectral fitting algorithm (GFIT), which employs an optimal estimation approach. The a priori profile is perturbed to produce a modelled spectrum to best match the measured spectrum (*Wunch et al.*, 2010). 24 out of the 25 sites that are currently in operation and are providing column-averaged dry-air mole fractions (DMF) of CO (XCO) are listed in

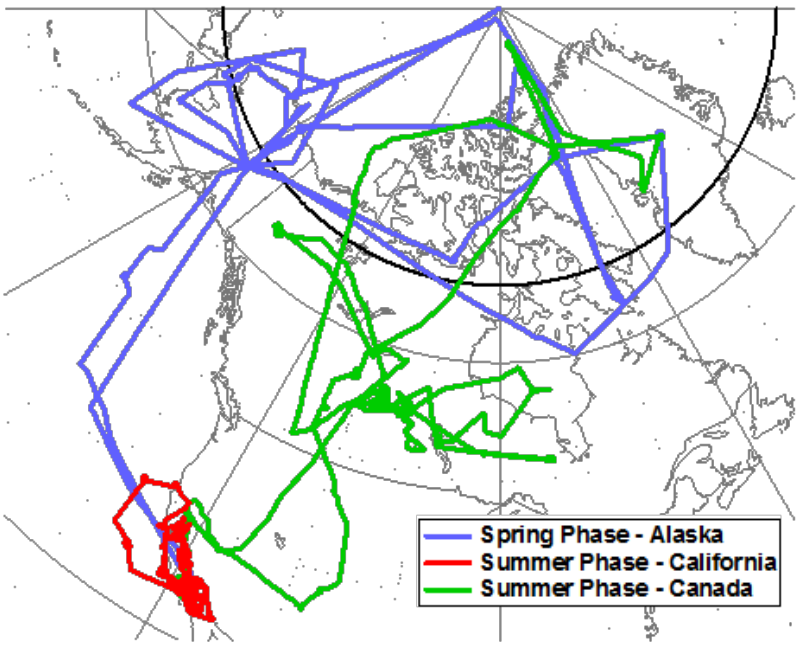

Figure 2.16: Flight path of the ARCTAS campaigns (*Jacob et al.*, 2010). Note that only the Summer Phase-California and Summer Phase-Canada campaigns are used in this thesis.

Table 2.1 (with Izaña, Tenerife excluded). The calibrated TCCON XCO columns have been validated by *Wunch et al.* (2010) using aircraft campaigns flying over many TCCON measurement sites. The XCO measurements are within 7% of aircraft/AirCore observations. The continuous XCO dataset throughout 2016 are used to evaluate modelled CO abundances in Chapter 4.

To compare the GEOS-Chem CO fields with TCCON XCO data, the TCCON column averaging kernels and a priori profiles are used to transform the model as follows:

$$\hat{c} = c_a + h^T a^T (x - x_a), \tag{2.8}$$

where $\hat{c}$ is the smoothed column DMF, c_a is the TCCON a priori DMF, h^T is the observation operator that vertically aggregates the CO concentrations into columns, a^T is the TCCON column average kernel, x is the DMF calculated from the GEOS-Chem model, and x_a is the TCCON a priori profile. It should be noted that x_a is expressed in wet mole fraction x_a^{wet} in GGG2014 version of the TCCON dataset. To apply Eq. 2.8 for any model comparisons, it is necessary to convert it to DMF given the mole fraction of water f_{H_2O} using:

$$x_a = \frac{x_a^{wet}}{1 - f_{H_2O}}, \tag{2.9}$$

Table 2.1: Information of all TCCON sites (*Wunch et al.*, 2010) operational in 2016. Site name, latitude, longitude, and altitude information are given. Classification shows the region where each site is grouped in Fig. 4.7

Site	Latitude	Longitude	Altitude (km)	Classification
Anmyeondo, Korea	36.5N	126.3E	0.03	East Asia
Ascension Island	7.9S	14.3W	0.01	SH Ocean
Bialystok, Poland	53.2N	23.0E	0.18	Northern high latitudes
Bremen, Germany	53.1N	8.9E	0.03	Europe
Caltech, USA	34.1N	118.1W	0.23	North America
Darwin, Australia	12.5N	130.9E	0.04	Oceania
East Trout Lake, Canada	54.4N	105.0W	0.50	Northern high latitudes
Garmisch, Germany	47.5N	11.1E	0.74	Northern high latitudes
Karlsruhe, Germany	49.1N	8.4E	0.12	Europe
Lamont, USA	36.6N	97.5N	0.32	North America
Lauder, New Zealand	45.0S	169.7E	0.37	Oceania
Ny Alesund, Spitsbergen	78.9N	11.9E	0.02	Northern high latitudes
Orleans, France	48.0N	2.1E	0.13	Europe
Paris, France	48.8N	2.4E	0.06	Europe
Park Falls, USA	45.9N	90.3W	0.44	North America
Reunion Island	29.9S	55.5E	0.09	SH Ocean
Rikubetsu, Japan	43.5N	143.8E	0.36	East Asia
Saga, Japan	33.2N	130.2E	0.01	East Asia
Sodankyla, Finland	67.4N	26.6E	0.19	Northern high latitudes
Tsukuba, Japan	36.1N	140.1E	0.03	East Asia
Wollongong, Australia	34.4S	150.9E	0.03	Oceania
Zugspitze, Germany	47.42N	11.0E	2.96	Europe

The EPA AQS network

The US EPA Air Quality System (AQS) (http://www.epa.gov/aqs) provides observations of surface pollutants such as O_3 and NO_2 (*Chai et al.*, 2006). These measurements of air pollutants are collected by the EPA, state, and local and tribal air pollution control agencies from more than 1000 sites across the US. The number of the qualified measurements are nearly twice as many in summer than in winter (*Chai et al.*, 2006). For O_3, modelled outputs are examined against hourly EPA AQS O_3 data. To reduce the representation biases with respect to the model, measurements under extreme O_3 events and severe weather conditions are excluded to avoid anomalous high or low O_3 observations (*Chai et al.*, 2013). As the EPA AQS measurements are mostly distributed over emission source regions, the dataset is used to examine the impact of optimized CO and NO_x emissions on the modelled O_3 distribution in Chapter 3.

Abundances of NO_2 are obtained by calculating the difference between the measurements obtained in the NO_x and NO modes. *Kharol et al.* (2015) has suggested that the NO_2 concentration obtained through this approach may be overestimated due the presence of other reactive nitrogen species such as PAN, HNO_3, HONO and organic nitrates (e.g. alkyl nitrates, RONO). Following *Kharol et al.* (2015), I applied a bias correction factor sf to the reported AQS NO_2 measurements, where

$$\mathrm{sf} = \frac{\mathrm{NO_2}}{\mathrm{NO_2} + \mathrm{RONO} + 0.95 \times \mathrm{PAN} + 0.15 \times \mathrm{HNO_3}}. \tag{2.10}$$

The concentration for each species used in the bias correction was obtained from the GEOS-Chem model. Moreover, the surface NO_2 concentration inferred from CTMs or satellite retrievals are underestimated compared to the in-situ measurements (*Wang and Chen*, 2013). This is because surface NO_2 from the model is usually area averaged into a broader area compared to the point source in-situ measurements. To account for this difference, an additional representativeness correction factor of 1.7 is applied in the AQS validation used in Chapter 5. The factor is based on the interpretation from land regression model used in *Kharol et al.* (2015).

SHADOZ O_3 ozonesonde

The Southern Hemisphere ADditional OZonesondes (SHADOZ) network consists of a number of ozonesonde sites across the tropics, subtropics, and SH. The ozonsondes provide measurements of O_3 from the surface up to about 30 km. I choose the 12 sites shown in Fig. 2.17 (which are all operation sites, with the exception of Hanoi, Vietnam) to evaluate the modelled O_3 concentrations. All ozonesonde profiles are available on the SHADOZ website: https://tropo.gsfc.nasa.gov/shadoz/. *Thompson et al.* (2007) examined the accuracy of the ozonesonde measurements and found that the uncertainty of the observed O_3 is within 5% throughout the troposphere. SHADOZ observations have been used for validation in many O_3 studies (e.g., *Bowman et al.*, 2006; *Miyazaki et al.*, 2012a; *Nassar et al.*, 2008). The data are used in Chapter 4 for evaluation of the modelled O_3 in the tropics and subtropics.

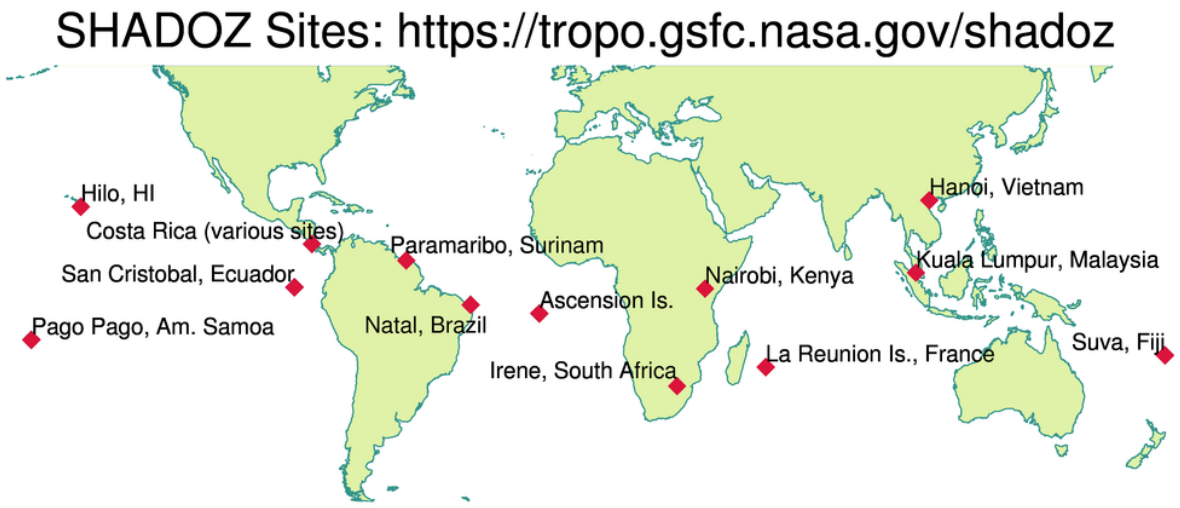

Figure 2.17: Geographical distributions of all SHADOZ ozonesonde cites that are currently in operation (*Witte et al.*, 2017).

2.2 Model

2.2.1 Chemical transport model

A CTM is a 3D numerical model simulating transport, chemistry, emission and deposition processes of chemicals in the atmosphere driven by prescribed meteorological fields (*Brasseur and Jacob*, 2017). CTMs such as GEOS-Chem solve the governing equations, such as the continuity equation, in Eulerian form, and are classified as "box" models. In contrast, CTMs such as FLEXible PARTicle dispersion model (FLEXPART) solve the governing equations in Lagrangian form, and are classifed as "puff" models. (*Jacob*, 1999) Moreover, as mentioned above, CTMs do not solve for the atmospheric dynamics, but instead use the archived meteorological data to simulate atmospheric transport. These meteorological fields are often based on assimilated fields so that the simulated weather patterns will be consistent with observations (*Brasseur and Jacob*, 2017). CTMs are therefore "offline" models, which do not allow the simulated tracers to exhibit any impact on the driving meteorological fields.

Continuity Equation

In Eulerian form, the continuity equation for a given chemical tracer is

$$\frac{\partial n}{\partial t} + \nabla \cdot \mathbf{F} = P - L, \tag{2.11}$$

where n is the number density (molec/cm^3) of the chemical species, $\frac{\partial n}{\partial t}$ represents the rate of change of the concentration of the species, $\nabla \cdot \mathbf{F}$ is the flux divergence which represents the influence of transport, and P and L are local production and loss rate (in units of molec/cm^3/s), which represents the source and sink term, respectively. In the lower atmosphere, where molecular diffusion can be neglected, I can rewrite Eq. 2.11 in the Eulerian advective form expressed in mixing ratio:

$$\frac{\partial \chi n_{air}}{\partial t} = -\nabla \cdot (\chi n_{air} \mathbf{U}) + P - L, \tag{2.12}$$

and then:

$$\frac{\partial \chi}{\partial t} = -\mathbf{U} \cdot \nabla \chi + P' - L', \tag{2.13}$$

where n_{air} is the number density of air, χ is the mixing ratio of the chemical species, $P' = \frac{P}{n_{air}}$ and $L' = \frac{L}{n_{air}}$ are the production and loss rate expressed in mixing ratio form. In this way, the continuity equation becomes a first order partial differential equation (PDE) in space and time. In the CTMs, the system is further discretized in space, which converts the PDE into a system of coupled ordinary differential equations (ODEs) in dimension of $d \times k$, where d is the total number of grid boxes in the model, and k is the total chemical species transported in the model (*Brasseur and Jacob*, 2017). To further simplify the continuity equations, operator splitting is applied to separate the transport and chemistry terms into two independent processes. The transport term is assumed to only interact between neighbouring grid boxes but not among species, whereas for the chemical term ($P' - L'$) there can be interaction between the species but not between neighbouring grid boxes (*Brasseur and Jacob*, 2017). In this way, Eq. 2.11 can be split into:

$$\left[\frac{\partial n}{\partial t}\right]_{trans} = -\nabla \cdot (n\mathbf{U}), \tag{2.14}$$

$$\left[\frac{\partial n}{\partial t}\right]_{chem} = P - L. \tag{2.15}$$

By solving Eqs. 2.14 and 2.16, given the total number of the involved species, their chemical reaction rates, emissions, the production and loss rate, and the meteorological fields, the concentration of modelled tracers can be propagated from time t_0 to time $t_0 + \delta t$ in each model time step δt with:

$$n(t_0 + \delta t) = M(n(t_0)) = \mathbf{C} \cdot \mathbf{T}(n(t_0)), \tag{2.16}$$

where M is the model, as represented by the chemistry operator, $\mathbf{C}$, and the transport operator, $\mathbf{T}$ (*Brasseur and Jacob*, 2017).

2.2.2 GEOS-Chem model

GEOS-Chem (www.geos-chem.org) is a global CTM driven by the meteorological reanalysis data from the Goddard Earth Observing System (GEOS) of the NASA Global Modeling and Assimilation Office (GMAO). In Chapter 3 I use version v34 of the GEOS-Chem adjoint model, which is based on v8-02-01 of the forward GEOS-Chem model. In Chapters 4 and 5, I use version v35l of the adjoint model, which is based on v8-02-01 of the GEOS-Chem model with relevant updates up to version 9 of the forward model. The adjoint model can be driven by either GEOS-5 or GEOS-Forward Processing (GEOS-FP) meteorological fields. The differences between GEOS-5 and GEOS-FP are in the vertical resolutions and physical parameterizations of the parent GCM at GMAO. GEOS-5 is also an older version that was available only until 2013. GEOS-FP is the current version that is available for simulations after 2013. As an offline CTM, the modelled chemical fields in GEOS-Chem will not impact the GEOS meteorological fields (*Brasseur and Jacob*, 2017).

The native resolution of the GEOS-5 and GEOS-FP fields is $0.5^\circ \times 0.667^\circ$ and $0.25^\circ \times 0.3125^\circ$ horizontally, respectively, with 72 vertical pressure levels from the surface to 0.01 hPa. For the global inversions in Chapter 3 and 4, in order to reduce the computational costs of the simulations, the GEOS-5 and GEOS-FP products are reshaped into a horizontal resolution of $4^\circ \times 5^\circ$ in longitude $\times$ latitude with 47 vertical pressure levels from surface to 0.01 hPa. The meteorological fields are updated every 3 or 6 hours depending on the variables. At a horizontal resolution of $4^\circ \times 5^\circ$, the time step in the model for advection, convection, emission updates and chemistry are 30, 30, 60, and 60 minutes, respectively. In the full chemistry simulation, there are 86 species and 307 chemical reactions representing the O_3-NO_2-hydrocarbon-aerosol chemistry. However, only 43 of the individual tracers are transported. The numerical scheme of the transport operator is based on that of *Lin and Rood* (1996). The chemical operator can use either the Gear or Rosenbrock chemical solver. For GEOS-Chem, the adjoint model does not explicitly simulate stratospheric chemistry. Instead, it uses the linearization of stratospheric ozone (LINOZ) chemistry by *McLinden et al.* (2000). Wet and dry deposition of trace gases is parameterized as described in *Liu et al.* (2001) and *Wang et al.* (1998a).

2.2.3 GEOS-Chem nested model

The model can be run at the native horizontal resolution, in a nested mode, over selected domains such as over North America, Europe, and Asia. For the NO_x assimilation in Chapter 5, I use the North American nested domain (140°W-40°W, 10°N-70°N) with version v35l of the adjoint model at a resolution of $0.5° \times 0.667°$ resolution. The regional model requires boundary conditions around the nested domain during the simulation period, which are generated from the $4° \times 5°$ global version of the model. The time step in North American nested-grid model for advection, convection, emission updates and chemistry are 10, 10, 20, and 20 minutes, respectively.

2.3 Emission inventories

For the global analyses in Chapters 3 and 4, anthropogenic emissions are taken from EDGAR 3.2FT2000 (*Olivier and Berdowski*, 2001). However, for certain regions, the global inventory is replaced with a number of regional inventories. In Canada, the model uses the Criteria Air Contaminants from the National Pollutant Release Inventory (www.ec.gc.ca/inrp-npri/). The US inventory is based on the US EPA National Emission Inventory for 2005 (NEI05). The Mexican inventory uses the Big Bend Regional Aerosol and Visibility Observational Study Emissions Inventory (BRAVO) (*Kuhn and Etyemezian*, 2003). The European inventory is from the Cooperative Programme for Monitoring and Evaluation of Long-Range Transmission of Air Pollutants in Europe (EMEP). In Asia, the model uses the emission estimates from *Streets et al.* (2006). Biomass burning emissions are taken from GFED-3 inventory (*van der Werf et al.*, 2006). Biogenic emissions are from version 2.0 of the Model of Emissions of Gases and Aerosols from Nature (MEGAN 2.0) (*Guenther et al.*, 2006). Soil emissions of NO_x are according to *Wang et al.* (1998b), and the parameterization of LNO_x follows *Price and Rind* (1992), with a global scaling, as described by *Murray et al.* (2012), to match climatological observations from the Optical Transient Detector and the Lightning Imaging Sensor (OTD/LIS).

For the regional NO_x studies in Chapter 5, the US anthropogenic emissions are from the 2008 National Emission Inventory (NEI08) of the US EPA. The Canadian and Mexican anthropogenic emissions, as well as the biomass burning inventory, are the same as in the global model. Biogenic emissions are based on MEGAN v2.1. LNO_x and soil NO_x emissions are as in the global model.

2.4 Inversion approaches

2.4.1 Bayesian inversions

Inverse modelling, in general, is a Bayesian inference problem, in which I seek to optimize selected modelled states given a set of observations. The variables estimated by the numerical model are called state variables (represented by the state vector $\mathbf{x}$). Mathematically, for a perfect model without bias, the forward model can be written as (*Keller*, 2014):

$$\mathbf{x}_{n+1} = M(\mathbf{p}, \mathbf{x}_n) \tag{2.17}$$

where M represents the numerical model, acting on model parameter (such as emissions) $\mathbf{p}$ and modelled states (such as concentrations) $\mathbf{x}_n$ at time n. M evolves the modelled state $\mathbf{x}$ from time n to time $n+1$. With the consideration of Eq. 2.17, the observation can be written as:

$$\mathbf{y} = H(\mathbf{x})) + \epsilon, \tag{2.18}$$

where $\mathbf{y}$ represent the observation vector, H is the observation operator that projects the modelled state into the observation space, and ϵ is the observation error, whose mean is zero and covariance matrix is given by $\mathbf{R}$.

Given a set of observations and prior knowledge of certain modelled states, Bayes' theorem describes the probability distribution of the a posterior states as:

$$P(\mathbf{x}|\mathbf{y}) = \frac{P(\mathbf{y}|\mathbf{x})P(\mathbf{x})}{P(\mathbf{y})}, \tag{2.19}$$

where $P(\mathbf{x})$ is the a prior probability density function (pdf) of the state, $P(\mathbf{y}|\mathbf{x})$ is the conditional pdf of $\mathbf{y}$ given the a priori knowledge of $\mathbf{x}$, and $P(\mathbf{x}|\mathbf{y})$ represents the a posteriori pdf based on the knowledge from the observations. In this thesis, I assume that the probability distributions in Eq. 2.19 are Gaussian. Hence, the conditional pdf of the observations can be expressed as:

$$P(\mathbf{y}|\mathbf{x}) \propto exp[-\frac{1}{2}(\mathbf{y} - H(\mathbf{x}))^T\mathbf{R}^{-1}(\mathbf{y} - H(\mathbf{x}))], \tag{2.20}$$

and the a priori pdf as:

$$P(\mathbf{x}) \propto exp[-\frac{1}{2}(\mathbf{x} - \mathbf{x_a})^T\mathbf{B}^{-1}(\mathbf{x} - \mathbf{x_a})]. \tag{2.21}$$

Hence, the a posteriori pdf is expressed as:

$$P(\mathbf{x}|\mathbf{y}) \propto exp[-\frac{1}{2}(\mathbf{x}-\mathbf{x_a})^T\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x_a}) - \frac{1}{2}(\mathbf{y}-H(\mathbf{x}))^T\mathbf{R}^{-1}(\mathbf{y}-H(\mathbf{x}))], \quad (2.22)$$

where $\mathbf{x_a}$ is the a priori state vector, $\mathbf{R}$ is the observation error covariance matrix, and $\mathbf{B}$ is the a priori/background error covariance matrix (*Rodgers*, 2000). I can also express Eq. 2.22 as a cost function:

$$J = \frac{1}{2}(\mathbf{x}-\mathbf{x_a})^T\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x_a}) + \frac{1}{2}(\mathbf{y}-H(\mathbf{x}))^T\mathbf{R}^{-1}(\mathbf{y}-H(\mathbf{x})), \quad (2.23)$$

The maximum a posteriori (MAP) state is retrieved through either maximizing the a posteriori pdf $P(\mathbf{x}|\mathbf{y})$ in Eq. 2.22 or by minimizing the cost function J in Eq. 2.23.

2.4.2 Optimal Estimation

Optimal estimation is a Bayesian inversion that is widely used in the remote sounding community. The retrieval of trace gas concentrations (such as CO) or meteorological fields (such as temperature) is performed by minimizing the difference between the spectrum from a RTM and the radiance observed by Earth observing satellites, subject to a priori constraints on the atmospheric quantity of interest. By assuming ϵ in Eq. 2.18 to be normally distributed and linearizing the observation operator $\mathbf{H}$, the solution obtained from minimizing the cost function in Eq. 2.23 is

$$\hat{\mathbf{x}} = \mathbf{x_a} + \mathbf{A}(\mathbf{x}-\mathbf{x_a}), \quad (2.24)$$

where

$$\mathbf{A} = \mathbf{BH^T}(\mathbf{HBH^T}+\mathbf{R})^{-1}\mathbf{H}. \quad (2.25)$$

Equation 2.24 is the same as Eqs. 2.1, 2.2, and 2.3, where $\mathbf{x_a}$ is the a priori state vector, $\mathbf{x}$ is the true state, $\mathbf{A}$ is the averaging kernel of the retrievals, and $\hat{\mathbf{x}}$ is the retrieved state vector.

2.4.3 4D-Var

The 4D-Var data assimilation scheme can be used to assimilate the 3D fields of the gases of interest in a CTM, which has two types of applications in my project. First, inverse modelling is applied when optimizing the emissions. In this case, surface emissions are the optimized parameters. The measured atmospheric concentrations are the observa-

tions. The forward model is the CTM that transforms the emissions into the atmospheric abundances. The a priori emissions are based on bottom-up inventories. The a posteriori emission then refers to the optimized top-down emission estimates. Second, inverse modelling is applied when optimizing initial conditions. In this case, the modelled concentrations, which evolve in time according to the forward model, are compared with the observations at various locations and times to optimize the initial model distribution so that the model better reproduces the observations when the initial conditions are projected forward in time, over the period for which the observations were assimilated. In the case where the satellite retrievals are not in the same space as the model, such as when the data are retrievals of column abundances, an observation operator must be used to map the state variables to the observation space. The mathematical expression of the 4D-Var cost function is

$$\begin{aligned} J(\mathbf{p}, \mathbf{x_0}) = & \frac{1}{2}[\mathbf{p} - \mathbf{p_a}]^T \mathbf{B}_p^{-1} [\mathbf{p} - \mathbf{p_a}] \\ & + \frac{1}{2}[\mathbf{x_0} - \mathbf{x_a}]^T \mathbf{B}_x^{-1} [\mathbf{x_0} - \mathbf{x_a}] \\ & + \frac{1}{2} \sum_{n=1}^{N} \gamma_r [\mathbf{y}_n - H(M(\mathbf{p}, \mathbf{x_0}))]^T \mathbf{R}^{-1} [\mathbf{y}_n - H(M(\mathbf{p}, \mathbf{x_0}))], \end{aligned} \tag{2.26}$$

where $\mathbf{p}$ are the model parameters, such as the CO emissions, $\mathbf{x_0}$ is the initial condition, such as the O_3 initial distribution, $\mathbf{p_a}$ and $\mathbf{x_a}$ are the a priori estimates of the parameters, $\mathbf{B}$ is the a priori error covariance matrix, $\mathbf{x}_n = M(\mathbf{p}, \mathbf{x_0})$ is the modelled state (e.g. CO concentration) at time steps n, $\mathbf{y}_n$ are observations (e.g. MOPITT CO profiles), and γ_r is the normalization factor used for balancing the contribution of the different chemical species to the cost function.

The minimization of the cost function J uses the method of Lagrange multipliers, which is an algorithm to find the local minima of a function subject to certain constraints. In this system, I would like to maximize $J(\mathbf{p}, \mathbf{x_0})$ subject to $\mathbf{x}_n = M(\mathbf{p}, \mathbf{x_0})$ (which can be re-expressed $\mathbf{x}_n - M(\mathbf{p}, \mathbf{x_0}) = 0$). Combining Eq. 2.17, Eq. 2.18 with Eq. 2.26, the Lagrangian function $\mathcal{L}(\mathbf{p}, \mathbf{x_0}, \lambda_0) = J(\mathbf{p}, \mathbf{x_0}) - \lambda(\mathbf{x}_n - M(\mathbf{p}, \mathbf{x_0}))$ is expanded as:

$$\begin{aligned} \mathcal{L}(\mathbf{p}, \mathbf{x_0}, \lambda_0) = \frac{1}{2} \Bigg(& (\mathbf{p} - \mathbf{p_a})^T {\mathbf{B_p}}^{-1} (\mathbf{p} - \mathbf{p_a}) + \sum_{n=1}^{N} (\mathbf{y}_n - H(\mathbf{x}_n))^T \mathbf{R}^{-1} (\mathbf{y}_n - H(\mathbf{x}_n)) \\ & - 2 \sum_{n=1}^{N} \lambda_n^T [\mathbf{x}_n - M(\mathbf{p}, \mathbf{x_0})] \Bigg). \end{aligned} \tag{2.27}$$

To minimize the cost function, $\mathcal{L}$ must be minimized to obtain the improved estimates. The derivative of $\mathcal{L}$ with respect to each variables after dropping the common multiplicative factor $\frac{1}{2}$ is shown below (*Keller*, 2014):

$$\frac{\partial \mathcal{L}}{\partial \mathbf{p}} = 2\left({\mathbf{S_p}}^{-1}(\mathbf{p} - \mathbf{p_b}) + \sum_{n=1}^{N} \lambda_n^T \frac{\partial M(\mathbf{p}, \mathbf{x_0})}{\partial \mathbf{p}} \right) \tag{2.28}$$

$$\frac{\partial \mathcal{L}}{\partial \mathbf{x}_n} = 2\left(-\mathbf{H}^T\mathbf{R}^{-1}[\mathbf{y}_n - H(\mathbf{x}_n)] - \lambda_n + \left(\frac{\partial M(\mathbf{p}, \mathbf{x_0})}{\partial \mathbf{x}} \right)^T \lambda_{n+1} \right) \tag{2.29}$$

$$\frac{\partial \mathcal{L}}{\partial \mathbf{x}_N} = 2(-\mathbf{H}^T\mathbf{R}^{-1}[\mathbf{y}_N - H(\mathbf{x}_N)] - \lambda_N) \tag{2.30}$$

$$\frac{\partial \mathcal{L}}{\partial \lambda_n} = 2(\mathbf{x}_n - M(\mathbf{p}, \mathbf{x}_{n-1})) \tag{2.31}$$

When the system is minimized, $\frac{\partial \mathcal{L}}{\partial \mathbf{p}}$, $\frac{\partial \mathcal{L}}{\partial \mathbf{x}_n}$, $\frac{\partial \mathcal{L}}{\partial \mathbf{x}_N}$, and $\frac{\partial \mathcal{L}}{\partial \lambda_n}$ are all set to be zero. $\frac{\partial \mathcal{L}}{\partial \lambda_n}$=0 is self-explanatory since it represents the GEOS-Chem forward model which simulates the tracers of interest from $\mathbf{x}_0$ to $\mathbf{x}_n$ based on the a priori estimates $\mathbf{p}$. $\frac{\partial \mathcal{L}}{\partial \mathbf{x}_n}$=0 and $\frac{\partial \mathcal{L}}{\partial \mathbf{x}_N}$=0 produce the adjoint model equations:

$$\lambda_n = \left(\frac{\partial M}{\partial \mathbf{x}_n} \right)^T \lambda_{n+1} - \mathbf{H^T R^{-1}}(\mathbf{y}_n - H(\mathbf{x}_n)), \tag{2.32}$$

and

$$\lambda_N = -\mathbf{H^T R^{-1}}(\mathbf{y}_N - H(\mathbf{x}_N)), \tag{2.33}$$

where $(\frac{\partial M}{\partial \mathbf{x}})^T$ is the adjoint of tangent linear model $\mathcal{M} = \frac{\partial \mathbf{M}}{\partial \mathbf{x}}$. The adjoint model starts calculate the adjoint forcing λ_N at the end of time window expressed in Eq. 2.33. After that, the adjoint forcing in the previous time steps λ_n are computed using Eq. 2.32 until the total cost function J and the state $\mathbf{p}$ and/or $\mathbf{x_0}$ are optimized using Eq. 2.28. In the end, I repeat the forward calculation with the new estimate of $\mathbf{p}$ and/or $\mathbf{x_0}$ and iterate until the convergence criterion is satisfied. (*Keller*, 2014)

In summary, the forward model with bottom-up emission inventories as well as initial concentrations for all chemical tracers are simulated. First, the initial conditions are integrated forward in time. The concentrations of various species will be compared with the employed satellite observations at the locations and times of the observations to compute the cost function and adjoint forcing. After that, the total cost function is minimized by reducing the gradient $\frac{dJ}{dx}$ using the GEOS-Chem adjoint model (*Henze et al.*, 2007). (x represents the modelled states which depends on M, $\mathbf{x_0}$, and $\mathbf{p}$.) Reducing the gradient $\frac{dJ}{dx}$ requires adjusting the optimized quantities (emissions or initial conditions)

at the beginning of the assimilation window. Then, I repeat the forward calculation with the new estimate of $\mathbf{p}$ and/or $\mathbf{x_0}$ and iterate until the convergence criterion of the cost function J is satisfied. In the GEOS-chem adjoint model, the Limited memory Broyden-Fletcher-Goldfarb-Shanno algorithm for Bound-constrained optimization (L-BFGS-B) is used to optimize the cost function (*Byrd et al.*, 1995).

2.4.4 Weak-constraint 4D-Var

For the O_3 assimilation in Chapter 5, I use the weak constraint 4D-var data assimilation system of GEOS-Chem. Unlike strong constraint 4D-var, which assumes that the model is perfect, weak constraint introduces a model error term (*Trémolet*, 2006) as an additional optimization factor to correct for biases in the modelled transport and chemistry. Implementing the model forcing in the total cost function for weak constraint 4D-var is key to mitigating the model biases. In this way, both the O_3 initial conditions and the model forcing terms are optimized throughout the assimilation window. The mathematical expressions of the weak constraint 4D-var system are as follows:

$$\mathbf{x_n} = M(\mathbf{x_{n-1}}, \mathbf{p}) + \eta_n, \tag{2.34}$$

$$\begin{aligned} J(\mathbf{x_0}, \mathbf{p}, \eta) = &[\mathbf{p} - \mathbf{p_a}]^T \mathbf{B}_p^{-1} [\mathbf{p} - \mathbf{p_a}] \\ &+ [\mathbf{x_0} - \mathbf{x_a}]^T \mathbf{B}_x^{-1} [\mathbf{x_0} - \mathbf{x_a}] \\ &+ \sum_{n=1}^{N} \gamma_r [\mathbf{y}_n - H(M(\mathbf{x_0}, \mathbf{p}))]^T \mathbf{R}^{-1} [\mathbf{y}_n - H(M(\mathbf{x_0}, \mathbf{p}))] \\ &+ \sum_{n=1}^{N-1} \eta_n^T \mathbf{Q}_\eta^{-1} \eta_n. \end{aligned} \tag{2.35}$$

Equation 2.34 represents the implementation of the model error for the forward model simulations. Here η_n is added to the model to account for errors in evolving the model state from time step $n-1$ to n. Equation 2.35 shows the cost function for weak constraint 4D-Var, which is similar to Eq. 2.26. It is noted that this expression drops the common factor $\frac{1}{2}$ and has the addition of the model error cost function term, where $\mathbf{Q}_\eta$ is the model error covariance matrix. The selection of $\mathbf{Q}_\eta$ is further discussed in Chapter 5.

2.5 Observation operators

Recall that $\mathbf{H}$ is the observation operator that projects the model state to the observation space. For MOPITT and TES, the observation operators are equivalent to Eqs. 2.1 and

2.2, respectively, and are given by

$$H_{\mathrm{MOP}}(\mathbf{x}_n) = \mathbf{z_a}^{\mathrm{MOP}} + \mathbf{A}^{\mathrm{MOP}}(\mathbf{x}_n - \mathbf{z_a}^{\mathrm{MOP}}), \tag{2.36}$$

and

$$H_{\mathrm{TES}}(\mathbf{x}_n) = \mathbf{z_a}^{\mathrm{TES}} + \mathbf{A}^{\mathrm{TES}}(\mathbf{x}_n - \mathbf{z_a}^{\mathrm{TES}}), \tag{2.37}$$

where $\mathbf{x}_n$ is the GEOS-Chem modelled output. Because of the high vertical resolution of the OSIRIS data, the observation operator for OSIRIS consists only of interpolation of the GEOS-Chem O_3 profile onto the OSIRIS vertical grid. The observation operator for OMI and OmO NO_2 involves a two step approach. First, to calculate the modelled tropospheric NO_2 column abundance, I integrate the modelled NO_2 profile to obtain vertical column densities. Second, to avoid the mismatch of the air mass factor between the model and the observation, I compare the modelled NO_2 SCDs with retrieved NO_2 observations, using the equation

$$H_{\mathrm{OMI/OmO}}(\mathbf{x}_n) = \mathrm{AMF} \cdot h(\mathbf{x}_n) \tag{2.38}$$

where h is the column operator that linearly aggregates the GEOS-Chem NO_2 partial columns into VCDs and AMF is the air mass factor computed from the GEOS-Chem model, where

$$\mathrm{AMF} = \frac{\int \psi(p) m(p) dp}{\int \psi(p) dp}. \tag{2.39}$$

By definition, $\psi(p)$ is the GEOS-Chem NO_2 mixing ratio depending on the model pressure p, and $m(p)$ is the scattering weights which described the sensitivity of the backscattered sunlight to the absorber. For IASI O_3, the observation operator is equivalent to Eq. 2.3.

$$H_{\mathrm{IASI}}(\mathbf{x}_n) = \mathbf{z_a}^{\mathrm{IASI}} + \mathbf{A}^{\mathrm{IASI}}(\mathbf{C_n x_n} - \mathbf{z_a}^{\mathrm{IASI}}) \tag{2.40}$$

where $\mathbf{z}_a$ is the single IASI a priori O_3 partial columns, $\mathbf{x}_n$ is the modelled O_3 concentrations, $\mathbf{C}_n$ the partial column operator that converts the modelled O_3 concentrations into partial columns, according to the information of pressure and altitude from GEOS-Chem, and $\mathbf{A}$ is the IASI averaging kernel. For OMI HCHO, I apply the same observation operator as OMI NO_2 (for which I assimilate tropospheric SCDs), given by $H_{\mathrm{OMI}}(\mathbf{x}_n)$=AMF$\cdot h(\mathbf{x}_n)$. The modelled HCHO is then transformed into SCDs using the model derived AMF. This observation operator is also applied and explained in *Kaiser et al.* (2018). For MLS HNO_3, since the averaging kernel of a retrieved level peaks sharply on the corresponding pressure levels (*Massart et al.*, 2012), the observation operator di-

rectly interpolates the modelled stratospheric HNO_3 profile onto the MLS observation grid using the same approach as for OSIRIS O_3.

Chapter 3

Quantifying CO and NO_x emissions in November 2009 and July 2010 using observations from MOPITT, OMI, TES, and OSIRIS

3.1 Introduction

Surface pollutants such as CO and NO_x have a significant impact on air quality as well as the evolution of the chemistry of the troposphere (*Zhang et al.*, 2008a). As discussed in Chapter 1, top-down estimates of regional emissions of CO and NO_x are uncertain. These uncertainties in the regional emission estimates reflect the influence of model errors, differences in the employed data assimilation approaches, and the observations assimilated in different studies. For instance, the uncertainties of tropospheric abundance of OH and its major precursor O_3 could profoundly impact the CO emission estimates. Multi-species chemical data assimilation introduced in Section 1.6 has the potential to reduce the modelled biases on CO-HO_x-NO_x-O_3 tropospheric chemistry by providing complementary information from multiple satellite observations. *Miyazaki et al.* (2012a) were the first to assimilate a suite of global space-based observations of atmospheric trace gases to constrain the chemical mechanism in their CTM in the context of their inversion analysis to quantify global NO_x emissions and modelled states of CO, O_3 and OH. In this chapter, I use the GEOS-Chem 4D-Var data assimilation system to conduct a MSA to quantify NO_x and CO emissions, following the approach of *Miyazaki et al.* (2012a). I assimilate O_3 data from TES and OSIRIS, NO_2 column abundances measured by OMI,

and retrieved CO profiles from MOPITT.

A key difference between the work here and that of *Miyazaki et al.* (2012a) is my use of the 4D-Var assimilation scheme. With the LETKF approach, *Miyazaki et al.* (2012a) had to employ localization to minimize the impact of long-range correlations, which could limit the chemical coupling among species. Furthermore, they used a short assimilation window of 90 minutes, which limits the model's transport impact on optimizing surface emissions of CO and NO_x. In contrast, I use a long assimilation window of two weeks. This was chosen to ensure that sufficient observations are ingested to constrain the model states. In the 4D-Var scheme, the state vector is optimized to produce a smoothed model trajectory, which would allow the model forecast to best reproduce the assimilated observations over the assimilation period (*Trémolet*, 2006). As a result of this time dependence, the 4D-Var scheme inherently produces a consistent chemical state across the assimilation period (*Eibern and Schmidt*, 1999). Due to expensive computational costs of running MSA, I restrict the study periods to November 2009 and July 2010. These durations are chosen because the availability of all the datasets are maximized during the corresponding period of the months. The November 2009 run is used to examine the quality of the MSA results, as this assimilation time period overlaps with the HIPPO-2 aircraft campaign. The experiments in July 2010 are used to examine the seasonal differences for MSA. The 2-week assimilation window is selected since optimizing O_3 initial conditions in the 4D-var scheme can only provide O_3 constraints within its lifetime in troposphere. I also compare the 2-week window with other assimilation window selections in Section 3.3.7. The initial conditions of all the advected tracers were obtained by running the GEOS-Chem forward model from July 2005. I do not include a "spin-up" assimilation period before each analysis period, which could introduce biases from initial conditions. But these experiments are still legitimate since I would like to test whether MSA could adjust the overall chemical state and reduce the initial condition bias simultaneously.

As this study focuses on global emission estimates of CO and NO_x, their surface emissions including anthropogenic, biomass burning, biofuel and soil (for NO_x only) sources are optimized. O_3 initial condition is also optimized when assimilating the two O_3 observations. This is motivated by the results from *Jones et al.* (2009) and *Bowman et al.* (2006). They showed a strong impact of changing NO_x surface emissions on O_3 and CO concentrations, which in turn suggested that changing O_3 would affect surface CO and NO_x emissions. Between the two O_3 observations, TES O_3 provides highest sensitivity around mid-troposphere, while OSIRIS O_3 provides the most information in UTLS. As explained in Chapter 1, combining both observations is conducive to providing

Table 3.1: Optimized quantities with the corresponding observations used in this study.

Optimized quantities	Assimilated observation(s)
CO emissions	MOPITT CO profiles
O_3 initial conditions	TES or OSIRIS O_3 profiles
Surface NO_x emissions	OMI NO_2 columns
Surface CO and NO_x emissions, O_3 initial conditions, LNO_x emissions	All instruments (including MOPITT, TES, OSIRIS, OMI)

complementary information for O_3 throughout the troposphere. LNO_x is also optimized, as it constitutes an important NO_x source in the free troposphere, especially in the tropics, and can impact free tropospheric O_3 (*Bowman et al.*, 2006).

The optimization options are shown in Table 3.1. CO emission, surface NO_x emissions and O_3 initial conditions are optimized when MOPITT CO retrievals, OMI tropospheric NO_2 column or TES/OSIRIS O_3 retrievals are assimilated respectively. In MSA (also referred to as the "all instrument run"), I optimize surface emissions of CO and NO_x, LNOx and O_3 initial conditions with all observations assimilated.

3.2 Model setup

3.2.1 Assimilation approach

In the 4D-var data assimilation system of GEOS-Chem, the cost function is given by:

$$\begin{aligned} J(\mathbf{x_0}, \mathbf{p}) = &[\mathbf{p} - \mathbf{p}_a]^T \mathbf{B}_p^{-1} [\mathbf{p} - \mathbf{p}_a] + [\mathbf{x}_0 - \mathbf{x}_a]^T \mathbf{B}_x^{-1} [\mathbf{x}_0 - \mathbf{x}_a] \\ &+ \sum_{r=1}^{4} \sum_{n=1}^{N} \gamma_r [\mathbf{y}_{n,r} - H_r(M(\mathbf{x_0}, \mathbf{p}))]^T \mathbf{R}_r^{-1} [\mathbf{y}_{n,r} - H_r(M(\mathbf{x_0}, \mathbf{p}))], \end{aligned} \tag{3.1}$$

where $\mathbf{y}_{n,r}$ is the observation vector that contains the four different observation types (r) (MOPITT CO, TES O_3, OSIRIS O_3, and OMI NO_2), at a given time n over the assimilation period, $\mathbf{p}$ represents the model parameters, which refers to surface CO and NO_x emissions in our case, and $\mathbf{x_0}$ is the initial condition state vector which optimizes the O_3 initial conditions. In this chapter, CO production due to oxidation of NMVOC is not constrained. In minimizing J, the model is compared to the observations at the observation time n, and the GEOS-Chem adjoint is used to project the model-data mismatch backward onto the initial conditions and model parameters (*Henze et al.*, 2007). $\mathbf{p_a}$ and $\mathbf{x_a}$ are the a priori estimates of the model parameters and the initial conditions, $\mathbf{B}_p$ and $\mathbf{B}_x$ are the a priori error covariance matrices for the model parameters and the

initial conditions, $\mathbf{R}_r$ is the observation error covariance for observations of species r, and $M(\mathbf{x_0}, \mathbf{p})$ represents the GEOS-Chem forward model that simulates the CO, O_3, and NO_2 concentrations. γ_r is a normalization factor used for balancing the contribution of the various observations to the cost function, which will be discussed in Section 3.2.3. H is the observation operator that projects the model state to the observation space. The detailed setup of each employed observation operator has been introduced in Section 2.5.

In the assimilation I set the a priori error covariances as diagonal matrices with relative error standard deviations of 50%. It should be noted that characterizing the covariances is challenging, particularly for the 4D-Var scheme. As a result, in the absence of reliable knowledge of the covariance structure, a diagonal covariance matrix is a reasonable assumption. My assumption of a uniform a priori error of 50% is an ad-hoc simplification, but still consistent with previous studies (e.g., *Jiang et al.*, 2013, 2015a, 2017; *Stavrakou et al.*, 2008). In addition, using a diagonal background error covariance matrix could speed up the minimization of the cost function, which has been a standard approach for most of the 4D-Var based chemical data assimilations (*Skachko et al.*, 2016). Although I impose the diagonal covariance matrices, the use of the adjoint to propagate information across the assimilation window allows observations of a particular species to influence other species in the model. The assumption of a spatially uniform a priori error enables me to better identify the impact of the multi-species observation and their chemical coupling on the regional emission estimates without the confounding influence of some regions being more tightly constrained to match the a priori emissions than others. My use of a uniform 50% error is similar to the approach of *Miyazaki et al.* (2012a). They assumed an initial a priori error of 40% for the emission estimates and then inflated the emission errors after each analysis step to ensure a minimum emission error of 30%. A notable difference between my assimilation configuration and *Miyazaki et al.* (2012a) is that they assumed a 10% error for the tracer concentrations, whereas my 50% a priori error is assumed for the emission and concentration estimates. I believe that a larger a priori error for the concentrations is appropriate in our configuration because of the much longer assimilation window used here compared to that in *Miyazaki et al.* (2012a) (two weeks compared to 90 minutes).

For the observation error covariances, I implement a super-observation approach for MOPITT CO and OMI NO_2 data. Following *Jiang et al.* (2013), I aggregate the MOPITT retrievals onto the model's $4^\circ \times 5^\circ$ grid and assume a uniform observation error of 20% to capture the influence of representativeness errors. For OMI, which has the greatest observational coverage of the datasets assimilated, I generate super-observations following *Miyazaki et al.* (2012b), with the exception that I use our own representative-

ness errors according to the cost function scaling discussed in Section 3.2.3. In general, the super-observation error for NO_2 is 25% of the mean measurement error.

3.2.2 Cost Function Weighting Observing System Simulation Experiment (OSSE)

One major consideration in my MSA is the choice of the cost function normalization factor γ, which weights observations of different species to balance their information contribution in the MSA framework. If each observation is considered equally ($\gamma = 1$ for all observations), the model will favor observations with high accuracy and the observations of the chemical species with maximum adjoint sensitivity. My state vector consists of CO and NO_x emissions and the initial O_3 abundances. Changes in the CO emissions in the state vector will have a limited impact on the modelled NO_2 mismatch with the OMI observations. Changes in the NO_x emissions will have a larger, but still limited impact on the modelled CO mismatch with the MOPITT observations. However, on the long timescales considered in the assimilation, changes in the CO and NO_x emissions will influence the model mismatch with the O_3 data. So the challenge in the MSA is to ensure that the MOPITT and OMI data provide a sufficient contribution to the cost function minimization to constrain the CO and NO_x emissions, respectively.

Table 3.2: Configurations of the the OSSEs for γ. Shown are the imposed bias in the CO and NO_x emissions and in the O_3 initial conditions. Also listed are the pseudo datasets assimilated in each OSSE.

OSSE	CO Bias	NO_x Bias	O_3 Bias	Pseudo Data
MOPITT	−50%	0%	0%	MOPITT
OMI	0%	+50%	0%	OMI
MSA1	−50%	+50%	+30%	MOPITT, TES, OSIRIS, OMI
MSA2	−50%	+50%	+30%	MOPITT, TES, OSIRIS, OMI
MSA3	−50%	+50%	+30%	OMI-like CO, O_3, NO_2

To better understand the need for balancing the cost function, I have conducted an OSSE using pseudo-observations during the first week of November 2009. Pseudo data for MOPITT, TES and OSIRIS, and OMI, were generated by sampling the GEOS-Chem a priori CO, O_3, and NO_2 fields, respectively, at the observation locations and times for each instrument. In generating the TES and MOPITT pseudo data I applied the real averaging kernels and a priori profiles. For OMI, I produced pseudo NO_2 column densities using the reported scattering weights from the real OMI NO_2 dataset. I also produced NO_2, O_3, and CO columns densities from an OMI-like instrument, for which I used the

OMI observation locations and times to sample the GEOS-Chem CO, O_3, and NO_2 fields and generated column densities by vertically integrating the modelled profiles (without averaging kernels or scattering weights). I then assimilated these data in the five OSSE experiments described in Table 3.2, assuming biased CO and NO_x emissions and biased O_3 initial conditions. For the MOPITT OSSE, I reduced CO emissions by 50%, whereas for the OMI OSSE I increased NO_x emissions by 50%. For both OSSEs, the model was run with perturbed emissions from early September to the beginning of November, followed by the assimilation of the pseudo data during the first week of November. For the MSA OSSEs, in addition to perturbing the CO and NO_x emissions, I also increased the O_3 initial conditions by 30% throughout the troposphere. Since I conducted the OSSEs for only one week in November 2009, I increased the precision of the pseudo data by assuming a measurement error that was 50% of the errors reported with the real data. I did not apply additional noise to the pseudo data since random errors would average out given the large amount of data assimilated, and because my focus was mainly on the impact of the weighting of the different terms in the cost function on the inferred sources.

Figure 3.1 shows the ratio of estimated to true emissions for Northern Africa, East Asia, Europe, North America, Australia, and the globe. As expected, the OMI-only and MOPITT-only OSSEs recovered the global emissions, but there were regional differences. The OMI-only inversion successfully recovered the regional NO_x emissions, but the MOPITT-only inversion did not have sufficient information with only one week's worth of MOPITT pseudo data to recover the true CO emission estimates for Europe and North America to within 10%. For the MSA1 OSSE, when no scaling was applied to the cost function (γ=1), the global mean CO and NO_x emission estimates reverted closer to the prior. The regional CO and NO_x emission estimates were more biased than for the single instrument case for most regions. This is because the OMI and MOPITT pseudo data contributed less toward the cost function minimization as the inversion primarily focused on optimizing the O_3 state. For the MSA2, I scaled the MOPITT and OMI terms with $\gamma_{\mathrm{OMI}} = 16$ and $\gamma_{\mathrm{MOP}} = 64$, respectively. The motivation for this particular scaling will be discussed below in Sec. 3.2.3. As seen in Fig. 3.1, with the increased weighting of the MOPITT and OMI pseudo data in MSA2, the assimilation was able to recover the global mean CO and NO_x emissions. Furthermore, the regional NO_x emission estimates were consistent with the estimates from the OMI-only assimilation. For the CO emissions, the estimates for North America and Australia were closer to the truth, whereas the North African and East Asian estimates were biased high. The European estimate was the same as the result for MSA1. In MSA3, there was no scaling applied

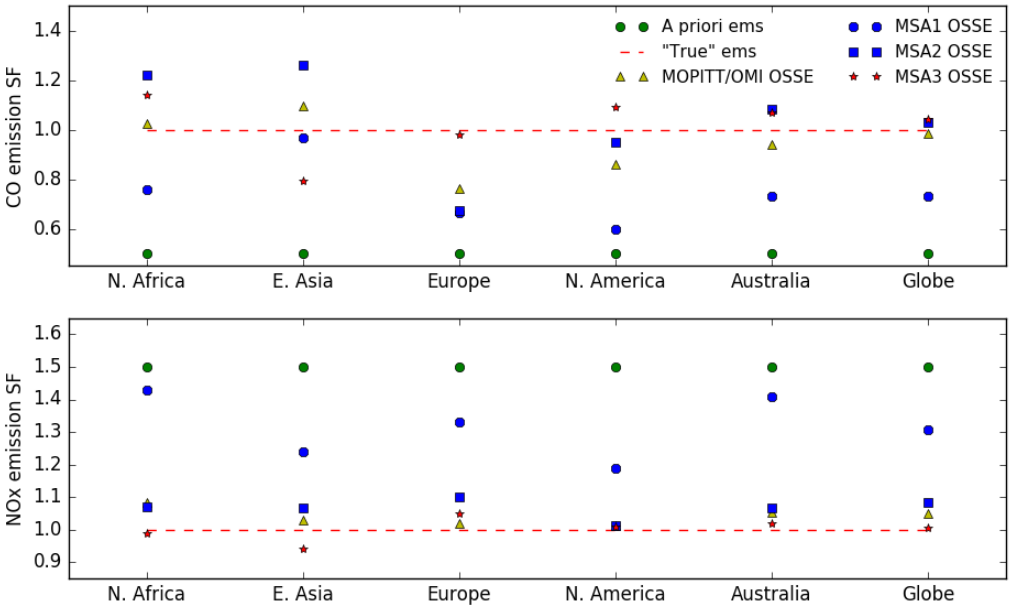

Figure 3.1: Emission SF for different experiments in major emission sources when comparing with the unperturbed "true" state. MSA1 and MSA2 OSSE assimilate the pseudo-observations of CO, NO_2 and O_3 sampled from MOPITT, OMI and TES & OSIRIS observations respectively. γ=1 was applied in MSA1, Scaling following Case G in Section 3.2.3 was applied in MSA2. In MSA3, the pseudo-observations of CO, NO_2 and O_3 are all sampled from OMI observations as tropospheric columns with γ=1.

to the cost function. However, unlike MSA1, the assimilation was able to recover the global mean CO and NO_x emissions. MSA3 also recovered the regional NO_x emissions. For CO, the assimilation recovered the emission estimates to within 10% for all regions except for Northern Africa and East Asia. As a result of the OMI-like observational coverage, MSA3 also produced the best estimate of European emissions. Similarly, I believe that because the OMI-sampled O_3 observations produced much greater observational coverage than TES, the assimilation was better able to adjust the O_3 state using the O_3 data, and consequently, better able to utilize the CO and NO_2 to optimize the CO and NO_x emissions, respectively. Overall, the choice of γ will depend on the type of data assimilated and how well they constrain the different quantities within the state vector. It should be noted that although the OMI-sampled tropospheric O_3 column data worked well in the OSSE, I believe that real tropospheric O_3 column data from OMI would be challenging to use in my MSA because of the potential impact of errors associated with the removal of the stratospheric contribution to the total column.

Table 3.3: Relative a posteriori contribution of the various observations to the total observation term in the cost function using different normalization factors γ_{MOP} and γ_{OMI}. Note that γ_{TES}=1 and γ_{OSIRIS}=1 for all experiments.

Case #	γ_{OMI}	γ_{MOP}	Cost function share (%)				Total cost function ($\times 10^7$)
			MOPITT CO	TES O_3	OMI NO_2	OSIRIS O_3	
A	1	1	1.04	59	0.12	39	5.4
B	4	1	1.04	59	0.27	40	5.5
C	16	1	1.02	57	1.3	40	5.9
D	400	1	0.72	43	18	38	11
E	16	4	3.7	67	1.3	28	5.5
F	16	16	22	48	1.2	29	7.2
G	16	64	37	43	1.1	18	9.0
H	16	400	78	16	0.42	5.7	17

3.2.3 Estimating the Cost Function Weighting

To determine the appropriate values for γ for the MOPITT and OMI data, I conducted a series of sensitivity tests in which I sequentially increased γ_{MOP} and γ_{OMI} from one to 64 by a factor of four. I also considered a more extreme case with $\gamma_{\mathrm{MOP}} = 400$ and $\gamma_{\mathrm{OMI}} = 400$. Table 3.3 shows the MOPITT and OMI data contributions to the observation term in the a posteriori cost function for eight of the sensitivity tests conducted. When no scaling is applied, O_3 observations from TES and OSIRIS provide the dominant contributions to the total observation term in the cost function (60% and 39%, respectively). The total contribution to the observation term from MOPITT CO is about 1% and that from OMI NO_2 is 0.12%. The dominating cost function contribution by O_3 observations would lead to all O_3 precursor fields such as modelled NO_2 being over adjusted to drive the assimilated O_3 changes. Consequently, the assimilated NO_x emissions will not strongly reflect the influence of the OMI NO_2 observations, which is consistent with the results of the OSSE discussed above. As expected, as γ_{MOP} and γ_{OMI} become larger, the contributions from MOPITT CO and OMI NO_2 increase, and the contributions from the two sets of O_3 observations decrease. When $\gamma_{\mathrm{OMI}} = 400$, the cost function share of OMI NO_2 reaches 18%. Similarly, increasing γ_{MOP} to 16 (with $\gamma_{\mathrm{OMI}} = 16$), the MOPITT CO a posteriori cost function contribution increases to 22%.

Figure 3.2 shows the a posteriori global CO and NO_x emission estimates for the different values of γ_{MOP} and γ_{OMI} in the MSA. With $\gamma_{\mathrm{MOP}} = 1$ or $\gamma_{\mathrm{OMI}} = 1$, the estimated surface CO or NO_x emission changes in the MSA are much smaller than that inferred using only MOPITT CO or OMI NO_2, respectively. Instead, with $\gamma_{\mathrm{MOP}} = 64$ and $\gamma_{\mathrm{OMI}} = 16$, we recovered the global CO emissions to within 10% of the estimate

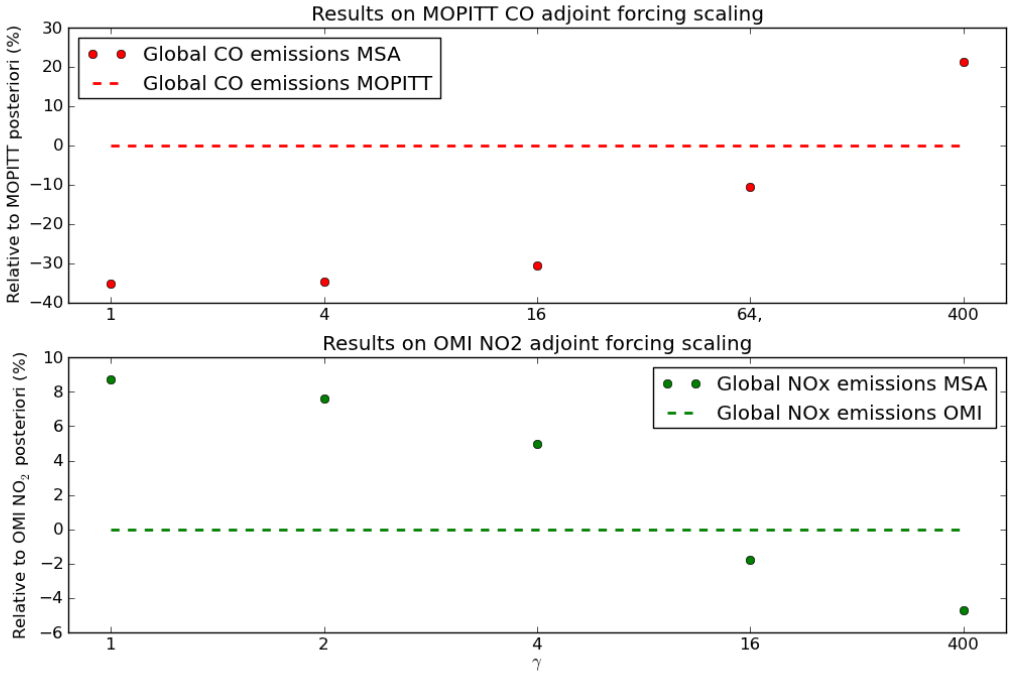

Figure 3.2: Changes (in percent) in the global a posteriori emissions of CO and NO_2 as a function of the normalization factors γ shown in Table 3.3. The changes are shown relative to the global emission estimates obtained from assimilating only MOPITT CO or only OMI NO_2 data to separately infer the CO or NO_x emissions, respectively.

obtained from the MOPITT-only assimilation and the global NO_x emissions to within 2% of the estimate inferred in the OMI-only assimilation. These are the cost function weightings applied when data from all the instruments are assimilated together (referred to as the all instrument run) for November 2009 and July 2010. I believe that the $\gamma_{MOP} = 64$ is acceptable because global mean OH has an uncertainty of 10-15% (e.g., *Krol et al.*, 1998; *Spivakovsky et al.*, 2000; *Wang et al.*, 1998b; *Naik et al.*, 2013), which leads to a similar uncertainty on the global mean CO emissions. The resulting global CO emissions are also consistent with both MOPITT CO assimilation and many other inverse modelling studies summarized in *Hooghiemstra et al.* (2012). Moreover, my choice of $\gamma_{OMI} = 16$ and $\gamma_{MOP} = 64$ ensures that no single dataset will dominate the optimization throughout the assimilation. The scaling applied here can be considered as an observation error reduction for MOPITT and OMI, which is plausible since super-observations are considered for the two observations. The importance of scaling each observation type was also discussed in *Müller and Stavrakou* (2005). In their joint assimilation of ground-based CO observations with GOME NO_2 data, they apply a different scaling on the CO

and NO_2 observations in order to produce a posteriori results that best match the CO and NO_2 in-situ measurements.

3.3 Results

3.3.1 Assimilated O_3 Distribution

The O_3 distribution plays a key role in the HO_x-NO_x-O_3-VOC chemistry and as discussed above, has a strong influence on the cost function minimization, therefore, I begin with evaluating the assimilated O_3 distribution. The a priori O_3 distribution at 500 hPa and the changes in O_3 that result from assimilating TES or OSIRIS O_3 observations are shown in Fig. 3.3. The a priori O_3 concentrations show a distinct minimum within the subtropics and a maximum between 20°-40° in both hemispheres. The changes in O_3 in the assimilation using only TES data have a strong zonal structure. Over the tropics and subtropics there are negative corrections exceeding 10 ppbv, whereas over the northern high latitudes there are positive corrections also exceeding 10 ppbv. In contrast, in the OSIRIS assimilation, the changes are negligible, with a maximum decrease in O_3 of 3 ppbv over northern Africa. This is because OSIRIS does not make observations at these altitudes. The changes in the OSIRIS assimilation at these altitudes are caused by downward transport of the changes in O_3 in the UTLS.

The a posteriori O_3 changes at 300 hPa (shown in Fig. 3.4) for the experiment assimilating TES O_3 data are similar in structure to the O_3 changes at 500 hPa. There are negative changes over the tropics and subtropics, and positive changes over the northern high latitudes. The averaged magnitude of changes in the UTLS varies between -15 to 8 ppbv. As expected, the O_3 changes at 300 hPa when using OSIRIS O_3 are larger than the changes obtained in the lower troposphere (LT), with negative corrections everywhere except for polar regions. The averaged magnitude of the changes is about -10 ppbv.

For the experiment combining both TES and OSIRIS instruments, the assimilated results are dominated by OSIRIS O_3 for both the 500 hPa and 300 hPa levels. Within the UTLS, OSIRIS provides more O_3 measurements with smaller observation error than TES. The optimization using 4D-var is most effective to respond to observations with much less observation error (OSIRIS in this case). In the LT, it is suprising that OSIRIS assimilation still has a large impact on O_3 changes. When using TES only, any differences between the modelled states and the observations will lead to changes where TES has the highest vertical sensitivity. Since the averaging kernel of TES O_3 peaks in the LT and the middle troposphere (MT) (*Bowman et al.*, 2006), the O_3 changes at 500 hPa

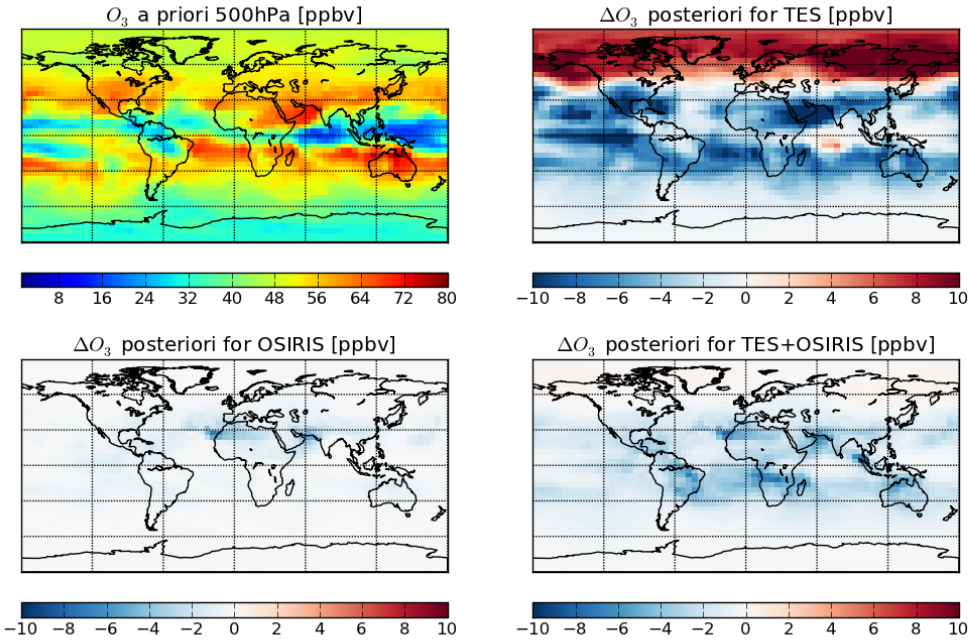

Figure 3.3: The a priori O_3 concentrations and the difference between the a posteriori and the a priori O_3 concentrations at 500 hPa for November 2009.

would be most dominant. After OSIRIS data is added in the assimilation, the O_3 adjoint sensitivity will be most pronounced in the UTLS. Since the O_3 bias is largest with a much longer lifetime in the UTLS compared to other levels, it mitigates the O_3 biases for other layers.

For the experiment combining all instruments, the assimilated O_3 distribution (not shown) in both the middle troposphere and UTLS is similar to that obtained in the TES+OSIRIS assimilation. However, at the surface, the all instrument assimilation produces increases of up to 5 ppbv in O_3 over the main anthropogenic emission regions in the northern hemisphere (see Fig. 3.5), whereas the changes in surface O_3 in the TES+OSIRIS assimilation are negligible (not shown). In the TES-only assimilation there are large changes in high latitude O_3, which I believe are due to the downward transport of the changes in O_3 in the middle troposphere (see Fig. 3.3). In the tropics and southern subtropics, the all instrument assimilation produce modest reductions in surface O_3. The increase in O_3 in the northern extratropics as well as the decrease in the tropics in the all instrument case are due to the reduced NO_x emissions in the assimilation. These results show that optimizing the O_3 initial conditions is a less effective means of adjusting surface O_3 concentrations than optimizing the O_3 precursor emissions. This is because the O_3

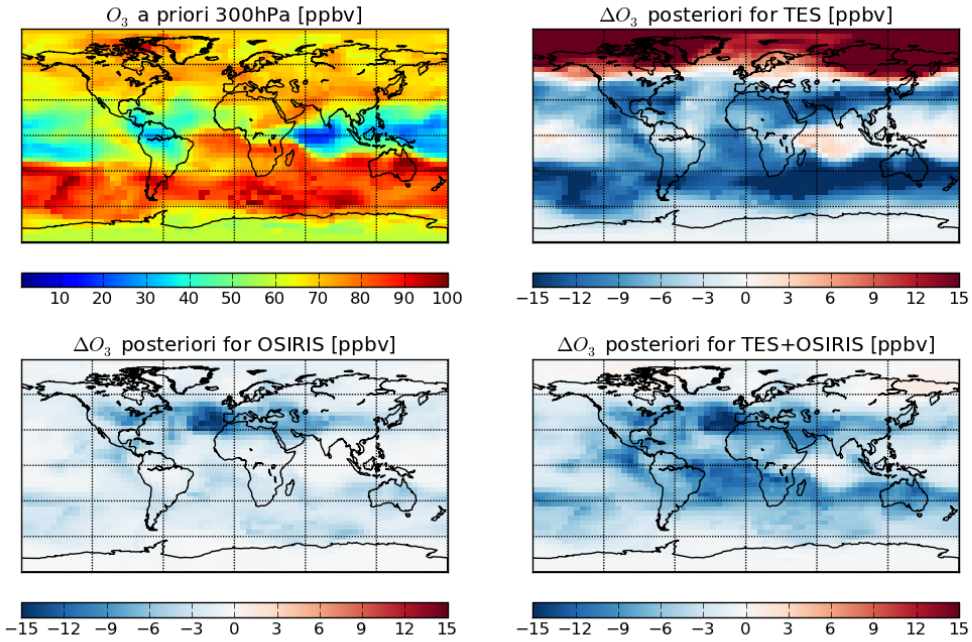

Figure 3.4: The a priori O_3 concentrations and the difference between the a posteriori and the a priori O_3 concentrations at 300 hPa for November 2009.

lifetime is short near the surface, so the perturbations in the initial conditions are quickly destroyed. The lifetime is expected to be longer at high latitude in November and, thus, I see the persistence at the high latitude surface as a signal of the mid-troposphere O_3 changes in the TES assimilation in Fig. 3.5.

Evaluation against HIPPO O_3 data

The assimilated O_3 in November 2009 is compared with HIPPO data to evaluate the O_3 distribution both in the LT and MT as well as in the UTLS. To map the modelled state to the observations, I sampled the assimilated O_3 at the time and locations closest to the HIPPO observations. The differences between the observations and the a priori and a posteriori simulations were then averaged into three latitude bins: 90°N-45°N, 45°N-45°S, and 45°S-90°S. As shown in Table 3.4, within the tropics and subtropics, the observed O_3 concentrations are lower than the a priori O_3 concentrations for both LT (by -4.4 ppbv) and UTLS (by -15.5 ppbv). When TES O_3 is added to the assimilation, the a posteriori O_3 concentrations are closer to the observed values over this region. However, the positive biases in the model are not removed completely. The experiment using TES O_3 also provides a posteriori O_3 concentrations that are closer to the HIPPO values over

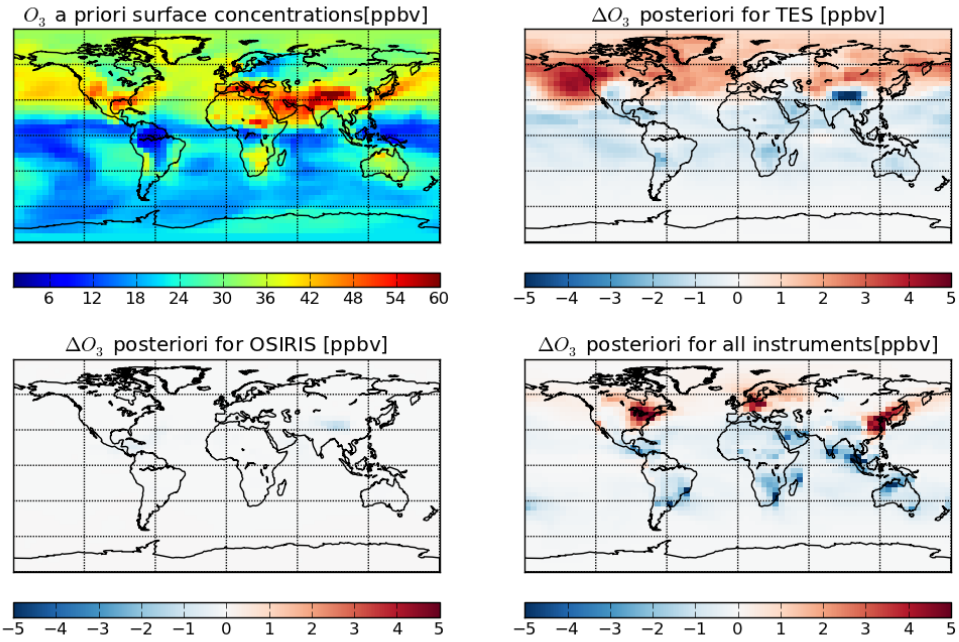

Figure 3.5: The a priori O_3 concentrations and the difference between the a posteriori and the a priori O_3 concentrations at the surface for November 2009.

the southern hemisphere. However, the assimilated O_3 using TES degrades the model agreement with the observations in the northern high latitudes. This could be due to the influence of the TES averaging kernels smearing O_3 from the stratosphere down into the troposphere. In late fall and early winter, the tropopause is lower and O_3 abundances are greater in the high latitude lower stratosphere, which exacerbates the smearing effects of the TES averaging kernels. The experiment combining TES and OSIRIS observations compensates for the adverse impact of the TES assimilation in the northern high latitudes and also significantly reduces the a priori model bias in the southern high latitudes. However, at low latitudes in both hemispheres, the TES-only assimilation provides the best agreement with the HIPPO data. The all instrument assimilation produces an O_3 distribution that is similar to that of the TES+OSIRIS assimilation, but with slightly larger absolute biases. Overall, despite the adverse impact of the TES assimilation in the northern high latitudes, both the TES+OSIRIS and the all instrument assimilation experiments effectively reduced the a priori bias.

Table 3.4: Comparisons (in units of ppbv) between the HIPPO O_3 observations and the a priori and a posteriori O_3 fields in November 2009. Shown are the mean difference (HIPPO - model) and the mean absolute difference |(HIPPO - model)|, in parentheses. Lower and Middle Troposphere (LT & MT) refers to 0-7 km in altitude, whereas Upper Troposphere and Lower Stratosphere (UTLS) refers to 7-13 km.

Region	A priori	TES O_3	OSIRIS O_3	TES and OSIRIS	All instruments
Global mean	-7.3 (36)	-3.3 (24)	-1.7 (23)	-1.4 (23)	-1.9 (24)
90°N-45°N LT & MT	-1.9 (11)	-5.8 (15)	-1.0 (7.3)	-1.6 (7.6)	-1.2 (7.2)
90°N-45°N UTLS	26 (52)	34 (50)	27 (41)	27 (40)	28 (42)
45°N-45°S LT & MT	-4.4 (19)	-1.2 (9.0)	-3.7 (9.5)	-3.3 (9.2)	-3.4 (9.9)
45°N-45°S UTLS	-16 (37)	0.9 (25)	-4.7 (29)	-1.7 (27)	-0.8 (25)
45°S-90°S LT & MT	-4.3 (17)	-3.8 (6.8)	-1.0 (7.3)	-1.6 (7.3)	-2.1 (7.1)
45°S-90°S UTLS	-36 (50)	-33 (47)	29 (55)	12 (53)	14 (51)

Evaluation against EPA AQS O_3 data

The assimilated surface O_3 in July 2010 is evaluated using EPA AQS O_3 data. I focus on O_3 in July because models tend to overestimate summertime surface O_3 abundances in the eastern US (*Reidmiller et al.*, 2009). Fig. 3.6 shows the differences between the observations and the a priori model, the TES-OSIRIS assimilation, the OMI NO_2 assimilation and the all instrument assimilation. The AQS data were binned onto the model grid using the weighted-average algorithm suggested by *Klonecki et al.* (2012). As expected the a priori model overestimates surface ozone across much for the US. *Canty et al.* (2015) suggested that these biases on modelled surface ozone could be due to discrepancies in vertical mixing as well as errors in the O_3 precursors. In the TES-OSIRIS assimilation in which I adjusted only the O_3 initial conditions there is no significant improvement on surface O_3, which is expected since the O_3 lifetime is short in the boundary layer. The all instrument run, however, shows significant surface O_3 improvements. In the eastern and western US (east and west of 95°W) the mean ozone bias decreased from 13.1 to 6.2 ppbv and from 5.4 to 0.3 ppbv, respectively. In fact, the OMI NO_2 assimilation alone would correct surface O_3 with similar patterns as the O_3 changes obtained by assimilation of all of the instruments. The surface O_3 biases were reduced to 7.9 and 1.9 ppbv for the eastern and western US, respectively, driven by the chemical response to the adjusted NO_x emissions. This confirms the utility of the MSA approach for providing additional corrections to the O_3 distribution through the emission optimization.

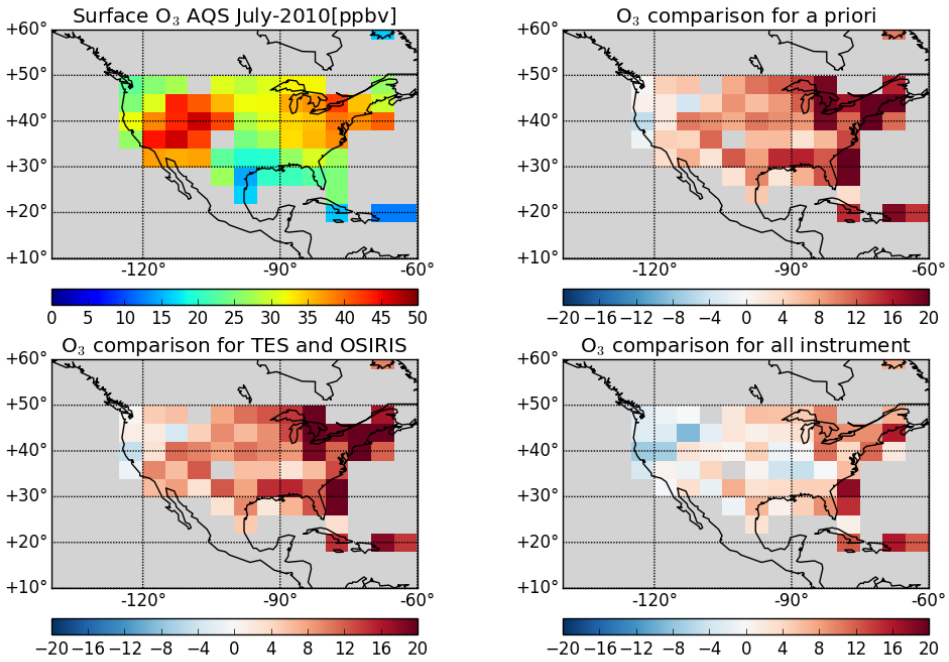

Figure 3.6: Surface O_3 distribution for AQS data interpolated on the GEOS-Chem model grid, as well as the differences between AQS and the modelled O_3 during July 2010.

3.3.2 Tropospheric OH concentrations

Figure 3.7 shows the mass weighted global mean OH concentrations for the a priori and the changes in OH from MOPITT assimilation as well as MSA. The global mean OH produced by the a priori is 11.9×10^5 molec/cm^3 for November 2009, and 13.9×10^5 molec/cm^3 for July 2010. The distribution of tropospheric mean OH is consistent the tropospheric water vapour abundances. In November, the highest OH concentrations are found in the tropics and southern subtropics due to higher actinic fluxes as well as higher water vapour concentrations. In July 2010, the OH maximum shifts into the northern hemisphere. For the assimilated results in comparison with the a priori, the global mean OH constrained by MSA is reduced to 10.5×10^5 molec/cm^3 (by 11.7%) and 12.8×10^5 molec/cm^3 (by 7.9%) in November and July respectively. In contrast, with the MOPITT-only assimilation I obtained reductions in the global mean OH of 0.9% and 1.8% for November and July, respectively. My results show that the MSA is conducive to constraining the tropical and subtropical OH, which could reduce the uncertainties of CO emission estimates due to biased tropospheric OH (*Müller et al.*, 2018). There are also significant OH changes due to CO and NO_x emission adjustments over the extratropics.

In November 2009, the a posterior OH is reduced by 6.5% between 30^{o}N and 60^{o}N. In July 2010, as the tropospheric chemical cycling of CO-HO_x-NO_x-O_3-OH becomes more active over the northern extratropics, the resulting OH is reduced by 10.5%. These OH changes inferred by MSA suggest that tropospheric OH could be constrained indirectly, relying on the assimilation of the OH-related species in the 4D-var assimilation system. I note, however, that improving the OH abundance in the assimilation should also account for the influence of VOCs on OH, which was not done in my analysis.

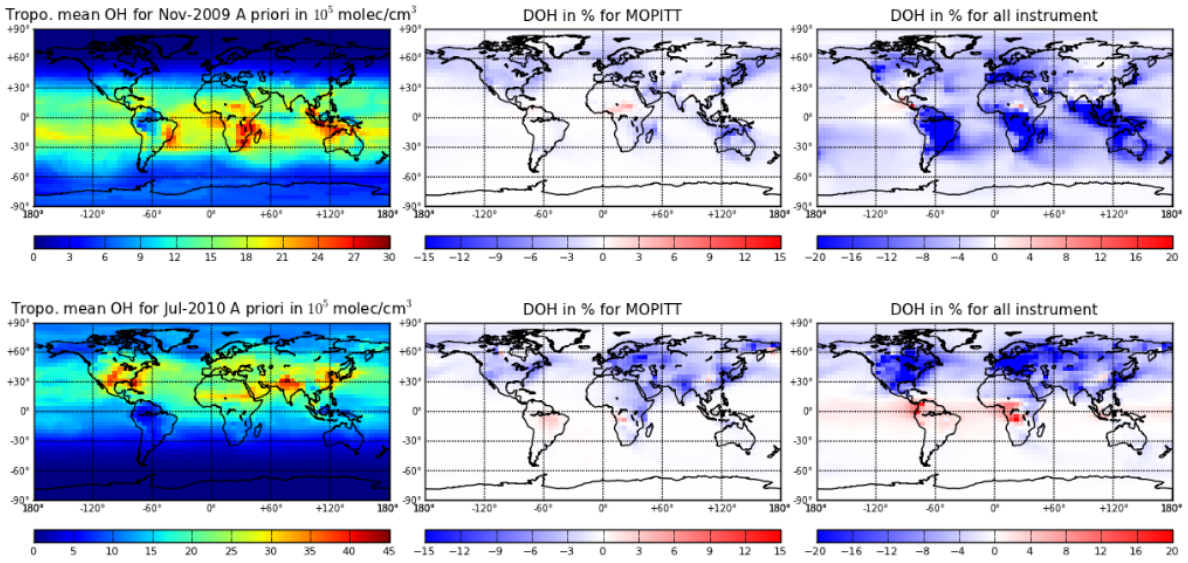

Figure 3.7: The a priori mass weighted tropospheric mean OH distribution (in 10^5 molec/cm^3) and the relative difference ($\frac{\mathrm{OH_{post}}-\mathrm{OH_{ap}}}{\mathrm{OH_{ap}}}\times 100\%$) between the a posteriori and the a priori OH for different experiments. The mass weighted tropospheric mean OH was calculated using Equation 1.16, following *Spivakovsky et al.* (2000).

3.3.3 Regional CO and NO_x Emissions in November 2009

The difference of anthropogenic and biomass burning CO emissions between the a priori and the a posteriori using MOPITT-only and all other experiments are shown in Fig. 3.8. The anthropogenic emissions optimized by MOPITT observations show an increased emissions (more than 5×10^{11}molec/cm^2/s) over East Asia, Europe, the eastern and western United States, and over Northeast Asia. The biomass burning emissions assimilating MOPITT CO shows that the emissions over Northern Equatorial Africa are decreased by more than 4×10^{11}molec/cm^2/s, while emissions over southern Africa, northern Australia, and northwestern North America are increased around 3.5×10^{11}molec/cm^2/s. When assimilating all chemical species, the optimized emissions over both anthropogenic

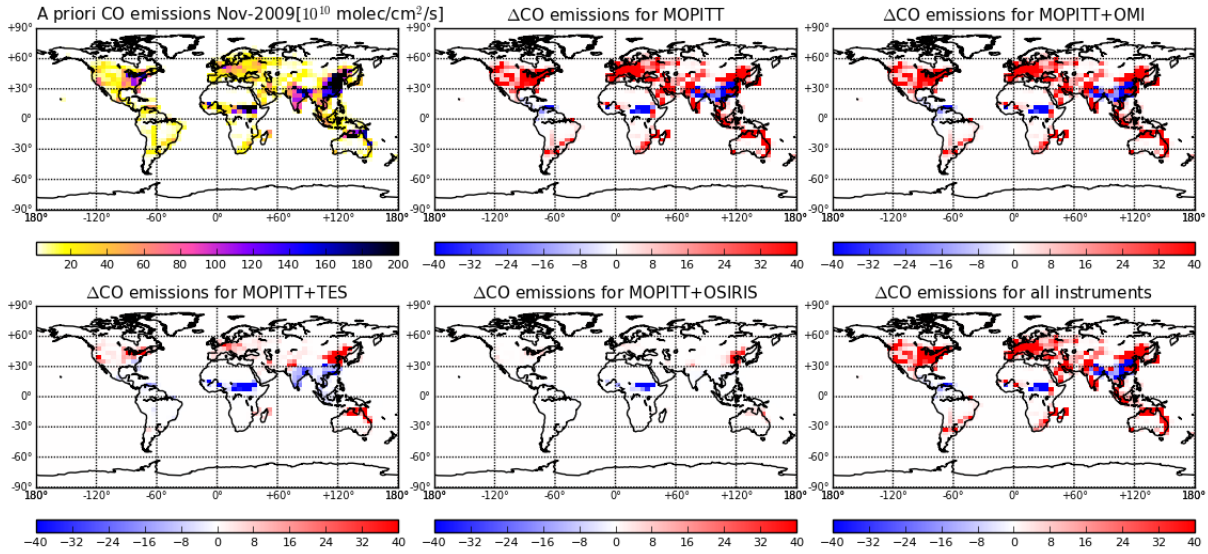

Figure 3.8: The a priori CO emissions (in 10^{10} molec/cm^2/s) and the a posteriori CO emission changes (including both anthropogenic and biomass burning components) between the a priori and the a posteriori for different experiments in November 2009.

and biomass burning emission regions are similar to the experiment using MOPITT CO. However, the magnitude of the estimated emission changes for the assimilation using all instruments are different, suggesting model biases on CO fields have been mitigated by MSA.

To test how other species' assimilation would affect CO emissions, the remaining panels of Fig. 3.8 display the optimized CO emissions for MOPITT + TES, MOPITT+OSIRIS and MOPITT+OMI. Among all three runs, adding OSIRIS O_3 observations has the most impact on perturbing the a posterior surface CO emissions. The a posterior surface CO emissions have retreated to the a priori level in almost all the regions except for East Asia and northern Africa. Two reasons might explain this significant damping effect on emission changes. The first is that the total cost function for the MOPITT+OSIRIS joint assimilation is dominated by OSIRIS O_3, the model then does not respond to CO emission optimization efficiently. The other reason is that less CO adjustments are required to match with the a posterior O_3 concentration for OSIRIS assimilation in the MT and LT. In comparison with OSIRIS O_3 retrievals, MOPITT+TES O_3 assimilation shows some impacts on perturbing the CO emissions in biomass burning regions. However, the relative changes between the MOPITT+TES case and the MOPITT case range around 20-30%, much smaller than the impact of adding OSIRIS assimilation.

To assess the a posteriori emissions quantitatively, the CO emissions have been aggregated over the different continental regions shown in the upper-left corner of Fig. 3.8.

I find the MOPITT+OMI and the all instruments cases are consistent, and are generally in agreement with the MOPITT-only assimilation. For example, for North America the MOPITT, MOPITT+OMI, and the all instrument estimates are 21.5 TgCO/month, 18.4 TgCO/month, and 18.5 TgCO/month, respectively. For Europe, the corresponding estimates are 12.6 TgCO/month, 12.0 TgCO/month and 10.5 TgCO/month, respectively. For the biomass burning emissions in northern equatorial Africa (N. Africa), I find that combining MOPITT with the O_3 observations results in larger departures from the a priori. The MOPITT+TES and MOPITT+OSIRIS emission estimates are 21.8% and 22.1%, respectively, lower than the a priori. In contrast, the all instrument estimate is only 4.3% lower than the a priori. Given the large biases in O_3 in the UTLS and the importance of convective transport in the tropics, it is possible that the larger changes relative to the a priori for MOPITT+TES/OSIRIS runs are due to the assimilation trying to reduce the overestimate in O_3 in the model by reducing the emissions that are convectively transported throughout the free troposphere, counteracting with CO emission increase brought by MOPITT CO assimilation. For the all instrument run, such ozone adjustment was brought by optimized NO_x emissions, which leaves assimilated CO emissions primarily responding to MOPITT CO observations.

The a priori NO_x combustion emissions and the estimated changes in the emissions relative to the a priori for the OMI-only assimilation and for OMI combined with the different datasets are shown in Fig. 3.9. For the OMI-only assimilation, there are reductions of NO_x emissions over the eastern US and Europe, but the changes are heterogeneous over East Asia, with increased and decreased emissions across the region. There is also a modest increase in emissions in the southern US and in the biomass burning region in northern equatorial Africa. Adding MOPITT CO observations to the assimilation does not change the spatial patterns of the optimized emissions significantly. In contrast, ingesting O_3 observations has a large impact on the NO_x emission estimates. The impact is more clearly seen in the regional emission estimates in Fig. 3.10. With TES data included, East Asian emissions are 24% lower than the a priori, whereas the OMI-only estimate is 11% lower. For northern equatorial Africa, the OMI-only case suggested a 5% increase in emissions, whereas the OMI+TES assimilation produced a 58% decrease in the emissions. These changes are occurring because the TES O_3 data override the information coming from the OMI tropospheric NO_2 observations. Consequently, the NO_x emission are compromised to reduce the bias in the modelled O_3 to give a better fit to the TES retrievals. In the all instrument case, in which I balance the contributions to the cost function, the estimated regional emissions are 20% lower than the OMI-only estimate. For the major anthropogenic emission regions, North America and East Asia, I find that

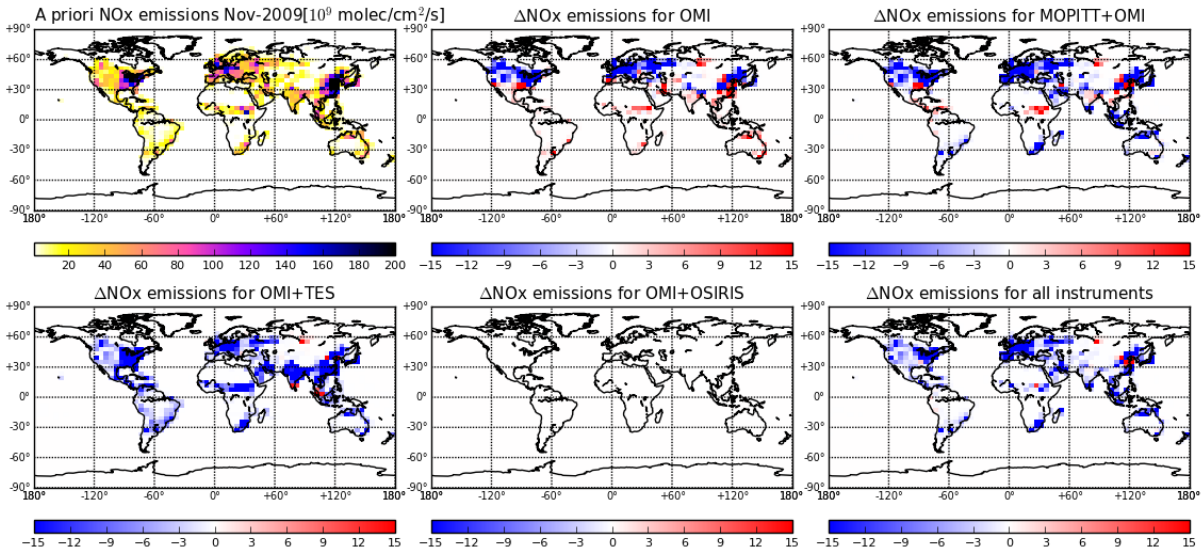

Figure 3.9: The a priori NO_x emissions (in 10^9 molec/cm^2/s) and the a posteriori NO_x emission changes for different experiments in November 2009.

the all instrument estimate is consistent with that inferred in the OMI-only assimilation. Europe is an exception, for which I obtain a large difference (18%) in emissions between the OMI-only and the all instrument cases. The situation with the European emission estimates will be discussed further in Section 3.3.4.

3.3.4 Regional CO and NO_x Emissions in July 2010

The regional CO emission estimates for July 2010 and November 2009 are compared in Table 3.5. In comparison to the results in November, I find differences between the emission estimates from the MSA and those from the MOPITT-only assimilation. For example, for North America and Europe, the MSA CO emission estimates are higher by 30% and lower by 65% than those in the MOPITT-only assimilation, respectively. In November, the MSA estimates for these regions were lower by 14% and 17%, respectively, compared to the MOPITT-only case. In East Asia, the other major anthropogenic source region, the source estimate in the MSA was lower than the MOPITT-only estimate by about 10% in both months. Globally, the total CO emission estimates between the two assimilations differ by 9% in both July and November. Regionally, the European emission estimate from the MSA differed the most from the MOPITT-only assimilation. My results are consistent with those of *Jiang et al.* (2015a), who also used the GEOS-Chem adjoint model, and found that summertime emission estimates for North America and Europe were most sensitive to the different OH fields. They argued that relative to East

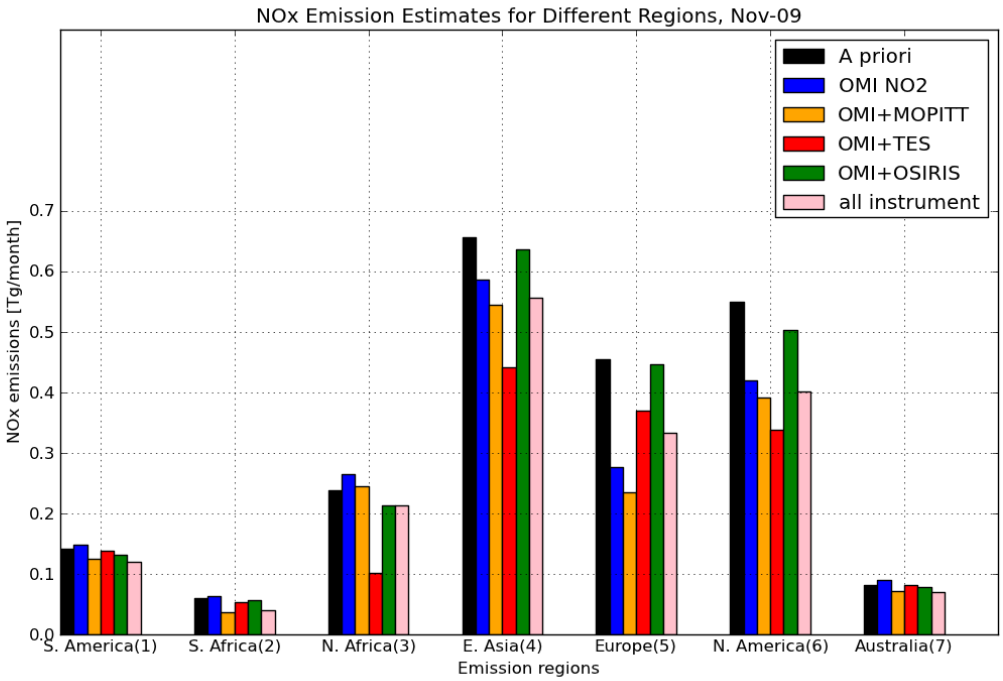

Figure 3.10: Surface NO_x emissions (in TgN/month) in different regions for different experiments in November 2009.

Asia, for example, North American and European CO emissions in GEOS-Chem were more strongly confined to the lower troposphere. This suggests that the free tropospheric CO from these regions was more chemically aged. Thus, inferring emission estimates from observations that exhibit high retrieval sensitivity to CO in the free troposphere could result in emission estimates that are sensitive to biases in OH, since the observations will contain a mix of lower tropospheric air with fresh emissions and free tropospheric air that is more chemically aged. In contrast, emission estimates inferred from observations in the boundary layer will be less sensitive to discrepancies in OH since the measurements will capture fresh CO emissions, with little chemical aging.

For the main anthropogenic NO_x emission regions (North America, Europe, and East Asia), I find that the emission estimates from the MSA for Europe and East Asia are lower than those from the OMI-only assimilation by 16% and 12%, respectively, in July, whereas the North American estimate is 27% higher. In contrast, in November, the North American and East Asian MSA estimates are 5% and 7% lower, respectively, whereas the European estimate is 18% higher. As with CO, I find that the European NO_x emission estimates in both seasons are sensitive to the MSA approach. As noted by *Jiang et al.* (2015a), European emissions are strongly confined to the boundary layer. This could

make estimating European NO_x emissions more sensitive to discrepancies in the vertical distribution of NO_x when using NO_2 column data. Additional work is needed to better understand the source of the differences in the European emission estimates between the MSA and OMI-only assimilation. The large difference between the MSA and OMI-only NO_x emission estimate for North America in July could be due to the fact that we do not optimize isoprene emissions in the assimilation. The isoprene oxidation pathways are NO_x dependent, and discrepancies in the isoprene emissions in the model could impact the local NO_x abundances. The other reason is that the model might try to correct biases owing to LNO_x, which is larger in summer, by over-adjusting surface emissions.

In July, the dominant biomass burning region in the tropics is southern Africa where the MSA suggested a 35% increase in CO emissions and a 43% increase in NO_x emissions over the a priori. In contrast, in November 2009, which is the beginning of the biomass burning season in northern Africa, the MSA suggested reductions of 4% and 15% in the CO and NO_x emissions relative to the a priori, respectively. It is interesting to note that the MOPITT-only and OMI-only assimilations produced increases of 7% and 5% in the CO and NO_x emissions relative to the a priori in November 2009, respectively. Interannual variability in Africa biomass burning emissions is low except for 2010 reported by *Giglio et al.* (2013). For the wildfire season in 2010, the emissions are associated with much higher than averaged burned area in southern Africa (*Giglio et al.*, 2013). My higher estimate is part of an increasing trend in burned area of +1.5 Mha/yr for southern Africa present in both GFED-3 and GFED-4 (*Giglio et al.*, 2010, 2013). *Giglio et al.* (2013) also argued that northern African emissions in 2009 and 2010 experienced significantly reduced burning, which is consistent with my estimated emission changes.

3.3.5 Evaluation of the emission estimates in November 2009 and July 2010

To evaluate the inferred emissions, I compared the a posteriori CO and NO_2 fields with the assimilated data. Fig. 3.11 shows that the MOPITT-only assimilation reduces the mean bias, relative to the MOPITT data, by more than a factor of two, from 12% to 5%. The MSA reduces the bias further, to 4%. The distribution of the MSA difference is also less skewed, with a standard deviation of 14% (compared to 16% for a priori). For NO_2, the OMI-only assimilation also reduces the mean bias by more than a factor of two, from 45% to 20%. However, with the MSA, the residual bias is reduced to only 32%. The larger residual bias in the MSA is expected since the assimilation is trying to adjust the NO_x emissions to accommodate the O_3 data as well as the NO_2 data. This makes the

Table 3.5: Regional surface CO and NO_x emission estimates for the a priori, MOPITT/OMI only, and all instruments runs. The emissions include both anthropogenic and biomass burning components.

	CO regional emissions (TgCO/month)						NO_x regional emissions (TgN/month)					
	November-09			July-10			November-09			July-10		
Regions	a priori	MOPITT	MSA	a priori	MOPITT	MSA	a priori	OMI	MSA	a priori	OMI	MSA
S. America	3.2	4.6	4.1	8.3	3.5	6.3	0.13	0.13	0.11	0.19	0.22	0.20
N. America	7.5	21.5	18.5	15.1	30.8	40.4	0.54	0.42	0.39	0.61	0.44	0.56
E. Asia	16.3	44.8	40.2	13.5	30.9	27.6	0.66	0.59	0.55	0.61	0.67	0.59
SE. Asia	1.8	3.4	3.0	1.7	1.5	1.6	0.06	0.09	0.04	0.06	0.07	0.08
Australia	5.1	19.8	18.1	1.5	2.9	1.5	0.08	0.09	0.07	0.06	0.07	0.06
Europe	4.4	12.6	10.5	5.0	27.4	9.7	0.45	0.28	0.33	0.41	0.45	0.38
N. Africa	9.2	9.8	8.8	1.3	5.0	2.3	0.19	0.20	0.16	0.07	0.11	0.07
S. Africa	1.2	6.5	5.1	17.9	23.4	24.1	0.05	0.05	0.02	0.28	0.44	0.40
Rest of World	8.6	14.4	15.4	17.9	39.7	36.4	0.31	0.33	0.24	0.44	0.46	0.44
Total	57.3	137.4	124.7	82.2	165.1	149.9	2.47	2.14	1.91	2.73	2.93	2.78

adjustment on surface NO_x emissions less effective on correcting the modelled NO_2 bias compared to the OMI-only assimilation. The distribution is more skewed than for CO due to the fact that the modelled NO_2 columns are negatively biased throughout much of the background atmosphere, which the assimilation could not correct. The model is positively biased over the main NO_x source regions, which the assimilation does correct. The better fit to the MOPITT data with the MSA may reflect the impact of a reduced OH bias on the CO simulation mentioned in Section 3.3.2. For July 2010 shown in Figure 3.12, my assimilation also reduces the mean CO and NO_2 biases in both seasons compared to their corresponding assimilated observations. In July, the relative mean bias is reduced from -0.45 (a priori) to -0.07 (all instrument) for CO comparing to MOPITT CO observations, from -0.32 (a priori) to -0.25 (all instrument) for NO_2 comparing to OMI NO_2 observations. Such evidence shows that my DA system improves the model states when comparing to the assimilated observations for both seasons.

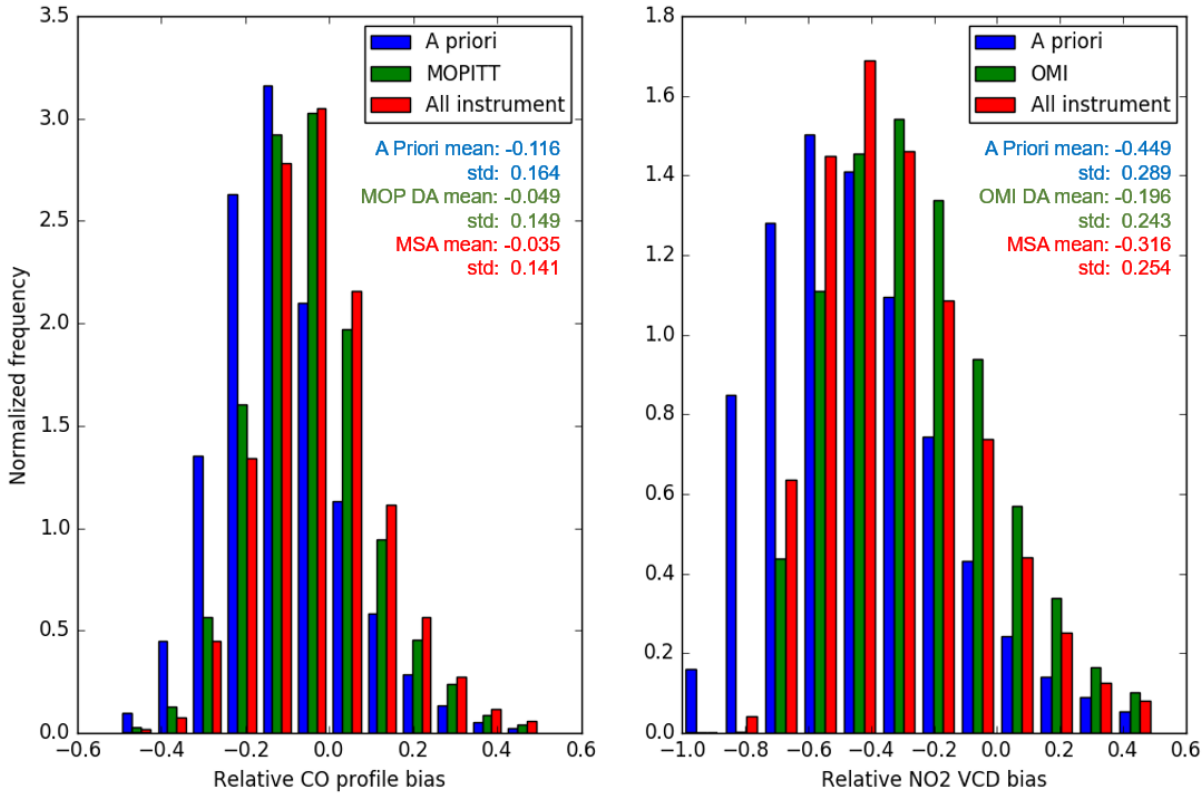

Figure 3.11: Relative bias distribution for November 2009 between the model and MOPITT (left panel)/OMI (right panel) for the a priori (blue) and the a posteriori using MSA (green).

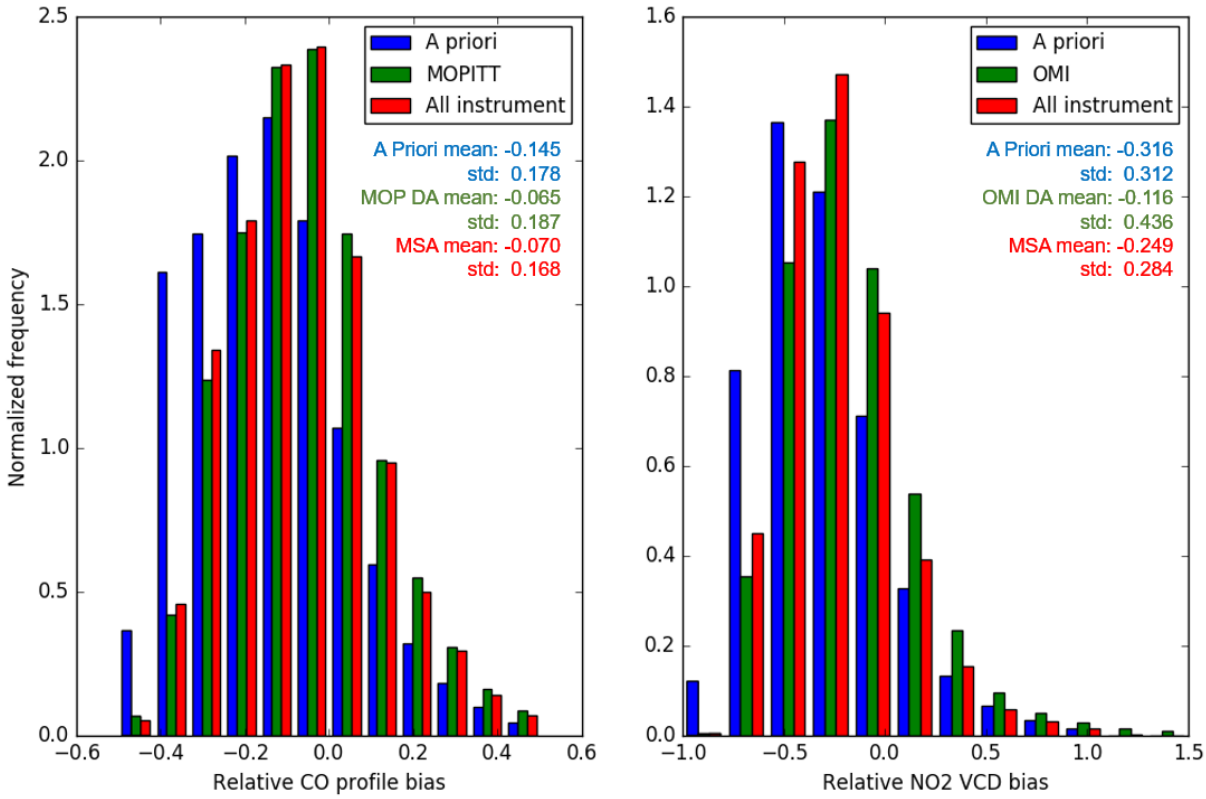

Figure 3.12: Same as Fig. 3.11, but for July, 2010.

3.3.6 Comparison with Previous Inversion Analyses

I compared the assimilated CO emission estimates with the inversion results shown in *Jiang et al.* (2017) and *Kopacz et al.* (2010). For North America I estimated a source of 40 Tg CO with the MSA and 31 Tg CO for the MOPITT-only assimilation for July, whereas *Jiang et al.* (2017) estimated 15 Tg CO with the MOPITT V6 TIR/NIR profile. For Europe, my July estimates were 10 Tg CO and 27 Tg CO from the MSA and MOPITT-only assimilation, respectively. In contrast, *Jiang et al.* (2017) estimated 5 Tg CO. So both my MSA and the MOPITT-only European estimates were larger than those in *Jiang et al.* (2015a). For East Asia I inferred 28 and 31 Tg CO for July for the MSA and MOPITT-only assimilation. *Jiang et al.* (2017) estimated 24 Tg CO. My East Asian emission estimates seem comparable to those of *Jiang et al.* (2017). For Europe and East Asia, *Kopacz et al.* (2010) estimated July 2004 combustion emissions of about 9 Tg CO and 12 Tg CO, respectively, which was comparable to my European MSA estimate of 10 Tg CO, but much lower than my East Asian estimate of 28 Tg CO for July 2010. For November 2004, *Kopacz et al.* (2010) reported European and East Asian combustion emissions of about 11 Tg CO and 21 Tg CO. For November 2009,

my MSA emission estimates for Europe and East Asia are 11 Tg CO and 40 Tg CO, respectively. So my European emission estimates are comparable to *Kopacz et al.* (2010). One issue when comparing my results to other studies is that all the inversion analyses were conducted at different time periods or with different assimilation configurations. For instance, both *Jiang et al.* (2017) and *Kopacz et al.* (2010) used v5-07-08 of GEOS-Chem OH fields, whereas the OH in my analysis is more similar to the v8-02-01 OH fields used by *Jiang et al.* (2015a). *Jiang et al.* (2015a) argued that CO emission estimates using v8-02-01 OH could be 40% higher than using v5-07-08 OH, which explains why my emission estimates are higher in many regions. *Jiang et al.* (2017) also included the biogenic source of CO in their state vector and optimized the VOC source together with the combustion emissions of CO at the model resolution. In comparison, *Kopacz et al.* (2010) separately optimized the combustion and VOC sources of CO. They optimized the combustion emissions at the model resolution and aggregated the VOC source with the methane source to produce a global mean chemical source of CO. For this study, I did not include biogenic emissions in my state vector, so my emission estimate accounts only for combustion-related emissions, which may also explain some differences on emission estimates.

3.3.7 Sensitivity to assimilation window selections

For the experiments mentioned above, I used a two-week assimilation period. The length of the assimilation window was selected to match the lifetime of mid-tropospheric O_3 in the tropics and the repeat time for TES observational coverage. I focused on O_3 since it was a key component of tropospheric chemistry and because the MOPITT and OMI provided greater observational coverage. Here, I vary the assimilation window of the MSA in November 2009 from two weeks to one week, and then to two days. In the experiments with the shorter windows, I am assimilating the data over for the same total two-week period. For example, for the experiment with the two-day window, after the optimized CO and NO_x emissions and O_3 concentrations are obtain at the end of the initial two-day period, they are used as initial parameters for the next two-day assimilation period. In this manner, I stepped through the original two-week period, at two-day intervals.

As shown in Table 3.6, the total regional CO and NO_x emission estimates are robust. Although the total emissions retreat to the a priori with the shorter assimilation windows, the differences relative to the standard two-week case are small. In my experiments, for the total CO emissions, the differences are -4% and -6% for the experiments with one-week and two-day window experiments, respectively. Similarly, for NO_x I found

Table 3.6: Regional surface CO and NO_x emission estimates in November 2009 for the all instrument case using assimilation windows of two weeks, one week, and two days.

	CO regional emissions (TgCO/month)			NO_x regional emissions (TgN/month)		
Regions	AW=2 weeks	AW=1 week	AW=2 days	AW=2 weeks	AW=1 week	AW=2 days
S. America	4.1	4.1	4.1	0.11	0.12	0.13
N. America	18.5	18.6	18.0	0.39	0.41	0.43
E. Asia	40.2	36.5	33.9	0.55	0.58	0.61
SE. Asia	3.0	2.9	3.0	0.04	0.06	0.05
Australia	18.1	15.5	18.9	0.07	0.07	0.07
Europe	10.5	10.9	10.4	0.33	0.32	0.35
N. Africa	8.8	9.3	9.1	0.16	0.20	0.17
S. Africa	5.1	5.4	5.2	0.02	0.03	0.03
Rest of World	15.4	16.5	15.0	0.24	0.27	0.31
Total	124.7	119.7	117.6	1.91	2.06	2.15

that total emission estimates differ by 8% and 13% for the one-week and two-day experiments, respectively. For the regional anthropogenic CO emission estimates, the largest discrepancy between the one-week window experiment and the standard two-week case was for the East Asian emissions, which were 9% lower in the one-week window experiment. This difference increased to -16% for the two-day window experiment. For the NO_x emission estimates, the absolute differences between one-week window experiment and the standard two-week case were less than about 5% for the major anthropogenic source regions. This agrees with *Zhang et al.* (2008b), who also found consistent optimized NO_2 despite varying assimilation windows. The maximum absolute difference increased to about 11% for the two-day assimilation window case. My results suggest that although I originally selected the two-week assimilation window based on the TES observational coverage, I could reduce the window length to one week when we integrate the data from all of the instruments.

3.4 Summary

A case study of multiple species data assimilation using 4D-var data assimilation scheme is applied through the GEOS-Chem adjoint model to optimize the surface emissions of CO and NO_x as well as the initial conditions of O_3. The observations from multiple chemical species are introduced, including MOPITT CO retrievals, TES and OSIRIS O_3 retrievals, as well as OMI tropospheric columns. I found that the multi-species assimilation reduced the absolute mean bias in modelled O_3, relative to HIPPO-2 data, to 10 ppbv or less in the middle and lower troposphere. The residual mean biases were less than 3.5 ppbv

everywhere, with the model overestimating O_3 concentrations, except in the high-latitude UTLS, where the model was biased low, with residual mean biases of 29 ppbv and 14 ppbv between $45°N - 90°N$ and $45°S - 90°S$, respectively. The large residual biases in the UTLS might be due to the low ($4° \times 5°$) horizontal resolution of the model (*Deng et al.*, 2015). At the surface, the MSA reduced mean surface O_3 over North America in July 2010 by 20.1% when validating against EPA AQS data.

For CO and NO_x emissions, I found that in November 2009, the estimates obtained from MSA were generally comparable to those inferred from the MOPITT-only and OMI-only assimilation, respectively. For CO emissions from the main anthropogenic source regions, the largest discrepancies were obtained for the emission estimates for Europe and North America. In November, the European and North American emission estimates from the MSA differed from the MOPITT-only estimates by -17% and -14%, respectively. For the NO_x emissions, the North American and European MSA estimates differed from the OMI-only estimates by -5% and 18%. In July, the differences between the MSA estimates and those based on the MOPITT-only and OMI-only assimilation were larger. The differences for the North American and European CO emission estimates increased to -65% and 31%, respectively. For the North American and European NO_x emission estimates, the July differences were -16% and 27%, respectively. The large difference between the MSA and MOPITT-only estimates for North American and European CO emissions are consistent with the findings of *Jiang et al.* (2015a), who found that summertime emissions from these regions were particularly sensitive by the choice of OH distribution specified in their model. They attributed this sensitivity to the fact that emissions from these regions are relatively more slowly exported to the free troposphere, so the emission signals from these regions in the free troposphere reflect more chemically aged air. Further investigation is needed to better understand the source and seasonal dependence of this sensitivity of the North American and European emissions.

Comparison of the a posteriori CO and NO_2 fields with the assimilated data showed that the MSA provided a better fit to the MOPITT data than the MOPITT-only assimilation. In contrast, the OMI-only assimilation produced a better fit to the OMI data than the MSA. The better agreement between the OMI-only assimilation and the OMI data was due to the fact that the assimilation adjusted the NO_x emission to minimize the model-data mismatch, only by the assumed prior and observation errors. Whereas in the MSA, the NO_x emissions were adjusted to improve the model agreement with the O_3 data as well as with the OMI NO_2 columns. For CO, the closer agreement between the MSA a posteriori fields and the MOPITT data could be due to the reduced chemical bias in the model, associated with the constrained OH fields in the MSA.

My results highlight the importance of producing consistent chemical states for inverse modelling of CO and NO_x emissions. However, this poses several challenges. A major issue is the biases in the modelled O_3 distribution, which affects tropospheric OH abundances and the NO_x-O_3 coupling. The NO_x inversion analysis responded strongly to the modelled O_3 to mitigate discrepancies in between the modelled and observed O_3. Consequently, it is important to effectively mitigate the O_3 biases through assimilation of O_3 data or through correcting modelled OH. Given the short lifetime of O_3 in the lower troposphere, it would be helpful to assimilate O_3 data that can provide global observational coverage on at least daily timescales such as IASI O_3. This would help mitigate the O_3 biases in the assimilation on short timescales and minimize the potential impact of the biases on the shorter-lived chemical tracers. Another issue is that only combustion-related sources were optimized in my inversion analyses. Tropospheric OH was also estimated relying on chemical feedback within the model. The global mean OH levels were decreased by 11.7% and 7.9% in November and July respectively. However, the biogenic source of CO from VOC oxidation is a large component of the CO budget, particularly for North America in the summer (*Hudman et al.*, 2008b), and there are large uncertainties in the bottom up biogenic inventories (*Marais et al.*, 2012). Studies (e.g. *Kleinman et al.*, 1994; *Zhang et al.*, 2004; *Travis et al.*, 2016) shown that these VOC sources, as well as anthropogenic VOC emissions (which are also uncertain), have a strong impact on tropospheric O_3 and OH abundances, and their oxidation is sensitive to local NO_x concentrations. Several inverse modelling studies (e.g. *Barkley et al.*, 2013; *Marais et al.*, 2012; *Kaiser et al.*, 2018) have utilized formaldehyde (HCHO) satellite observations to constrain isoprene emissions (the dominant biogenic CO source). Future multi-species assimilation studies should consider incorporating HCHO data to constrain the CO and VOC emissions as well as tropospheric O_3 and OH concentrations.

Chapter 4

Constraints on the oxidative capacity of the atmosphere: Implications for top-down CO emission estimates

4.1 Introduction

As discussed in Chapter 1, the chemical sources and sinks of atmospheric CO were still poorly characterized, which is largely due to imperfect chemistry in CTMs. A key chemical constituent that is important for tropospheric CO is the hydroxyl radical (OH), which determines the oxidative capacity of the atmosphere. *Jiang et al.* (2015a) and *Müller et al.* (2018) showed that the modelled OH concentrations were a main factor impacting inferred regional CO emission estimates. *Jiang et al.* (2015a) used the same data assimilation configuration with two different prescribed OH fields and obtained differences of up to 46% for the emission estimates for the United States and Europe. *Müller et al.* (2018) used five different prescribed OH fields found that the a posteriori CO field most consistent with independent CO observations was that inferred from the OH distribution with a low ratio (of 0.85) of northern hemispheric to southern hemispheric CO.

In this chapter, we extend the analysis from Chapter 3 to better quantify tropospheric OH and thus reduce the OH-related biases in top-down estimates of CO emissions. Compared to Chapter 3, I try to better constrain the impact of the oxidation of NMVOCs on CO and OH by including HCHO data in the assimilation. I also attempt to improve the optimization of the lightning NO_x source by including HNO_3 data in the UTLS. Hence, I

assimilate CO profiles from MOPITT, HCHO and NO_2 tropospheric columns from OMI, O_3 partial columns from IASI and OSIRIS, HNO_3 profiles from MLS. I also extend the assimilation over the full annual cycle in 2016 by running the model January 2016 to the end of February 2017. To avoid biases due to poor initial conditions at the beginning of the total assimilation period, the January 2016 period can be essentially treated as a spin-up period to correct the background concentrations. The assimilation window is 2 weeks, following the same approach in Chapter 3. For the full year analyses, this means that between January 2016 to February 2017 there are 28 windows in total. The assimilated concentration at the end of window N is saved and used as the initial conditions of the $N+1$ window.

4.2 Model Setup

4.2.1 Assimilation approach

Similar to Chapter 3, the mathematical expression for the 4D-var cost function is as follows:

$$\begin{aligned} J(\mathbf{x_0}, \mathbf{p}) = &[\mathbf{p} - \mathbf{p_a}]^T \mathbf{B}_p^{-1} [\mathbf{p} - \mathbf{p_a}] + [\mathbf{x_0} - \mathbf{x_a}]^T \mathbf{B}_x^{-1} [\mathbf{x_0} - \mathbf{x_a}] \\ &+ \sum_{r=1}^{6} \sum_{n=1}^{N} \gamma_r [\mathbf{y}_{n,r} - H_r(\mathbf{x}_n)]^T \mathbf{R}_r^{-1} [\mathbf{y}_{n,r} - H_r(\mathbf{x}_n)], \end{aligned} \tag{4.1}$$

where $\mathbf{p}$ is the emission parameters including CO, and NO_x emissions, and $\mathbf{x_0}$ is the initial condition parameter which accounts for O_3, HCHO and HNO_3 concentrations at the beginning of the assimilation window. It should be noted that I experimented with optimizing either the isoprene emissions or the HCHO initial conditions in the MSA to provide constraints on the chemical source of CO. $\mathbf{p_a}$ ($\mathbf{x_a}$) and $\mathbf{B}_p$ ($\mathbf{B}_x$) are the a priori estimates and error covariance matrices for $\mathbf{p}$ ($\mathbf{x_0}$), respectively. $\mathbf{y}_{n,r}$ represents the six assimilated observations including MOPITT CO, IASI O_3, OSIRIS O_3, OMI NO_2, OMI HCHO, and MLS HNO_3, at a given time n over the assimilation period, $\mathbf{R}_r$ is the observation error covariance for observation r, which accounts for both measurement error and representation error. H is the observation operator converting model outputs to the observation grid, which has been explained in Section 2.5. The total cost function is then reduced by emission optimization and initial condition optimization. Selection of $\mathbf{B}_p$, $\mathbf{B}_x$, and $\mathbf{R}_r$ will determine how the model will reduce the cost function based on the information given, i.e., how the optimized quantities would respond when trying to minimize the resulting cost function. These will be explained in the observation error

matrix setup below.

In this study, the optimized parameters include surface emissions of CO and NO_x, LNO_x emissions and initial conditions of O_3 and HNO_3. The specific optimized quantities for each experiment are explained in Table 4.1. I set the a priori error covariance matrices, $\mathbf{B}_p$ and $\mathbf{B}_x$, by assuming 50% uncertainty in the bottom-up inventories following *Jiang et al.* (2017), and in initial modelled states following Section 3.2. The only exception is for lightning NO_x emissions, since the LNO_x adjustment is over-sensitive to modelled O_3 and NO_2 biases in the UTLS during the summer. I assign an uncertainty of 20% for LNO_x to avoid over-adjustment. Such a setting would also allow the model to reduce the modelled NO_2 biases by primarily adjust surface NO_x emissions. Due to the coarse resolution of the a priori emissions and initial conditions, no spatial perturbations or correlations were assumed in either $\mathbf{B}_p$ or $\mathbf{B}_x$. As in Section 3.2, I assume diagonal error covariance matrices and thus do not manually impose any correlations between the species. Despite the fact that there is no correlation in the uncertainties among the species, a given optimized quantity can respond to all assimilated observations according to the chemical coupling that is captured by the adjoint forcing in Eq. 2.32.

Similar to the setup of the observation error covariance mentioned in Section 3.2, the following assumptions are made for the observation error of each species. For MOPITT CO, the observation error is set to 20% to account for both measurement and representation error, following *Jiang et al.* (2015a). For IASI O_3, the observation error is equal to 15% of the measured ozone column, same as the approach of *Peiro et al.* (2018) and *Emili et al.* (2014). For OMI, MLS and OSIRIS observations, the observation error for each species uses the measurement error reported in the corresponding retrieval product. Since the spatial and temporal resolution of all the assimilated observations tend to be finer than the employed model, to avoid over-sampling as well as spatial correction among multiple observations, a "super-observation" approach is used to characterize the adjoint forcing of each model grid box.

In the super-observation approach used by *Miyazaki et al.* (2012b) and *Klonecki et al.* (2012), they pre-processed all filtered measurements by computing a weighted measurement (with its weighted uncertainty) for all the measurements falling into a model grid cell. They then conducted the analysis by comparing the modelled states with the super-observation. In contrast, I first compute the adjoint forcing by comparing each observation with the modelled state. I then calculate the super-adjoint forcing for each model grid cell by computing the mean of all the adjoint forcings falling into the model grid cell. One advantage of my approach is that the adjoint forcing mean would effectively serve as a weighted average of all the measurements because the measurement with large

observation error would essentially lead to small adjoint forcing (see Eqs. 2.32 and 2.33 for the adjoint forcing equations). Another advantage is that this super-observation approach can be applied to all assimilated observations to avoid disagreement in the super-observation algorithms among the different assimilated observations. In the end, the cost function weighting (noted as γ in Eq. 4.1) is applied to balance the adjoint sensitivity among observations of different species.

Table 4.1: Optimized quantities for each experiments used in this study, with cost function weight schemes for each multi-species data assimilation. *Note that isoprene emissions are optimized for MSA-isop run only.

Experiments	Optimized parameters and cost function weighting schemes
CTRL run	None
Single Instrument runs:	
MOPITT CO assmilation	CO emissions
IASI or OSIRIS O_3 assimilation	O_3 initial conditions
OMI NO_2 assimilation	Surface NO_x emissions
OMI HCHO assimilation	Surface isoprene emissions
MSA (including MOPITT, IASI, OSIRIS, OMI and MLS data):	CO, NO_x and isoprene* emissions, O_3, HCHO, and HNO_3 initial conditions
MSA-std	γ_{MOP}=16, γ_{IASI}=1, γ_{OSI}=1,$\gamma_{\mathrm{OMI_{NO_2}}}$=16, $\gamma_{\mathrm{OMI_{HCHO}}}$=4, γ_{MLS}=1
MSA-γ=1	γ_{MOP}=1, γ_{IASI}=1, γ_{OSI}=1, $\gamma_{\mathrm{OMI_{NO_2}}}$=1, $\gamma_{\mathrm{OMI_{HCHO}}}$=1, γ_{MLS}=1
MSA-isop	γ_{MOP}=16, γ_{IASI}=1, γ_{OSI}=1, $\gamma_{\mathrm{OMI_{NO_2}}}$=16, $\gamma_{\mathrm{OMI_{HCHO}}}$=4, γ_{MLS}=1

4.2.2 Experiments

Three groups of full-year analyses shown in Table 4.1 were performed for the year 2016 as well as the first two months of 2017. The CTRL run refers to the forward model simulation based on the a priori estimates without assimilation. Single instrument run refers to assimilation on one species using its corresponding observation. The MSA runs refer to the posteriori modelled state when all the observations are assimilated. I propose four MSA experiments with different model configurations. In detail, the standard run of MSA (MSA-std) applies the cost function weighting with a similar assumption as Chapter 3 with γ_{MOP}=16 for CO, and γ_{OMI}=16 for NO_2 and γ_{OMI}=4 for HCHO. The MSA-std run is the default MSA run discussed in Section 4.3. The MSA-γ=1 case refers to the multi-species assimilation without the cost function scaling on any observations. MSA-isop turns on the isoprene emission optimization in the MSA framework, whose cost function scaling applied is the same as MSA-std.

4.3 Results

4.3.1 Assimilation results

Changes in tropospheric O_3

The annual mean O_3 distribution for the a priori, the IASI-only assimilation, and the MSA are shown in Figure 4.1. In the mid-troposphere, the IASI-only assimilation produced increases in O_3 (about 10-15 ppbv) in northern high-latitudes and decreases (exceeding 15 ppbv) in the subtropics. There are also enhancements in O_3 over equatorial Africa. It should be noted that no observations were assimilated poleward of 75°N, which means the O_3 changes in northern high-latitudes are driven by the model transport of O_3 to these regions. The same spatial pattern of O_3 corrections in the mid-troposphere is present in the lower troposphere, but with a decrease in the magnitude of the changes. In the UTLS, however, there are large negative corrections poleward of 30°. The assimilated results from the MSA showed the same correction patterns as the IASI-only assimilation in mid-troposphere and UTLS. Near the surface, there are additional changes in O_3 in the MSA that are produced by adjusting the O_3 precursor emissions. For instance, the assimilated ozone is lower by 3-5 ppbv in North America, which is a result of the chemical response to the decreased NO_x emissions in the MSA (shown in Fig. 4.13).

The O_3 time series in different latitude and altitude bins are shown in Figure 4.2. It should be noted that the modelled O_3 time series is produced by optimizing the initial conditions for two-week intervals throughout the year. Since the initial conditions for each adjacent two-week assimilation window (e.g., O_3 abundances on January 1st vs January 16th) are not continuous, the resulting assimilated time series exhibit larger month-to-month variations than the a priori run. In the northern extratropics, the a priori model has a strong seasonal cycle in the upper troposphere that decrease significantly in amplitude at lower altitudes. Both the IASI-only assimilation and the MSA produce a stronger seasonal cycle in the middle and lower troposphere, with enhanced O_3 in March-April-May (MAM). In the southern extratropics, the seasonal cycle is largest in the UTLS and the amplitude does not change in the assimilation experiments. However, both assimilations reduce O_3, with the MSA reducing O_3 from 110-160 to 80-130 ppbv over the seasonal cycle. In the middle troposphere, the IASI-only assimilation and the MSA produce consistent O_3 abundances, which are slightly lower than the a priori. Near the surface in the southern extratropics, the assimilation experiments have a negligible impact on O_3 abundances. In the tropics and subtropics, the largest reductions in O_3, relative to the a priori, are in the SH, where I find that the MSA reduces O_3 abundances

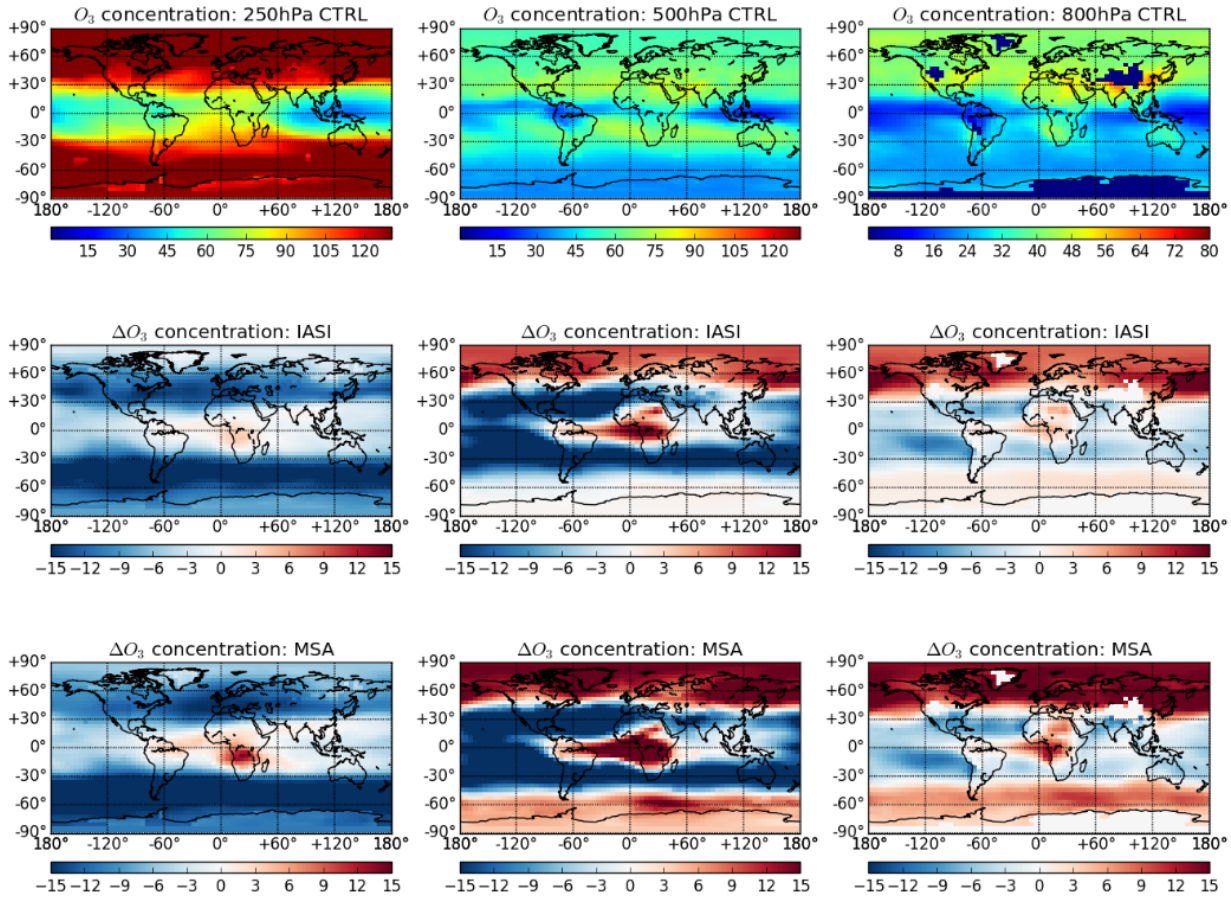

Figure 4.1: Modelled O_3 abundances (in ppbv) for the a priori (CTRL) run (top row) and the assimilated changes in ozone (in ppbv) relative to the a priori for the IASI-only assimilation (middle row) and the MSA (bottom row). Shown are O_3 changes at 250 hPa (left column), 500 hPa (middle column), and 800 hPa (right column).

at all levels.

If I define the tropopause at 100 hPa within ±32° latitude and at 250 hPa in the extratropics, the total tropospheric O_3 burden for the a priori, the IASI-only assimilation, and the MSA are 360, 343 and 311 Tg O_3/yr, respectively. The a priori O_3 burden is greater than the observation-based estimate of 337 Tg that is based on ozonesonde observations for 2010-2014 (*Gaudel et al.*, 2018). In contrast, the O_3 burden in the IASI-only assimilation is consistent with the ozonesonde-based estimate, whereas the burden in the MSA slightly underestimates the burden.

Changes in isoprene emissions and HCHO abundances

The a priori isoprene emissions and the changes in the emissions in the HCHO-only and the MSA-isop assimilations are shown in Fig. 4.3. In the HCHO-only assimilation, iso-

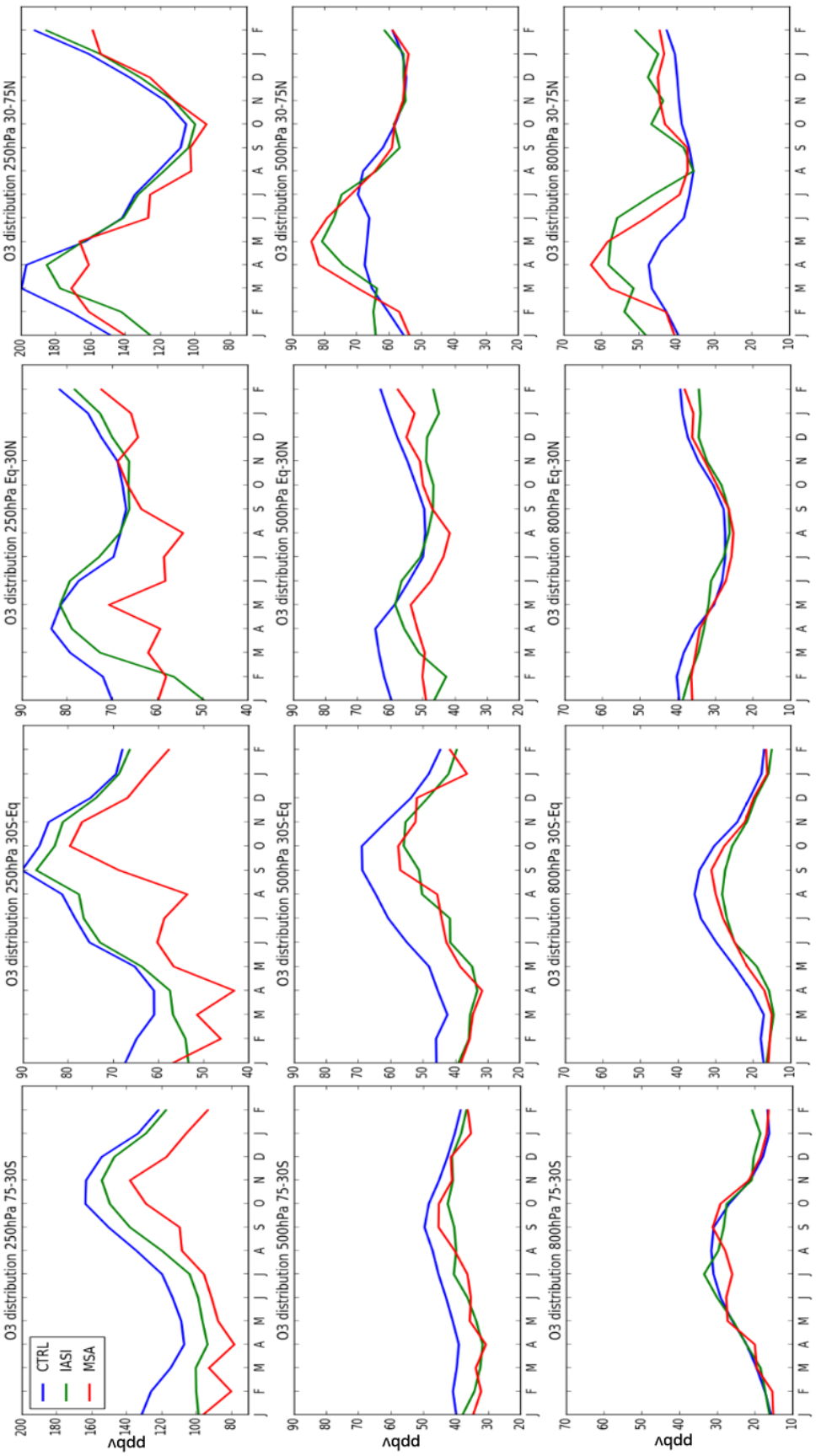

Figure 4.2: Monthly O_3 time series (in ppbv) for CTRL, IASI O_3 assimilation and MSA-std at 800 hPa, 500 hPa, and 250 hPa between at 75°S-30°S, 30°S-Eq, Eq-30°N, 30°N-75°N.

prene emissions are reduced in the southeastern US, the Amazon, southeast Asia, and in central Africa. The reductions of the isoprene emissions in the the HCHO-only assimilation are consistent with recent HCHO inversion studies (e.g., *Barkley et al.*, 2013; *Marais et al.*, 2012; *Kaiser et al.*, 2018). However, in the MSA-isop assimilation isoprene emissions are increased everywhere, which is inconsistent with my current a priori understanding that MEGAN may be overestimating isoprene emissions. Furthermore, in the MSA-isop assimilation, the chi-squared statistics of assimilated HCHO increased by 30% after the assimilation, which suggests that the assimilated HCHO is more inconsistent with the observation than the a priori HCHO. This inconsistency is due to a discrepancy in the isoprene oxidation scheme in the model. In the version of the adjoint employed here, isoprene oxidation acts as a sink for NO_x, resulting in a negative correlation between tropospheric O_3 and isoprene emissions (*Mao et al.*, 2013). Consequently, in the MSA-isop assimilation the optimization increases the isoprene emissions to reduce the positive bias in O_3 in the tropics and subtropics. In a more recent version of the oxidation scheme, as described by *Mao et al.* (2013), NO_x is recycled to the atmosphere instead of being lost to the formation of isoprene nitrates, which can produce a positive correlation between O_3 and isoprene emissions. However, the adjoint of the updated isoprene oxidation scheme is not available in the version of the adjoint model used in this study. This study therefore chose not to optimize the isoprene emissions in the standard MSA assimilation (MSA-std), although the HCHO retrievals are ingested in the assimilation. Instead, HCHO initial condition optimization in MSA-std is to provide indirect constraints on CO production from the chemical oxidation of NMVOCs.

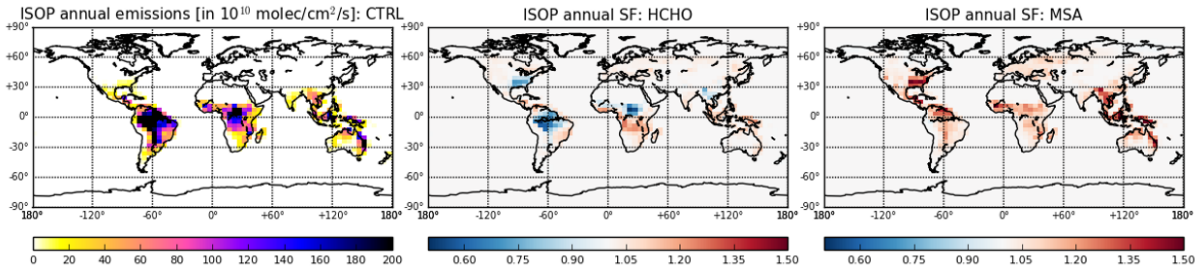

Figure 4.3: Modelled a priori isoprene emissions (left panel) as well as the scaling factors for the emissions in the HCHO-only assimilation (middle panel) and the MSA-isop assimilation (right panel). The scaling factor is the ratio of the a posteriori emissions to the a priori emissions.

The a priori HCHO abundances in the lower troposphere are shown in Figure 4.4 together with the changes in HCHO in the HCHO-only assimilation and the standard MSA. High HCHO concentrations are present over the main isoprene emission regions

shown in Fig. 4.3. In the HCHO-only assimilation HCHO is significantly reduced over the main source regions in the tropics. There are increases in HCHO over northern Canada, boreal Eurasia, and eastern China. There are also slight increases over the tropical oceans. For the MSA, the a posterior HCHO changes in the tropics follow the pattern of HCHO changes in the HCHO-only assimilation with large reductions over the continental source regions. However, in the MSA background HCHO is also reduced over the oceans, due to the O_3 assimilation. The consistency in the changes in HCHO over the main source regions between the HCHO-only assimilation the standard assimilation suggests that excluding the isoprene emissions from the state vector in MSA-std is not strongly biasing the HCHO analysis.

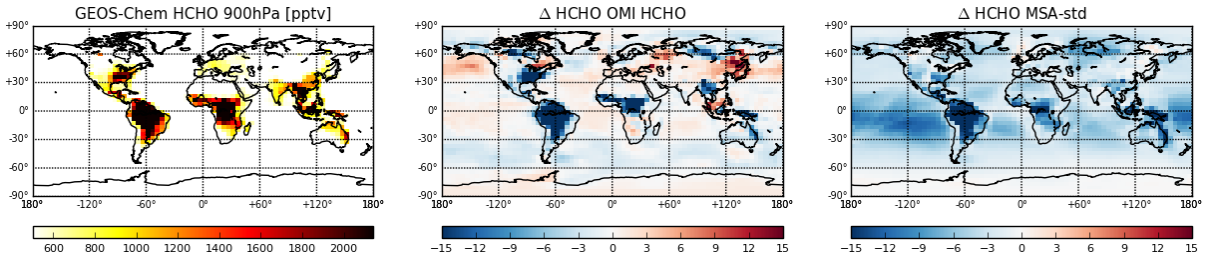

Figure 4.4: Modelled HCHO concentrations (in pptv) for CTRL run, as well as HCHO changes (in pptv) for OMI HCHO assimilation and MSA-std at 800 hPa.

Changes in tropospheric OH

The annual mean, mass weighted modelled tropospheric OH is shown in Fig. 4.5. As shown in a priori distribution, tropospheric OH is most abundant over the tropics and subtropics (30°S-30°N). The model also shows local minima in tropical OH over the Amazon (80°-60°W), equatorial Africa (25°-35°E), and Indonesia (130°-145°E). These minima are associated with the titration of OH due to the oxidation of NMVOCs. Among the different assimilation experiments, the MOPITT-only assimilation tends to reduce the modelled OH globally by 3-5% due to the chemical feedback from the increased modelled CO throughout the assimilation period. The OMI NO_2-only assimilation shows more localized reductions on OH over North America, Europe, East Asia, and southeast Asia due to the OH response to changes in the modelled NO_2. The OMI HCHO-only assimilation enhances the OH abundance by more than 10% over regions with high NMVOC emissions, such as the southeastern US, the Amazon, and equatorial Africa. As shown in Fig. 4.3, the HCHO-only assimilation tends to reduce isoprene emissions in these regions, which results in increased OH due to the isoprene titration of OH in this version of the adjoint model. Additionally, there are positive changes in OH over the northern high

latitudes in the HCHO-only assimilation. The IASI-only assimilation produces negative OH changes within the subtropics, except over the Amazon and equatorial Africa. These changes together with the enhancement in OH found over northern high-latitudes are OH responses to the assimilated O_3. However, for the OMI HCHO assimilation and the IASI O_3 assimilation, the positive OH changes over the northern high-latitudes are unrealistic since they would further exacerbate the existing high biases in OH. The MSA shows the OH responses to all of the assimilated quantities. Modelled OH is reduced by 10-15% over the tropics (except for equatorial Africa) and by 5-15% over North America and Europe. Moreover, OH responses in the MSA tend to reflect the changes due to the O_3 assimilation in the tropics and subtropics and to the CO and NO_2 assimilation in the northern extratropics. It should also be noted that the MSA does not produce positive changes in OH over the northern high latitudes, which mitigated some of the posterior biases introduced by the OMI HCHO and IASI O_3 assimilation.

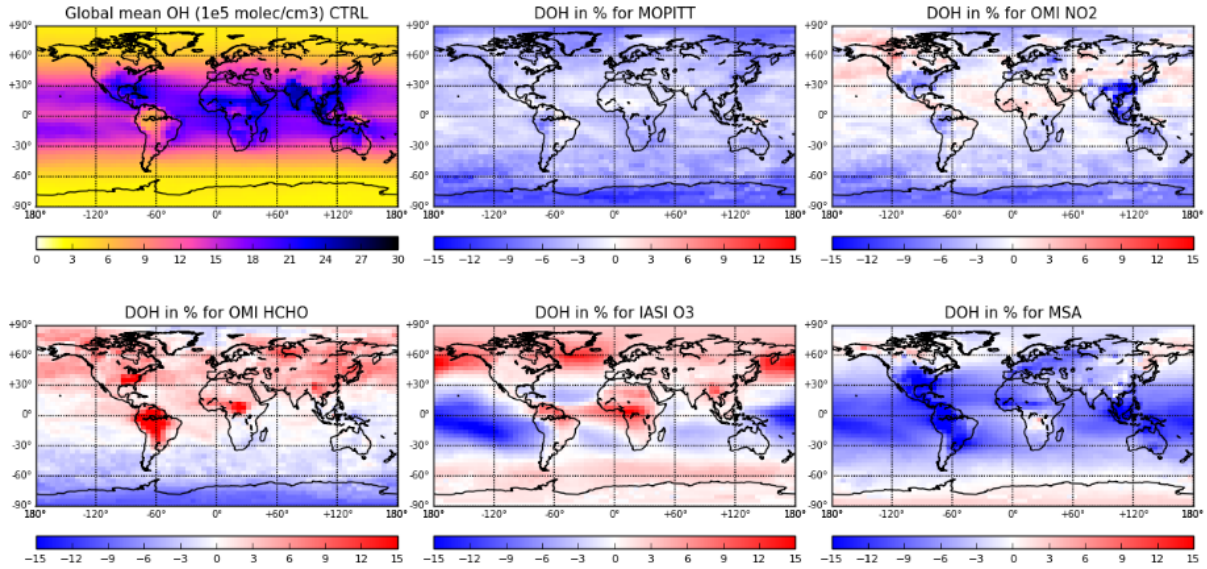

Figure 4.5: Tropospheric mass weighted mean OH (in 10^5 molec/cm^3) in for CTRL run, as well as relative changes in percentage ($\frac{post-ap}{ap} \times 100\%$) compared to the a posteriori for MOPITT assimilation, OMI NO_2 assimilation, OMI HCHO assimilation, IASI O_3 assimilation and MSA-std.

Figure 4.6 compares the time series of global mean, monthly mean, mass-weighted tropospheric OH as well as the ratio between NH and SH OH (given as the NH/SH ratio) for the different assimilation experiments and for the *Spivakovsky et al.* (2000) OH climatology. The OH climatology from *Spivakovsky et al.* (2000) is still widely used for inverse modelling of CO and CH_4 emissions since it provides a 3D OH distribution that was empirically derived from observations. The time series of OH for the a priori has a

strong seasonal cycle with high OH abundances in July-August-September (JAS) and low abundances in December-January-February (DJF). However, the a priori global mean OH is consistently higher (by 0.8-1.5×10^5 molec/cm^3) than *Spivakovsky et al.* (2000) OH climatology throughout the assimilation period. The MOPITT-only assimilation reduces the global mean OH in response to the enhanced assimilated CO, but is still much higher than the *Spivakovsky et al.* (2000)'s OH climatology. In contrast, the MSA produces even larger reductions that capture the climatology well in June-July-August (JJA). The global mean OH inferred from the MSA is 12.0×10^5 molec/cm^3, which is close to *Spivakovsky et al.* (2000)'s estimate of 11.6×10^5 molec/cm^3, whereas the a priori estimate was 13.0×10^5 molec/cm^3. The MSA global mean OH estimate is within the multi-model mean of 11.1±1.6×10^5 molec/cm^3) in ACCMIP (*Naik et al.*, 2013). The estimated methane lifetime for the a posteriori OH from the MSA is 8.68 years, which falls within the ACCMIP range of 9.8±1.6 years reported by *Voulgarakis et al.* (2013). Compared to the observationally-based estimate of the methane lifetime (for conditions for 2000) of 10.2±0.9 years (*Naik et al.*, 2013), the lifetime obtained here suggests that the OH abundance in the MSA is still biased high.

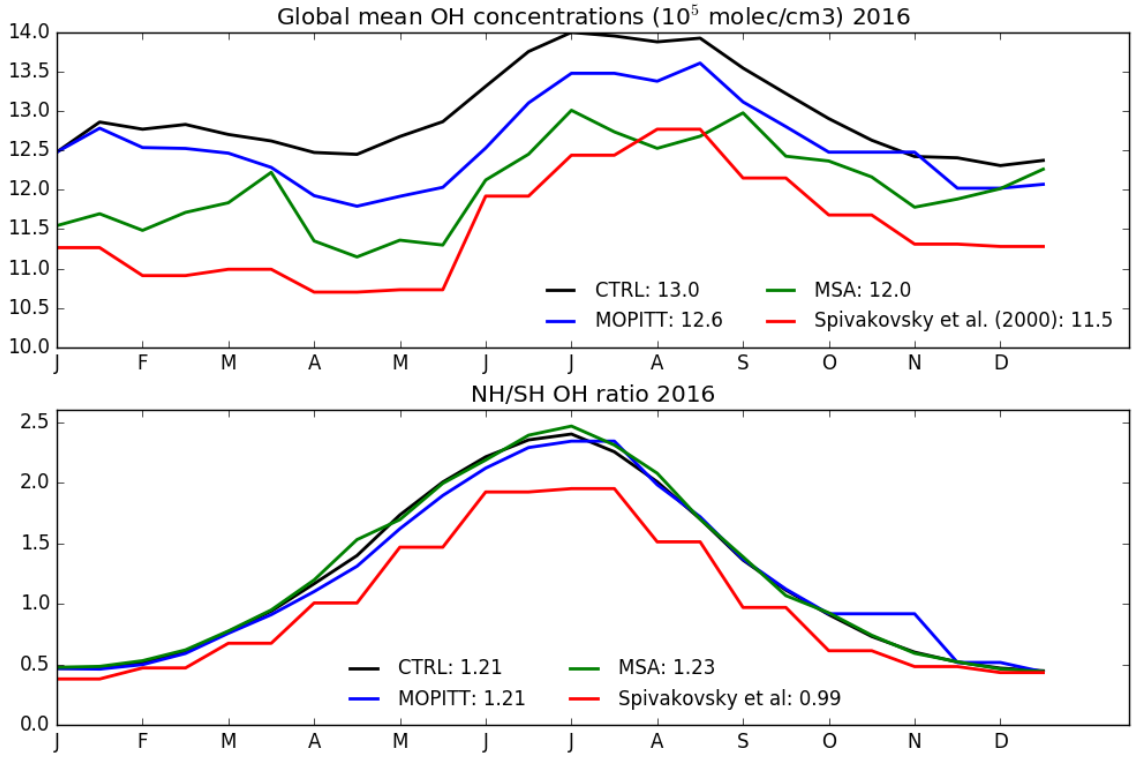

Figure 4.6: Time series of global mean, monthly mean, mass-weighted tropospheric OH (top panel, in units of 10^5 molec/cm^3) and the hemispheric ratio (NH/SH) of OH (bottom panel) for a priori simulation, the MOPITT-only assimilation, MSA, and the OH climatology from *Spivakovsky et al.* (2000).

Another key metric to evaluate the global OH distribution is the NH/SH ratio. Similar to *Spivakovsky et al.* (2000), both the a priori and the assimilation experiments have a higher NH/SH ratio (more than a factor of 2) in JJA, and a lower NH/SH ratio (0.5-0.6) in DJF. The annual mean NH/SH ratio for the a priori OH is 1.21, which is larger than the *Spivakovsky et al.* (2000) ratio of 1.0, but consistent with the ACCMIP multi-model mean *Naik et al.* (2013). For the assimilated results, neither the MOPITT-only assimilation nor the MSA significantly changed the interhemispheric ratio. Indeed, the MSA slightly increasing the ratio from 1.21 to 1.23. The insensitivity of the NH/SH ratio in the MOPITT-only assimilation and the MSA makes sense since the OH changes (shown in Fig. 4.5) are fairly symmetric across the equator. In contrast, *Miyazaki et al.* (2015), they found that their multi-species assimilation (for 2005-2012) reduced OH abundances in the NH and slightly increased it in the SH, resulting in a NH/SH ratio of 1.18. The differences between my results and those of *Miyazaki et al.* (2015) could be related to differences in transport in our atmospheric models as well as the data assimilation schemes used in the analyses. *Miyazaki et al.* (2015) used a Kalman filter scheme that sequentially adjusts the model state and could therefore mitigate transport-related biases. In contrast, my 4D-Var scheme adjusts only the initial conditions and assumes that the model transport is perfect.

4.3.2 Evaluation of the assimilation against TCCON XCO data

Figure 4.7 shows the time series of TCCON XCO binned into six different regions. The individual sites included in each region are listed in Table 2.1. The CO columns simulated by the a priori (CTRL) run underestimate XCO relative to the TCCON data by 15-45%. However, although the model is systematically low, it captures the seasonal cycle in the NH well, with higher CO in winter and lower CO in summer, reflecting the greater OH abundances in summer. During the first two months, both assimilations do not improve XCO by more than 15% due to the spin-up of the system. But from March 2016 to February 2017, both the MOPITT-only assimilation and the MSA better match the TCCON XCO across the NH. Over North America, the mean XCO from TCCON was 86 ppb; whereas for the a priori, the MOPITT-only assimilation, and MSA, they were 69, 83, and 84 ppb, respectively. Similarly, over Europe the mean XCO from TCCON was 89 ppb, whereas for the a priori it was 74 ppb and for both the MOPITT-only assimilation and MSA they were 89 ppb. In the SH, the XCO peaks during SON, which is associated with the transport of biomass burning emissions from the tropical continental regions. Both the MOPITT-only assimilation and the MSA reproduce the

peak in XCO over Oceania during the biomass burning season, but are biased low for the rest of the year, indicating the system's inability to correct background CO in the SH. This underestimate of SH background CO could be due to model transport errors. It could also indicate that OH in the model is still overestimated. Both assimilations have difficulties reproducing XCO over the SH oceans. These sites include Ascension Island and Reunion Island, which experience the influence of African biomass burning emissions. Discrepancies in the transport of these emissions could contribute to the discrepancies shown in Fig. 4.7. Indeed, *Arellano et al.* (2006) noted a difficulty in reproducing observations over Ascension Island in their inversion analysis of MOPITT CO. They suggested that this discrepancy could be due to errors in the altitude of the convective outflow from Africa. Recently, *Stanevich* (2018) assimilated CH_4 retrievals from GOSAT to characterize errors in the GEOS-Chem model and found that the inferred model errors over Africa suggested that there could be too much deep convective transport over Africa and insufficient shallow outflow to the Atlantic in the model.

4.3.3 Evaluation of the assimilation against ATom-1 and ATom-2 aircraft data

This section evaluates the modelled OH, O_3, and CO over the Atlantic and Pacific oceans using data from the ATom-1 (in Fig. 4.8) and ATom-2 (in Fig. 4.9) aircraft campaigns. As shown in Fig. 2.15, ATom-1 flew from August 4th to August 31st in 2016. ATom-2 flew from January 26th to February 22nd in 2017. For the comparison, I bin the data over each ocean basin into NH and SH bins to show the seasonal contrast between ATom-1 and ATom-2. The observed OH, O_3, and CO measurements are compared with the model at the nearest model time and location.

OH evaluation

As shown in Figs. 4.8 and 4.9, the a priori is generally able to reproduce the ATom observations in the lower and middle troposphere over the Atlantic Ocean. The largest a priori biases are in the upper troposphere, which the assimilations are not able to reduce. Over the Pacific Ocean, the largest discrepancies are in the NH during ATom-1 and in the SH during ATom-2. MOPITT CO assimilation does not produce significant OH responses. Only 5% decreases are found in mid-troposphere for the a posteriori OH in the Pacific leg in NH, which is related to the strong CO-OH coupling in the NH summer. In contrast, the OH response in the MSA framework is much stronger, with improvements found for the a posterior OH in both legs in both hemispheres. For ATom-1,

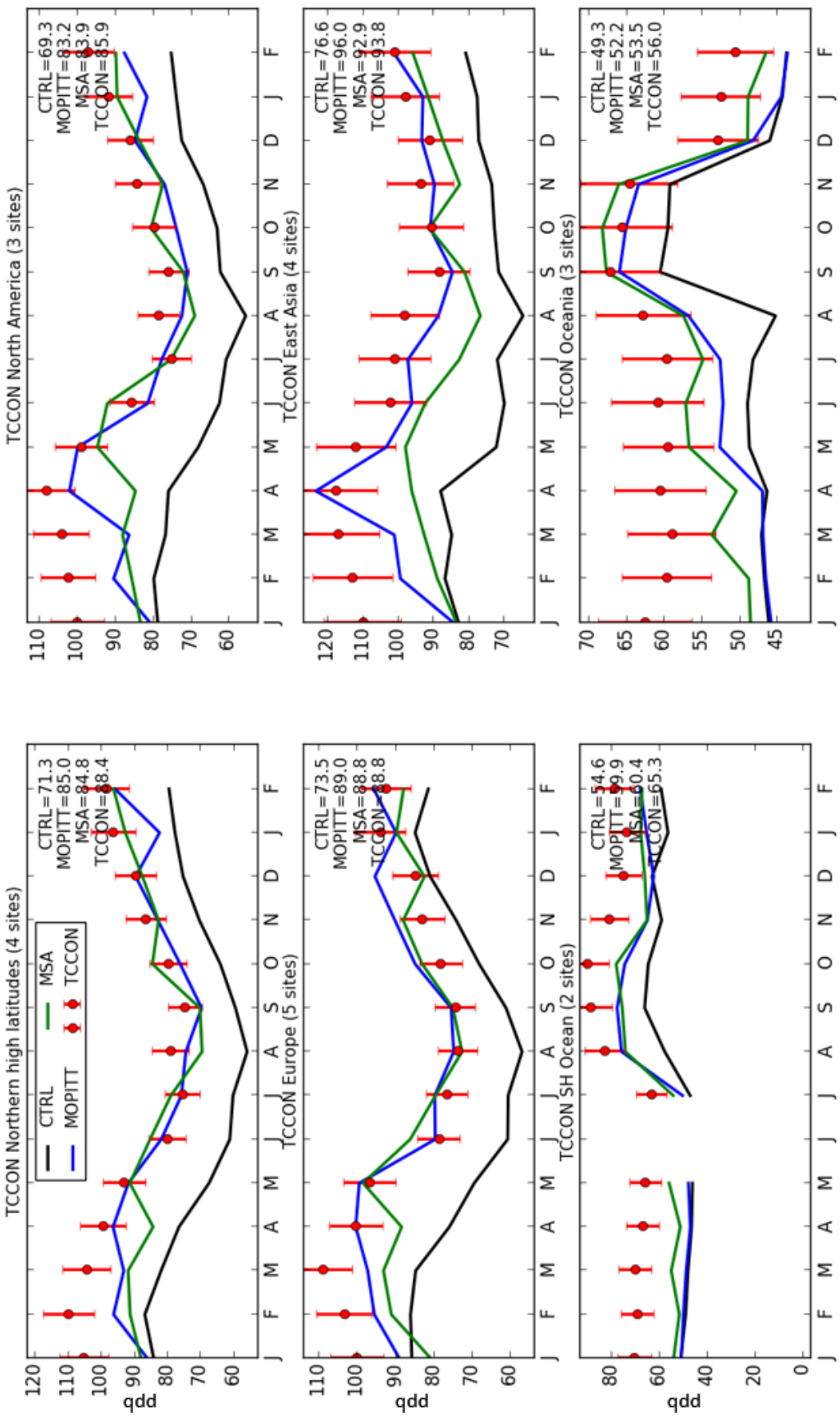

Figure 4.7: Comparisons of modelled CO column abundances (XCO) from the CTRL, the MOPITT assimilation, and MSA-std with TCCON data. Individual TCCON sites that are included in each of the six regions are listed in Table 2.1. The annual mean XCO (in ppb) in each region is shown at the top right in each panel.

the MSA reduces the overestimate on OH over the NH Pacific by 15-30%, but produced an underestimate in OH in the lower troposphere. Similarly, the MSA improved the modeled OH in the middle troposphere over the SH Pacific during ATom-2, but enhanced the bias near the surface.

O_3 evaluation

For both ATom-1 and ATom-2, the a priori run (CTRL) consistently overestimates the tropospheric O_3 by 25-60 ppbv, with the largest biases in the UTLS. In general, the IASI-only assimilation and the MSA produce similar improvements in the modelled O_3. For example, the assimilations reduce the O_3 overestimation by 15-30 ppbv in the SH for both Atlantic and Pacific legs of ATom-1. In the NH during ATom-1, the MSA produces larger bias reductions in the middle troposphere than the IASI assimilation. However, in the NH during ATom-2, the IASI-only assimilation produce larger bias reductions. This could be due to the fact that in NH summer (ATom-1 period), the assimilated constraints on the O_3 precursors in the MSA lead to larger adjustments to the O_3 distribution than assimilation of only IASI data. In contrast, in winter (ATom-2 period), when the O_3 lifetime is longer, assimilation of the precursor constraints may not provide much additional benefit beyond that from assimilation of only the IASI data.

CO evaluation

As with the TCCON comparison, the a priori model (CTRL) significantly underestimate the observed CO, with biases in a magnitude of 10-46 ppbv throughout the troposphere. Both the MOPITT-only assimilation and the MSA produce enhanced CO abundances in both hemispheres. However, the MOPITT-only assimilation best reduces the bias in the NH, whereas the MSA produces CO abundances in closer agreement with the observations in the SH. One exception occurs over the Atlantic Ocean in the SH during ATom-2, for which the MOPITT-only assimilation reproduces the observations well in the lower and middle troposphere. This peak in CO is likely associated with the transport of biomass burning emission from northern tropical Africa to the southern tropical Atlantic. In Chapter 3, it was shown that the MOPITT-only assimilation provided a closer fit to the MOPITT data than the MSA due to the fact that in the MSA the optimized CO must be consistent with the MOPITT data as well as other observational constraints. This could explain the smaller biases between the ATom data and the MOPITT-only assimilation in the NH. It is difficult, however, to explain the large residual biases between both assimilations and the ATom data given the good agreement obtained with the TCCON

data. A noticeable discrepancy between the model and the observations is the difference in the vertical gradient in CO over the SH Pacific in ATom-2. The observations show increasing CO with altitude, which the model does not exhibit and which the assimilations do not correct. The higher CO abundances in the upper troposphere may be due to long-range transport of CO in the UTLS, and an error in transport cannot be corrected for in the assimilation since only the CO emissions are optimized in the system. This suggests data assimilation schemes that consider model error in their cost function such as weak-constraint 4D-Var could be used to further mitigate the model biases due to transport, which will be discussed in Chapter 6.

4.3.4 Evaluation of the assimilation against SHADOZ data

Tropospheric O_3 in the tropics and subtropics are compared with SHADOZ Ozonesonde observations in Fig. 4.10. The a priori simulation is positively biased in the middle and upper troposphere by 20-60%. The IASI-only assimilation improves the modelled O_3 mostly in the middle troposphere with an improvement of 4-12 ppbv. However, O_3 biases in UTLS are largely unchanged. In contrast, the MSA, due to the assimilation of both tropospheric and stratospheric observations (IASI and OSIRIS), is able to improve O_3 from the middle troposphere to lower stratosphere. In the UTLS, the assimilation of OSIRIS O_3 and MLS HNO_3 observations in the MSA reduced the modelled O_3 biases by 15-25 ppbv. I also found that the MSA was most effective at improving the O_3 in the UTLS in the subtropics, such as at Hilo, La Reunion Island, and Irene.

4.3.5 CO emission estimates

Figure 4.11 shows the annual mean CO emissions including both anthropogenic and biomass burning sources based on the bottom-up inventories during March 2016 to February 2017, as well as its estimated emission scaling factors for the MOPITT and the MSA assimilations. The emission scaling factor is the ratio of the a posteriori to a priori emissions. The bottom-up inventories show large anthropogenic emission sources in North America, Europe, and East Asia. These continental regions account for 10.5%, 11.4%, and 56.4% of the total annual anthropogenic emission budget, respectively. The biomass burning sources in Amazonia, equatorial Africa, and Southeast Asia contribute 10.7%, 64.4% and 32.1% to the global biomass burning budget, respectively. After the assimilation, the MOPITT CO inversion increased emissions almost everywhere, with the largest correction over the eastern US, Europe, India, East Asia, and equatorial Africa. The CO emissions estimated in the MSA also shows increased emissions over these regions, but

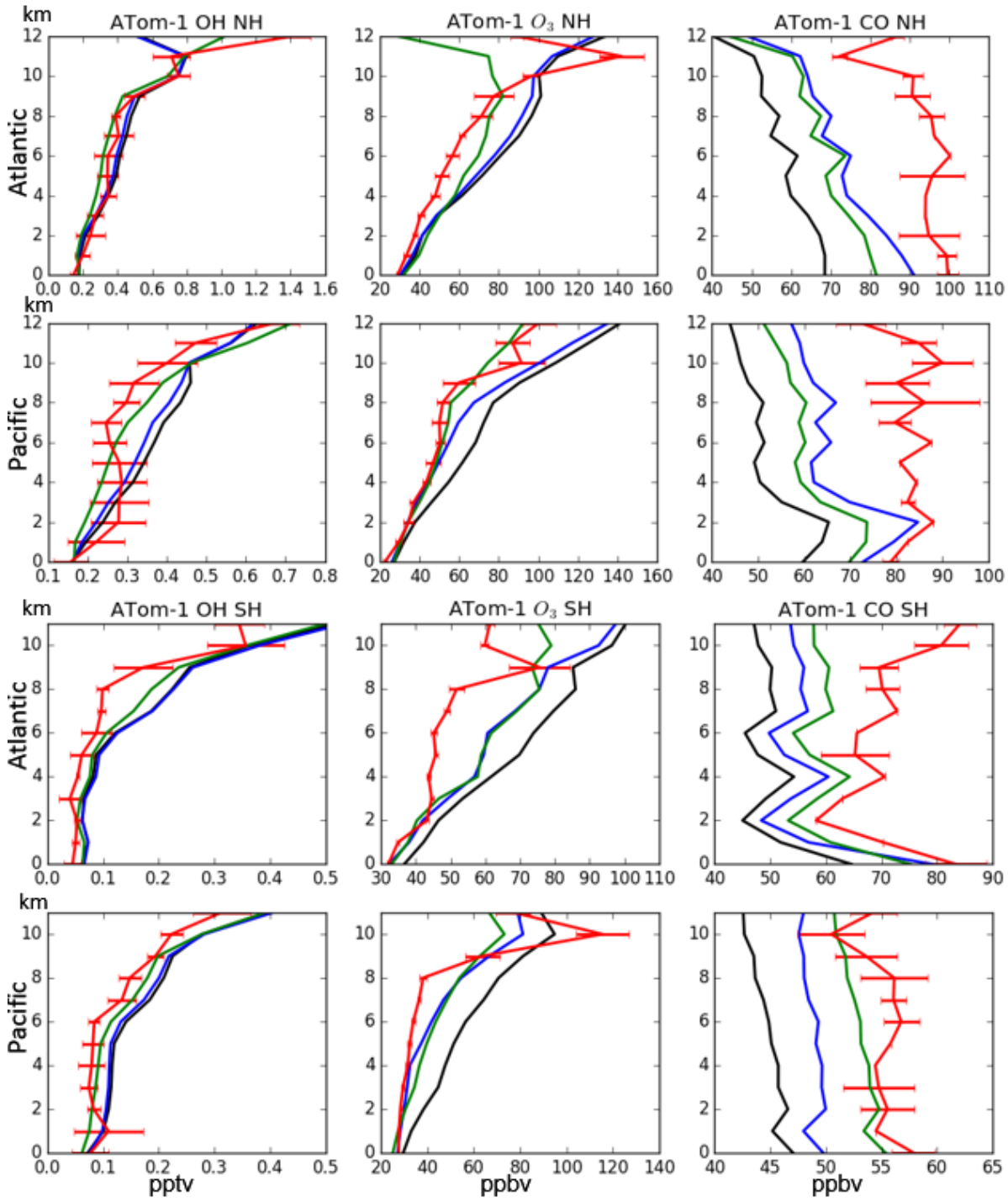

Figure 4.8: Comparisons of ATom-1 observations with modelled OH (left column), O_3 (middle column), and CO (right column) profiles from the a priori (CTRL) run, the single instrument assimilation, and the MSA-std. The single instrument assimilation refers to the MOPITT-only assimilation for CO and OH, whereas it refers to the IASI-only assimilation for O_3. The vertical profiles show the mean abundance of OH (in pptv), O_3 (in ppbv), and CO (in ppbv) averaged over the northern (top two rows) and southern (bottom two rows) hemispheres and over the Atlantic and Pacific oceans.

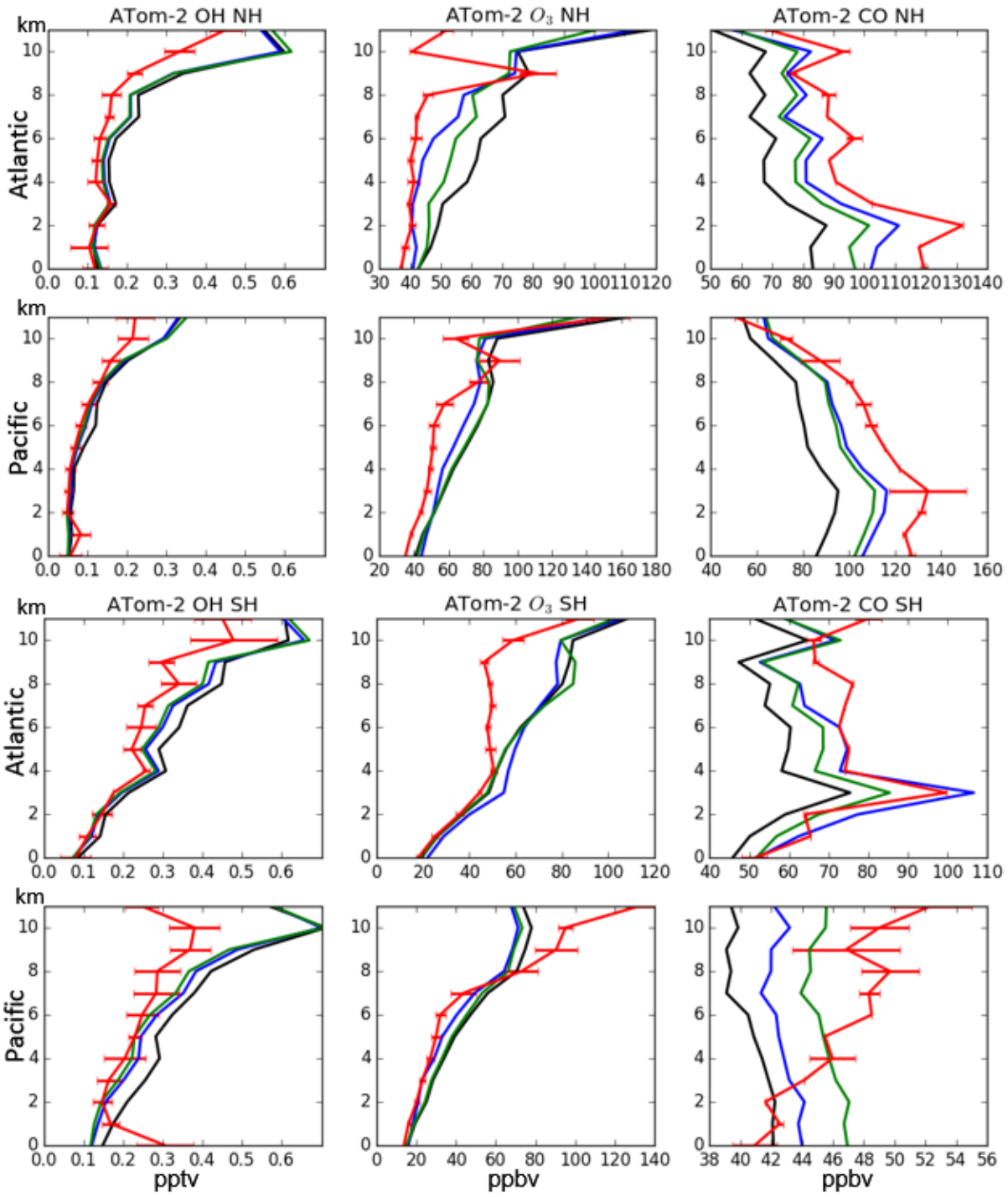

Figure 4.9: Same as Fig. 4.8, but for ATom-2.

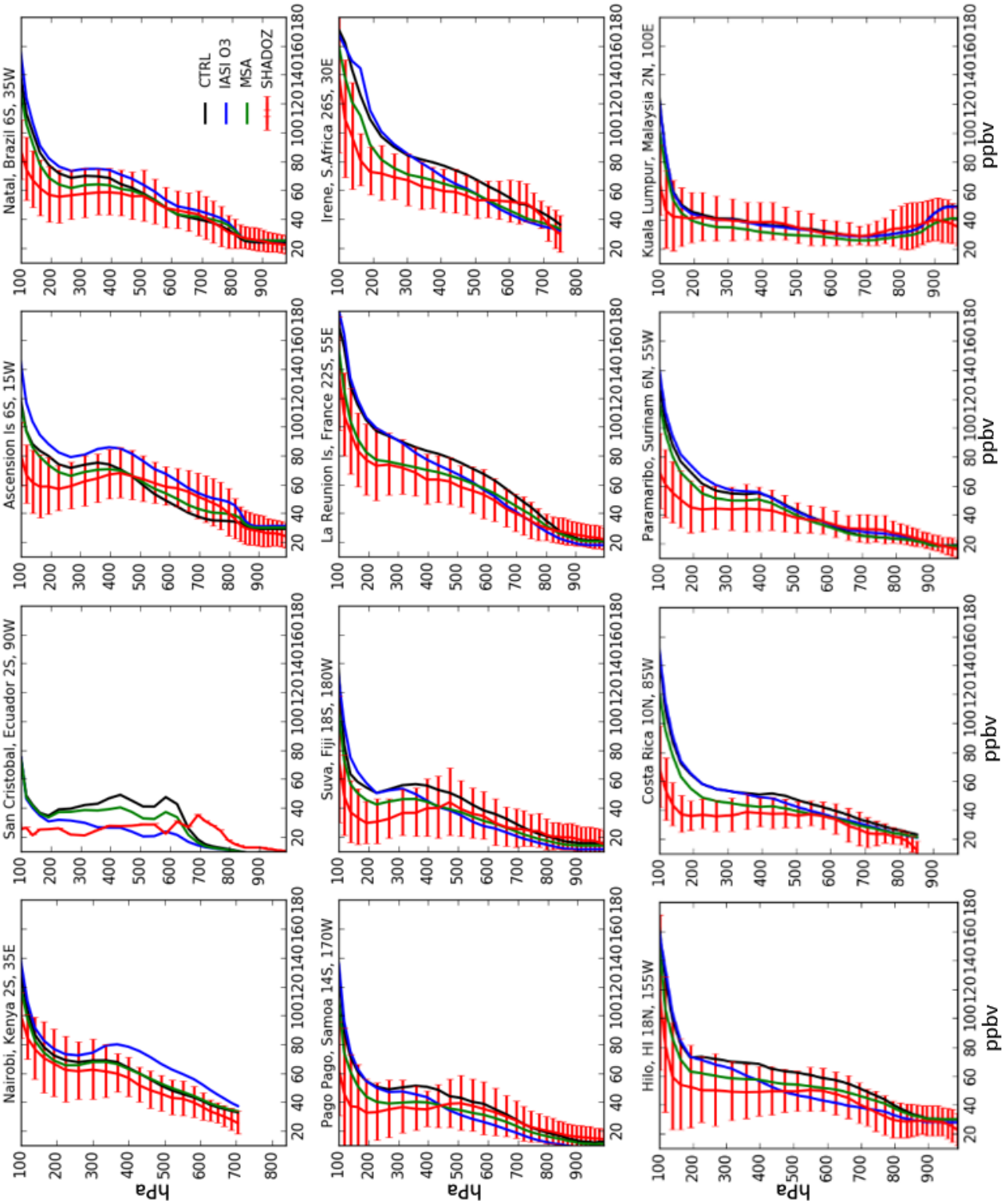

Figure 4.10: Comparisons of ozonesonde observations (red line) from 12 SHADOZ sites with O_3 profiles from the a priori (CTRL) run (black line), the IASI-only assimilation (the blue line), and the MSA-std (green line).

with much smaller changes in magnitude. Moreover, the MSA indicates that the bottom-up emission inventories did not have strong biases in regions such as the southeastern US, Amazonia and southeastern China, which is different from the MOPITT assimilation estimates. This is because the MOPITT assimilation over adjusts the surface emissions to compensate for model biases, such as discrepancies in the modelled OH. In contrast, the biases in the model chemistry are are mitigated in the MSA by assimilating the other species that are chemically related to CO. The reduction in the global OH budget shown in Fig. 4.6 would explain why the MSA CO emission estimates are lower than those from the MOPITT-only assimilation. Moreover, the different CO estimates obtained in the southeastern US and Amazonia are due to the assimilation of OMI HCHO to constrain the biogenic source of CO.

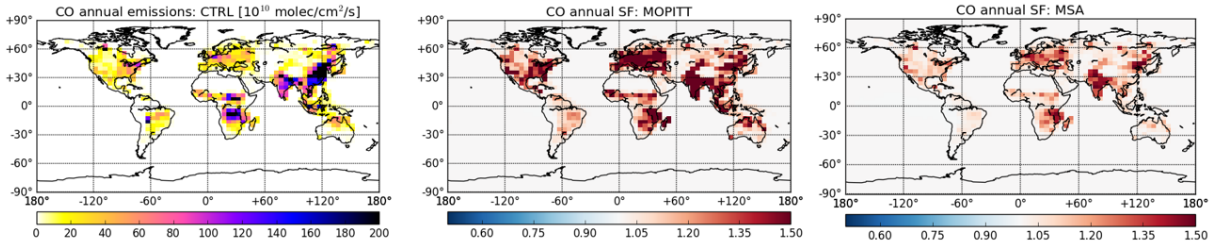

Figure 4.11: Annual mean total CO emissions (in 10^{10} molec/cm^2/s) for CTRL, and CO emission SF (which is the ratio of the a posteriori emissions to the a priori emissions) for MOPITT assimilation and MSA-std. Total emissions include anthropogenic and biomass burning components only.

The global and regional CO emission monthly timeseries for the a priori (CTRL) run, the MOPITT assimilation, and the MSA are shown in Figure 4.12. The focus here is only on the anthropogenic emissions in the US (the contiguous 48 states), Europe, and Asia (including East Asia, India and Southeast Asia) as well as the biomass burning emissions in Amazonia, equatorial Africa, southern Africa, and India/Southeast Asia. The a priori emissions did not show extreme seasonal variations in the northern mid-latitudes. Slightly higher CO emissions are present in DJF in East Asia as well as in JJA in North America. Compared to the MOPITT assimilation, the monthly CO emissions are reduced by 1-2.5 TgCO/month for the MSA, with the seasonal cycle more consistent with *Kopacz et al.* (2010) and *Jiang et al.* (2015a). The European anthropogenic emissions estimated by the MOPITT assimilation have a distinct peak in March and April, which were 2.2 TgCO/month larger than the emission estimates by the MSA. *Jiang et al.* (2015a) argued that the European emissions tended to be confined within the boundary layer and thus the air in the free troposphere over Europe is more chemically aged. As a result, inversions using column or profile CO data (that sample the free troposphere)

to estimate European emissions will be strongly influenced by OH. The MSA European emissions were consistently lower than for the MOPITT assimilation, with the exception of October. For East Asia, both the MOPITT and MSA have show larger CO emissions (by 20-30 TgCO/month) than the a priori in MAM. In other seasons, there was an 8-20% percent increase in anthropogenic CO emissions in both the MOPITT assimilation and the MSA.

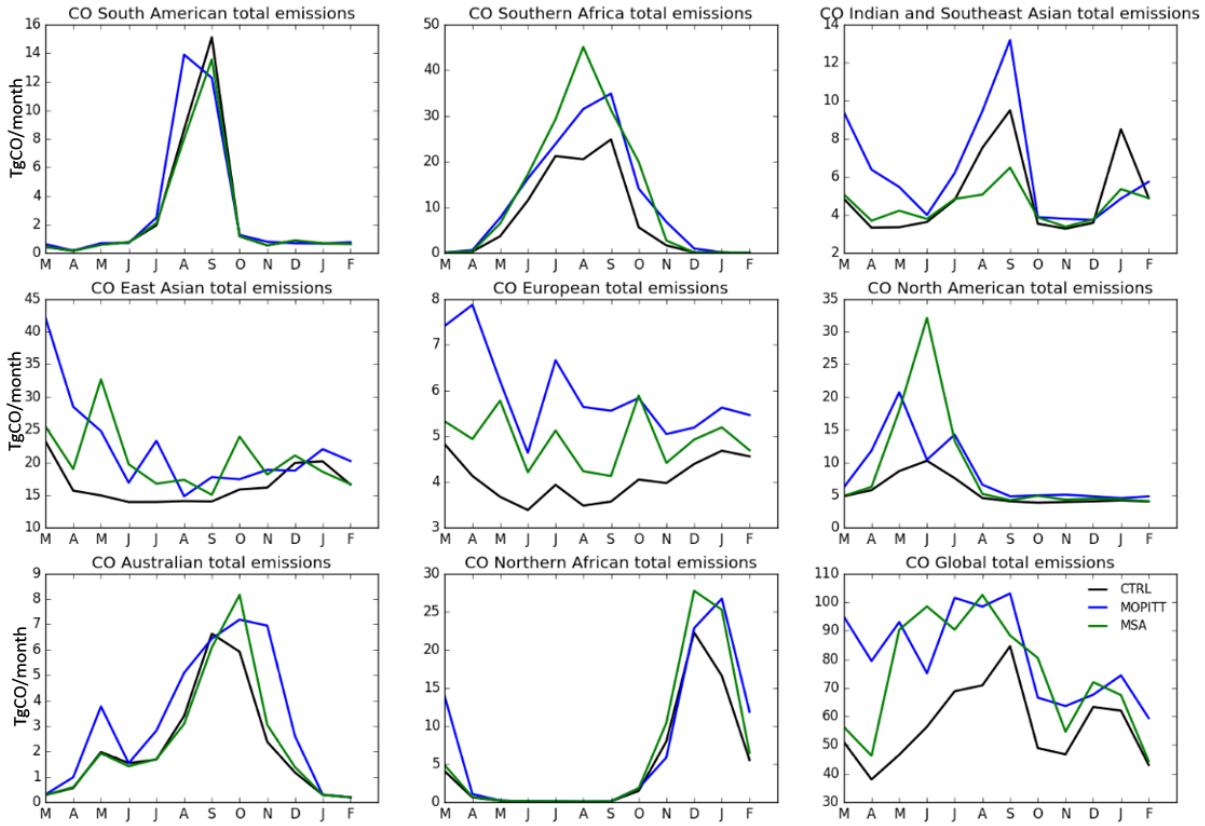

Figure 4.12: Regional monthly total CO emission time series (in TgCO/month) for CTRL, MOPITT assimilation and MSA-std.

For biomass burning emissions, regions such as Amazonia, northern and southern Africa, southeastern Asia and Australia show a much larger seasonal cycle with distinct peaks associated with their wildfire seasons. Both the MOPITT assimilation and the MSA suggest that biomass burning in northern and southern Africa are underestimated during their corresponding wildfire seasons. In both regions, the MSA shifted the peak emission earlier by one month relative to the MOPITT assimilation. In contrast, the South American a priori emissions are consistent with a posteriori estimates. However, the MOPITT assimilation shifted the peak in the emission one month earlier, whereas the MSA has the same timing of the peak in the emissions as the a priori. The region

with the most pronounced change in the emission estimates is southeastern Asia. The MOPITT assimilation suggested that the biomass burning peak in JAS for is significantly underestimated in the a priori. In contrast, the MSA suggested that the peak in the a prior is severely overestimated by 40%. Which scenario is more realistic is difficult to confirm since *Jiang et al.* (2017) showed that the interannual variability of CO emissions over southeastern Asia can vary by more than 200%. According to *Huijnen et al.* (2016) and *Dalsøren et al.* (2009), both the deep convection and the chemical conditions (e.g., the OH abundance) would affect CO inversions in southeastern Asia. Despite that, the overall total biomass burning emission estimates suggest that the wildfire emissions from the GFED-3 emission inventory is underestimated. This was also confirmed by *Giglio et al.* (2013) since they found that the updated biomass burning inventory (GFED-4) showed an increase of 15% of total burned area compared to GFED-3 in the Moderate-resolution imaging spectroradiometer (MODIS) era (after 2000).

The global and regional CO emission statistics for all experiments employed in my study are compared with other recent CO inversion studies (in Table 4.2) that have tried to mitigate the impact of chemical biases associated with OH (e.g., *Müller et al.*, 2018; *Jiang et al.*, 2017; *Miyazaki et al.*, 2015). The annual emission estimates for *Müller et al.* (2018), *Jiang et al.* (2017), and *Miyazaki et al.* (2015) refers to their CO estimates for 2013, 2015, and 2016 respectively. The global annual emissions for the a priori, the MOPITT assimilation, and the MSA-std are 680, 975 and 890 TgCO/year respectively. In fact, the total CO emission estimates from the MOPITT assimilation, the MSA-std, and the MSA-isop are all within the range of the total emissions (831-1009 TgCO/year) estimated by these recent CO inversion studies. In contrast, CO emissions estimated from MSA-γ=1 is much lower (771 TgCO/year) due to the ineffective cost function weighting towards the MOPITT CO observations. It should be noted that the CO emissions reported in *Müller et al.* (2018) are lower partially because they employed a global mean OH (9.3×10^5molec/cm^3) that is 22.5% lower than the OH estimates from the MSA-std. In contrast, the 2016 CO emission estimates from the TCR-2 reanalysis from *Miyazaki et al.* (2015) has larger emission estimates. Regionally, the MSA-std estimates are most similar to those from *Jiang et al.* (2017) who also used the GEOS-Chem model for their inversion analysis, but conducted a two-step assimilation approach to mitigate the impact of OH biases on their inversion.

Table 4.2: Regional surface CO emission estimates for the a priori, MOPITT only, and MSA runs with different setups, compared with three recent CO inverse modelling studies.

	Experiments in MSA studies					Other studies		
Regions	CTRL	MOPITT	MSA-standard	MSA-γ=1	MSA-isop	*Jiang et al.* (2017)	*Müller et al.* (2018)	*Miyazaki et al.* (2015)
				Anthropogenic emissions (TgCO/year)				
US (48 states)	42	55	47	48	45	55	39	65
Europe	46	68	56	49	62	61	30	67
Asia	226	298	264	240	235	260	208	266
Globe	401	562	473	424	443	485	511	663
				Biomass burning emissions (TgCO/year)				
South America	30	34	28	28	26	30	15*	48
Boreal Euroasia	11	20	21	15	18	14	N/A	26
N-Africa	58	88	76	91	68	70	53	62
S-Africa	90	137	153	124	131	119	107	152
India and SE-Asia	21	32	12	14	14	14	10	14
Globe	280	434	417	347	349	358	320	346
				Total Combustion emissions (TgCO/year)				
Globe	681	996	890	771	792	843	831	1009

4.3.6 NO_x emission estimates

Annual mean surface NO_x emission estimates from the a priori as well as the scaling factors for the OMI-only assimilation and the MSA-std assimilation are shown in Figure 4.13. High NO_x emissions mostly come from anthropogenic emission sources in North America, Europe, and East Asia. In comparison, surface NO_x emissions related to wildfires are relatively low, distributed over the Amazon, equatorial Africa, and Indonesia. The OMI NO_2 assimilation suggests anthropogenic NO_x emission inventories over the northern extratropics were significantly overestimated by 30-60%, especially over the eastern US and Europe. Biomass burning NO_x emission inventories for equatorial Africa, India and southeastern Asia were also overestimated by 10-50%. However, these emission changes inferred from the OMI-only assimilation tend to be excessive since the data assimilation system has to correct biases in the modelled concentrations due to discrepancies in the chemistry and emission inventories by adjusting only the surface NO_x emissions. In contrast, the results for the MSA-std assimilation suggest that the overestimation of the bottom-up inventories is much less severe (by 5-20%) for US, Europe and southeastern Asia. The MSA also suggests that the emission inventories over East Asia are slightly underestimated by 10%. Some NO_x emissions associated with Canadian and Russian wildfires are also picked up in the a posteriori emissions for the MSA.

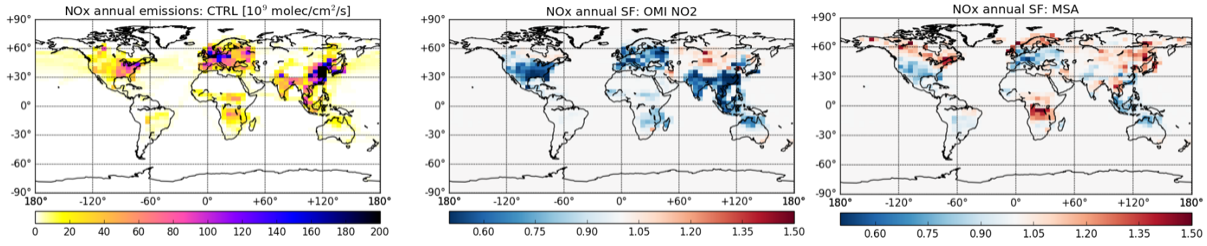

Figure 4.13: Annual mean total surface NO_x emissions for CTRL, and surface NO_x emission scaling factor for OMI NO_2 assimilation and MSA-std. Total surface emissions include anthropogenic and biomass burning components only.

In terms of the global budget, surface NO_x emissions from North America, Europe and East Asia accounts for 14.0%, 17.8% and 31.2% of the total emissions. Biomass burning emissions in boreal forests has increased from 3.7 to 4.7 TgN/year for the MSA. In the southern hemisphere, the a posteriori NO_x emissions for the Amazon, equatorial Africa, and Indonesia were changed by -15.2%, +44.7%, and -27.6% in the MSA compared to the a priori. Overall, the annual mean a posterior NO_x emission using the MSA is 24.9 TgN/year in 2016, which is consistent with the surface NO_x emission estimates of 26.5 TgN/year reported by *Miyazaki et al.* (2015). The annual production of NO_x from

LNO_x in the MSA-std assimilation is 5.9 Tg/year, which is 32.9% lower than the a priori, suggesting that there is a high positive biases in LNO_x in the model. The a posteriori estimates also agree with the LNO_x source estimates by other studies such as *Murray et al.* (2012) (6.0 TgN/year) and *Miyazaki et al.* (2014) (6.4 TgN/year).

4.4 Summary

Multi-species data assimilation experiments were performed using the GEOS-Chem model with a 4D-var data assimilation scheme to optimize the tropospheric CO-HO_x-NO_x-O_3 chemistry throughout 2016. Building on the work in Chapter 3, CO and NO_x emissions, as well as O_3 and HNO_3 initial conditions were optimized by assimilating satellite observations of CO from MOPITT, NO_2 from OMI, HCHO from OMI, O_3 from IASI and OSIRIS, and HNO_3 from MLS. Significant changes in tropospheric O_3 were obtained in the assimilation, mostly due to the influence of the IASI O_3 assimilation. The modelled HCHO response to the OMI HCHO observations was consistent between the OMI-only assimilation (in which I assimilated only HCHO and adjusted isoprene emissions) and the MSA-std (in which I assimilated HCHO but did not optimize isoprene emissions). However, the resulting surface isoprene emissions were largely inconsistent between the OMI-only assimilation and the MSA-isop assimilation (which I also optimized isoprene emissions), due to biases in the isoprene oxidation scheme in the model. The MSA-std produced reductions in OH of to 15% over North America and Europe as well as over regions within the tropics and subtropics. The assimilations were evaluated with independent in-situ aircraft, TCCON, and ozonesondes data. Comparisons with TCCON measurements and ATom aircraft data showed that CO enhancement after the assimilation reduced the negative biases in CO throughout the troposphere. The modelled O_3 evaluation with ATom aircraft data and SHADOZ O_3 ozonesonde data showed that assimilation resulted in improvement of O_3 in the middle and upper troposphere. Furthermore, the OH responses to the assimilation showed significant improvement relative to the OH climatology from *Spivakovsky et al.* (2000). The global tropospheric mean OH was reduced from 13.0 to 12.0×10^5molec/cm^3 which better matched the *Spivakovsky et al.* (2000) estimate of 11.6×10^5 molec/cm^3. The ability to constrain the modelled OH state is a key achievement of the MSA. The full a posteriori OH field can now be used for various CO and CH_4 inverse modelling studies in the future. As a result of the changes in OH, the CO emission estimated using the MSA-std suggested that the bottom-up inventory underestimates CO emissions by 13-30%, much less than underestimate of 43% indicated by the MOPITT-only assimilation. The surface and LNO_x emission estimates

for the MSA were 24.9 and 5.9 TgN/year respectively, both of which agreed well with NO_x inversion studies (e.g., *Miyazaki et al.*, 2015; *Murray et al.*, 2012).

Chapter 5

Estimating NO_x emissions for North America using OmO NO_2 retrievals

5.1 Introduction

Atmospheric NO_x is a major air quality contaminant that adversely impacts human health. It is also a key precursor for tropospheric O_3. Satellite observations of NO_2 have been used extensively to infer surface emissions of NO_x (e.g. *Martin et al.*, 2003; *Jaeglé et al.*, 2005; *Stavrakou and Müller*, 2006; *Miyazaki et al.*, 2012b, 2017). According to bottom-up emission inventories, there has been a significant negative trend in NO_x emissions in North America and Europe during the past two decades, and satellite retrievals of NO_2 indeed show large reductions in NO_2 during this period. However, recently *Jiang et al.* (2018) has suggested that the US Environmental Protection Agency's emissions inventory has overestimated the negative trend in North American NO_x emissions. They found that their top-down emission estimates suggested a reduction in the rate of decrease of North American emissions after 2011, which they attributed to an increasing relative contribution of industrial and off-road diesel emissions (which were not as strictly regulated as on-road gasoline emissions) and a slower than expected reduction in on-road diesel emissions. The work of *Jiang et al.* (2018) highlighted the need to better understand the trends in NO_x emissions, and utility of inverse modelling of space-based measurements of NO_2 to provide top-down constraint on the emissions.

Retrievals of NO_2 from GOME, SCIAMACHY, OMI, and GOME-2 provide a continuous NO_2 observation record over the last 20 years (*Boersma et al.*, 2008). Newly launched instruments such as TROPOMI as well as future instruments such as Tropospheric Emissions: Monitoring POllution (TEMPO) will produce NO_2 measurements

with much higher precision and finer spatial and temporal resolution. However, a challenge with using these data for inverse modelling of NO_x emissions is that the inversion analyses require observations of tropospheric NO_2 abundances, whereas the measurements provide constraints on the total atmospheric column abundance of NO_2. To isolate the tropospheric contribution to the total column, a number of different approaches have been employed. *Martin et al.* (2003) first applied the reference sector method, which assumed that the total NO_2 columns over the remote and unpolluted atmosphere, where tropospheric NO_2 abundances are low, is representative of the stratospheric NO2 columns. *Boersma et al.* (2008) assimilated OMI column data in a CTM to provide an optimized description of the stratospheric NO_2 columns. However, discrepancies in the estimated stratospheric column in each of these approaches could bias the tropospheric residuals. For example, the reference sector approach did not properly account for variability in the stratosphere, and the assimilation approach can introduce model biases in the data product. Recently, *Adams et al.* (2016) used a limb-nadir matching technique with stratospheric NO_2 retrievals from OSIRIS to subtract the stratospheric contribution to the OMI total column. Here I conduct the first inverse modelling analysis of these OSIRIS-minus-OMI (OmO) tropospheric NO_2 columns.

The objective in this chapter is to evaluate the utility of the new OmO retrievals for quantifying surface NO_x emissions. I focus on assimilating the data during 2008-2011 using the nested version of the model for North America. The use of the high-resolution nested grid model will better capture the strong heterogeneity in tropospheric NO_2 and thus reduce the representation error in the model chemistry. To reduce the impact of O_3-related biases on the inversion, I also assimilate IASI O_3 retrievals to create observation-based boundary conditions for the nested assimilations. Note that since the focus of this study is to evaluate the utility of the OmO retrievals for inverse modeling of NO_x emissions, I separately assimilate the O_3 and NO_2 observations, which is different from the MSA approach introduced in Chapters 3 and 4.

5.2 Model setup

5.2.1 Assimilation approach

The strong constraint 4D-var cost function for regional NO_x emission optimization is expressed as:

$$J(\mathbf{p}) = [\mathbf{p} - \mathbf{p}_a]^T \mathbf{B}^{-1} [\mathbf{p} - \mathbf{p}_a] + \sum_{n=1}^{N} [\mathbf{y}_n - H(\mathbf{x}_n)]^T \mathbf{R}^{-1} [\mathbf{y}_n - H(\mathbf{x}_n)], \tag{5.1}$$

where $\mathbf{p}$ represents the surface NO_x emissions, $\mathbf{y}_n$ represents the OmO NO_2 observation being assimilated at a given time n, H is OmO NO_2 observation operator described in Section 2.5, $\mathbf{B}$ is the a priori error covariance matrix that accounts for uncertainty of the initial estimates of surface NO_x emissions, and $\mathbf{R}$ is the observation error covariance. Here $\mathbf{B}$ is specifed with a uniform uncertainty of 50%, which is consistent with recent NO_x inversion studies (e.g., *Travis et al.*, 2016; *Lin et al.*, 2017; *McDonald et al.*, 2018) that have argued that the anthropogenic sources of NO_x are overestimated by as much as 50% over the southeastern US. This setup is also consistent with other regional NO_x inversion studies (e.g., *Qu et al.*, 2017; *Miyazaki et al.*, 2017). In specifying the observation error covariance, I used a similar super-observation approach to that described in Section 4.2. The uncertainty of the tropospheric VCD is $5 \pm 3 \times 10^{14}$ molec/cm^2 for all the assimilated observations over North America, which is consistent with the standard deviation shown in Fig. 2.13. The uncertainty of each OmO observations are computed based on the column abundance with its associated uncertainty from both OMI total columns and OSIRIS stratospheric columns.

To optimize the O_3 boundary conditions I first conducted a global assimilation analysis at a resolution of $4^\circ \times 5^\circ$ using IASI data, and archived the optimized fields around the North American boundary. To obtain a good constraint on the O_3 distribution throughout the analysis period, I used the weak constraint 4D-Var approach, for which cost function is given as:

$$\begin{aligned} J(\mathbf{x}_0, \eta) = {} & [\mathbf{x}_0 - \mathbf{x}_a]^T \mathbf{B}_x^{-1} [\mathbf{x}_0 - \mathbf{x}_a] + \sum_{n=1}^{N} [\mathbf{y}_n - H(\mathbf{x}_n)]^T \mathbf{R}^{-1} [\mathbf{y}_n - H(\mathbf{x}_n)] \\ & + \sum_{n=1}^{N-1} \eta_n^T \mathbf{Q}_\eta^{-1} \eta_n, \end{aligned} \tag{5.2}$$

where $\mathbf{x_0}$ represents the O_3 initial conditions, $\mathbf{y}_n$ represents the IASI O_3 observation

being assimilated at a given time n, H is the IASI O_3 observation operator described in Section 2.5, $\mathbf{B}$ is the a priori error covariance matrix that accounts for uncertainty of the initial O_3 distribution, $\mathbf{R}$ is the observation error covariance matrix, and $\mathbf{Q}$ is the model error covariance matrix, which was discussed in Section 2.4.4.

Here $\mathbf{B_x}$ is specified as a diagonal matrix with a relative error standard deviation of 50%. This assumption provides a loose a priori constraint and allow the O_3 initial conditions to respond more strongly to the observations so that the entire O_3 state across the assimilation window would better match the assimilated observations. A high uncertainty on the initial O_3 distribution is also consistent with other O_3 assimilation studies (e.g., *Miyazaki et al.*, 2012a; *Barré et al.*, 2012; *Emili et al.*, 2014). The observation error covariance $\mathbf{R}$ is specified as in Section 4.2. I assume that $\mathbf{Q} = q^2\mathbf{I}$ conducted a series of parameter tuning experiments to determine q. The O_3 assimilation analysis shown in Chapter 3 suggested that the model can exhibit O_3 biases varying from 0.5 ppbv to 20 ppbv. Hence, I tested values of q of 0.1, 0.5, 3, 6, and 10 ppbv. To test difference choices of q, we conduct some weak-constraint 4D-var based O_3 assimilation over August 2016. The assimilated O_3 fields were then validated against ATom-1 O_3 observations. It should be noted that when q is 10 ppbv, the cost function failed to search any local minima for several iterations. For other four possible values of q (0.1, 0.6, 3 and 6), the assimilated total O_3 burden changed by less than 5%. Compared to the strong-constraint 4D-Var, using $q = 0.5$ provided the most improvement over the northern extratropics when validated against ATom-1 O_3 observations. Hence, I used $q = 0.5$ for the model forcing error error in this study. Ideally though, $\mathbf{Q}$ should be estimated from an ensemble of runs with perturbed dynamics to capture the model error based on the spread of the ensemble. However, such an ensemble is not available at present.

5.2.2 Experiments

As mentioned above, I first optimized the O_3 distribution using the global model at $4°\times5°$ to provide improved boundary conditions for the nested simulation. I then carried out a case study for July 2008 in which I conducted the experiments listed in Table 5.1. I separately assimilated the OMI NO_2 and OmO NO_2 retrievals, and then assimilated the OmO NO_2 data with the IASI O_3 assimilated boundary conditions. The goal of these experiments was to understand the differences between the OMI and OmO assimilations, as well as how the O_3 impacted the optimized NO_x emissions. For the three assimilation runs, the assimilation window was 3 days. This assimilation window is much shorter than that used in the MSA studies in Chapters 3 and 4 to better capture the short lifetime of

Table 5.1: Experiments deployed in the regional NO_x inverse modeling studies.

Experiments	Optimized quantities	Assimilated observation(s)	Assimilated period
CTRL	None	None	Case study: July-2008 Continuous run: JJA 2008-2011
OMI assimilation	Surface NO_x emissions	OMI NO_2 tropospheric SCDs	Case study: July-2008
OmO assimilation	Surface NO_x emissions	OmO NO_2 tropospheric SCDs	Case study: July-2008
OmO+IASI -BC assim.	Surface NO_x emissions $+O_3$ initial conditions	OmO NO_2 and IASI O_3	Case study: July-2008, Continuous run: JJA 2008-2011

NO_x and to take advantage of the high data density of both IASI and OmO observations. I chose the period July 2008 for the case study to so that I can use the ARCTAS aircraft observations to evaluate the modelled NO_2. I also evaluated the model using surface observations from the EPA AQS and NAPS air quality network. After the July 2008 case study, I repeated the experiments for each summer (JJA) in 2008 to 2011.

The initial concentrations for the model on June 1st, at the beginning of the assimilation period each summer, was obtained by running the forward model from August 31st in the previous year without any assimilation. The boundary conditions for each experiment were similarly generated from a forward run of the global model, except for the OmO+IASI BC simulation, in which the O_3 boundary conditions were specified from the global IASI O_3 assimilation.

5.3 Results

5.3.1 O_3 changes in the IASI assimilation

The modeled O_3 changes over North America due to the IASI O_3 assimilation are shown in Fig. 5.1. The main changes are found in middle and upper troposphere. O_3 over northern California and the eastern US is reduced, whereas it is increased in the subtropics and high-latitudes. At the surface, the assimilated results have a similar pattern of changes as in the middle troposphere, but with much smaller magnitudes (less than 2 ppbv), due to downward transport from the free troposphere. Figure 5.1 also shows the altitude-latitude cross section of the a priori O_3 across North America. As can be seen, O_3 is at a maximum at the surface and in the upper troposphere. The surface maximum is due to anthropogenic emissions, whereas the upper tropospheric maximum is due to lightning NO_x emissions (*Parrington et al.*, 2008). The assimilated changes in O_3 suggest that there could be a discrepancy in the vertical distribution of O_3 production from lightning

NO_x emissions, with the model producing too much O_3 in the middle troposphere.

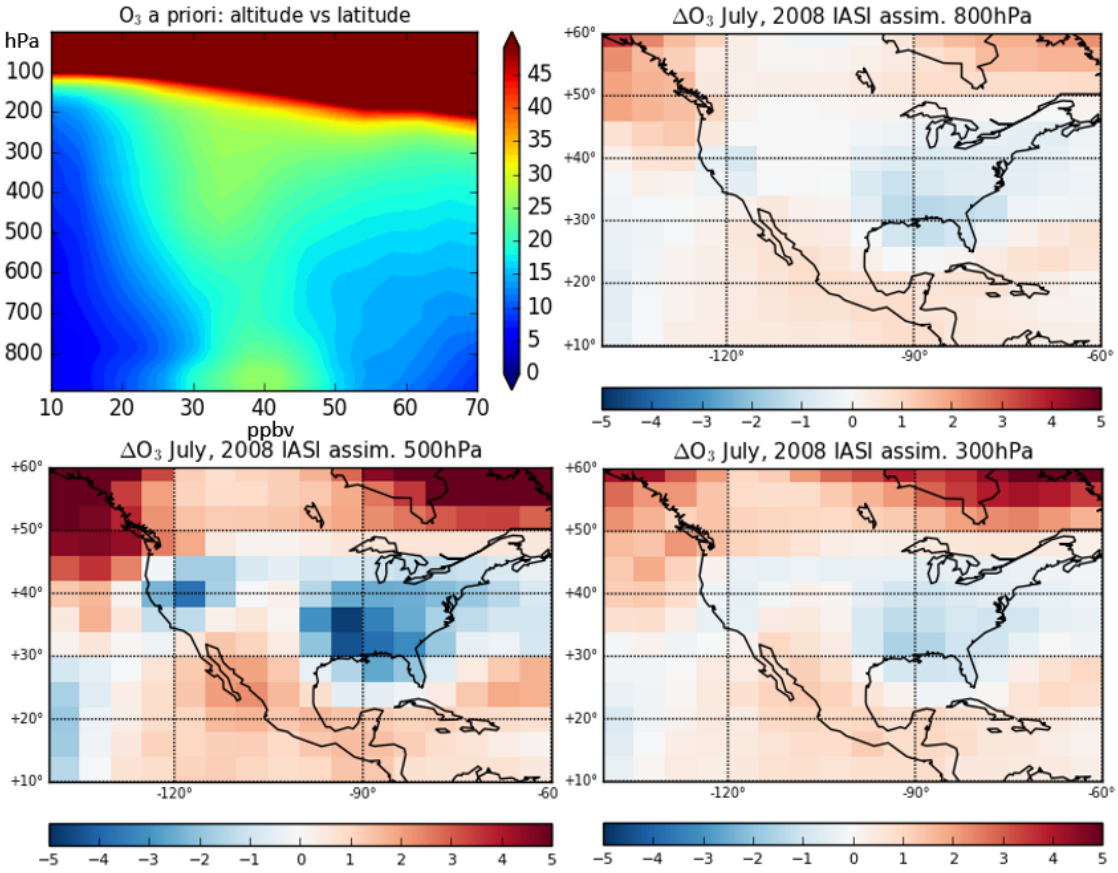

Figure 5.1: The a priori O_3 distribution (top left) and the difference (both in ppbv) between the a posteriori and the a priori O_x at 800 hPa (top right), 500 hPa (bottom left), and 300 hPa (bottom right) for the IASI O_3 assimilation.

5.3.2 Evaluation against OmO NO_2 data

Figure 5.2 compares the modeled NO_2 tropospheric columns with the OmO NO_2 observations for boreal summer in 2008. The a priori NO_2 columns are overestimated with mean differences of $0.18\pm 3.18\times10^{15}$ molec/cm^2. There is significant overestimation over major cities such as Los Angeles, San Francisco, Mexico City, Chicago, and Toronto as well as over larger regions such as the eastern US and Alberta. This overestimation was also seen in the assimilated results shown in Chapters 3 and 4. After the assimilation, the mean difference between the modelled and observed NO_2 columns was decreased to $0.03\pm2.92\times10^{15}$ molec/cm^2, which indicates that the assimilation is fitting the OmO data well. However, there are large residual biases in the mid-western US that the emission optimization cannot remove. This may suggest that the inversion is more effective at

adjusting urban NO_x sources than rural sources. Also, the remaining NO_2 biases could be related to two factors. First, discrepancies in the AMF between the model and observations would result in NO_2 column differences that could be not reduced using emission optimization. Second, lightning NO_x is a major source of NO_x emissions in summer. Since LNO_x was not optimized in these experiments, some of the residual NO_2 biases could be due to high LNO_x emissions. In Chapter 4 I estimated global a posteriori LNO_x emissions that were 33% lower than the a priori. The overestimation of the NO_2 columns in the Midwest coincides with the region of the O_3 reductions in IASI assimilation, suggesting that there could indeed be excessive LNO_x emissions in the middle and upper troposphere over the eastern US.

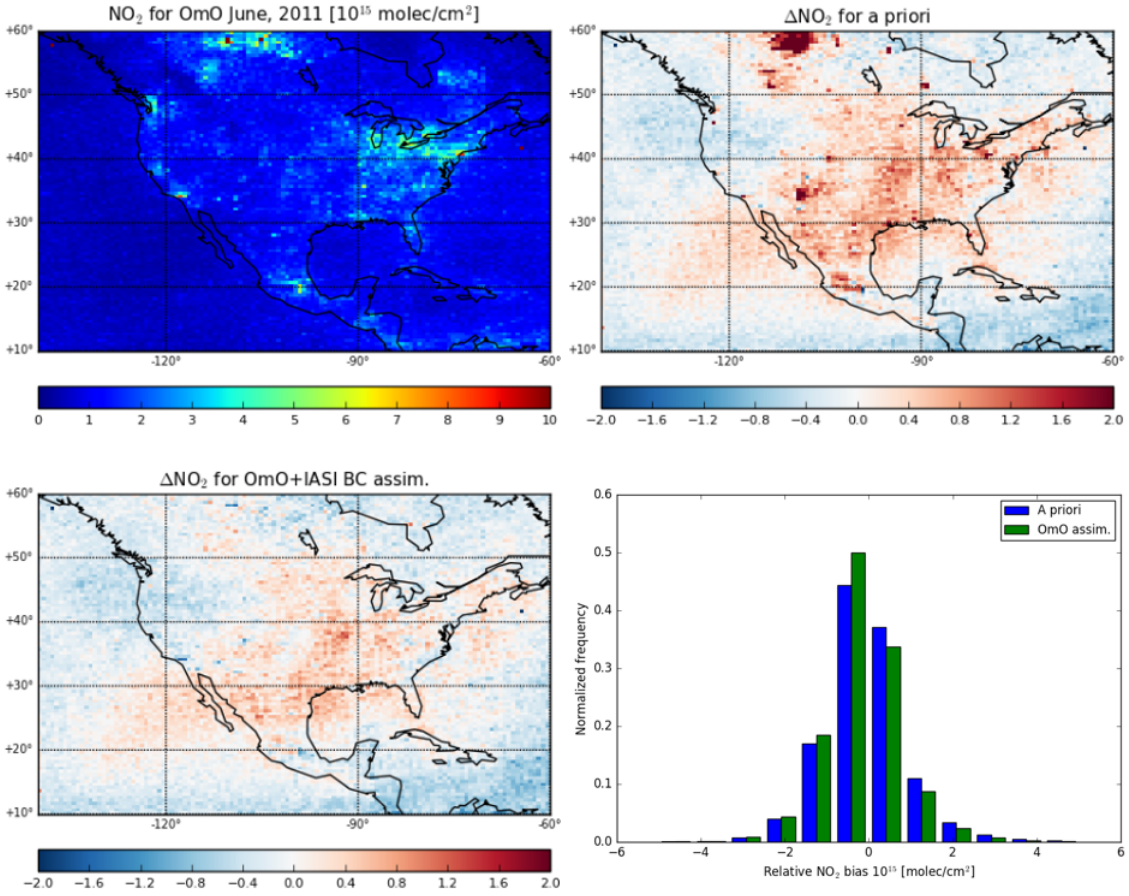

Figure 5.2: Top left: Distribution of the mean NO_2 column abundances (10^{15} molec cm^{-2}) from the OmO retrievals in JJA 2008. Top right: Differences between the a priori model and the OmO retrievals. Bottom left: Differences between the a posteriori NO_2 from the OmO+IASI BC assimilation and the OmO retrievals. Bottom right: Histogram of the modelled minus observation differences (10^{15} molec cm^{-2}) for a priori model (blue bars) and the OmO+IASI BC a posteriori fields (green bars).

5.3.3 Evaluation against ARCTAS NO_2 aircraft data

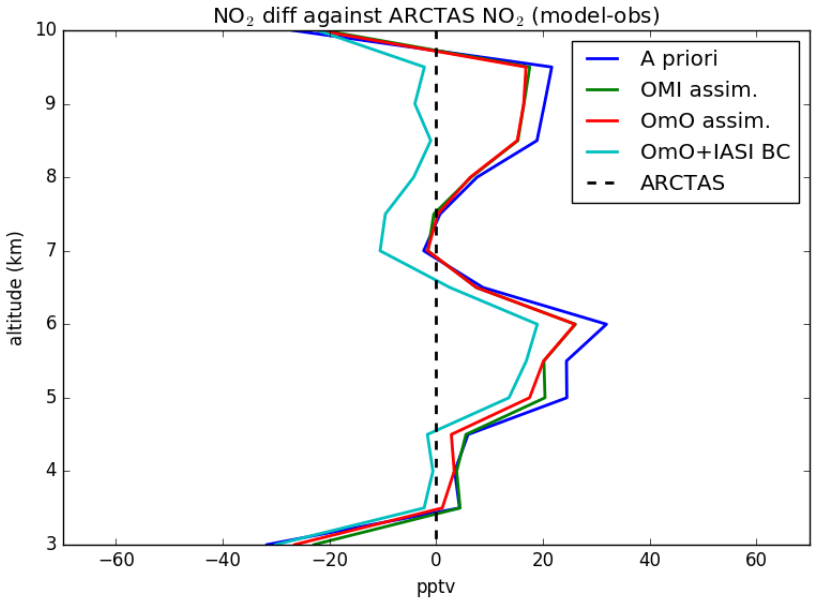

Figure 5.3: Vertical profile of the =mean differences (in pptv) between modelled NO_2 and ARCTAS observations of NO_2. Shown are the mean differences for the a priori (blue line), the OMI assimilation (green line), the OmO assimilation (red line), and the OmO+IASI BC assimilation (aqua line)

Figure 5.3 shows vertical profiles of relative mean differences between the modelled NO_2 and the ARCTAS observations. From the surface to about 3 km, the assimilations did not have much of an impact on the bias relative to the aircraft observations. According to the ARCTAS flight plan, these measurements were made on the outbound flights from Los Angeles, Yellowknife, and Cold Lake. For these observations, both representation error of the modelled NO_2 as well as observation errors could be large (*Miyazaki et al.*, 2017). Hence, I excluded the boundary layer data in the comparisons shown in Fig. 5.3. Between 3-10 km, the a priori NO_2 was biased high by up to 34 pptv. The OMI and OmO assimilation results were similar, and both reduced NO_2 biases above by about 4 pptv in 5 km and between 8-9.5 km. For ARCTAS observations poleward of 60°N, the two assimilations cannot improve the modelled NO_2 significantly because the observations are away from emission source regions. In contrast, the OmO assimilation using assimilated O_3 boundary conditions further reduce the NO_2 bias by about 3 ppbv between 4-6 km, and by 20 pptv between 8-9.5 km. The background NO_2 in the upper troposphere is remarkably improved thanks to the chemical feedback of the assimilated O_3 on NO_2. The relative improvement in the upper troposphere also reflects the fact that IASI tends to be

more sensitive to O_3 abundances in the UTLS, which leads to more effective corrections on upper tropospheric O_3, and hence on modelled upper tropospheric NO_2.

5.3.4 Evaluation against AQS and NAPS O_3 data

Figure 5.4 shows the surface O_3 evaluation against AQS and NAPS observations. Since the purpose of O_3 assimilation is to improve the O_3 fields for boundary conditions in the regional NO_x inversions, I first interpolate the modelled O_3 from the global model to the nested grid and then compared the re-gridded fields at the times and locations closest to the observations. The AQS and NAPS O_3 observations reveal that there is high level of surface O_3 in California, in the Midwest, and in the eastern US. Without O_3 assimilation, the global model overestimates surface O_3 in the Midwest and eastern US by 9.8 and 16.1 ppbv, respectively. These summertime surface O_3 biases were also found in the AQS validation analysis in Chapter 3. Assimilation of the IASI O_3 at $4°\times5°$ reduces the model biases by 2.5 ppbv in the midwest and 2.8 ppbv in the eastern US. The mean model bias across the AQS and NAPS network for the $4°\times5°$ a priori simulation, the a posterior IASI O_3 assimilation, and the a posteriori OmO+IASI BC O_3 assimilation are 13.9, 11.7, and 9.3 ppbv, respectively. When I simulate the surface O_3 using the nested model with the O_3 boundary conditions archived from the IASI assimilation (a priori+IASI BC), the total surface O_3 bias is reduced to 9.9 ppbv. This shows that improving the model resolution alone does mitigate some of modelled biases. But the improvement is modest (with the bias decreasing from 11.7 to 9.9 ppbv). However, the NO_x emission optimization in the OmO+IASI BC assimilation only provides a slight additional reduction in the mean bias (to 9.3 ppbv). This could be due to the fact that the O_3 and NO_2 assimilations are conducted independently as separate optimizations instead of the joint assimilation approaches used in Chapters 3 and 4. The larger reduction in the surface ozone bias obtained in Chapter 3 compared to the results here confirms that jointly assimilating NO_2 and O_3 to exploit the NO_x-O_3 coupling provides a strong constraint on surface ozone.

5.3.5 Validation against AQS and NAPS NO_2 observations

Shown in Fig. 5.5 are the AQS and NAPS and observations of NO_2, gridded on the GEOS-Chem nested grid, and the differences between the model and the observations. The AQS and NAPS NO_2 data are less dense than the O_3. There are NO_2 measurements from AQS+NAPS measurements distributed in California, the midwest, the eastern US, and western Canada. The multi-year mean of the AQS and NAPS measurements shows

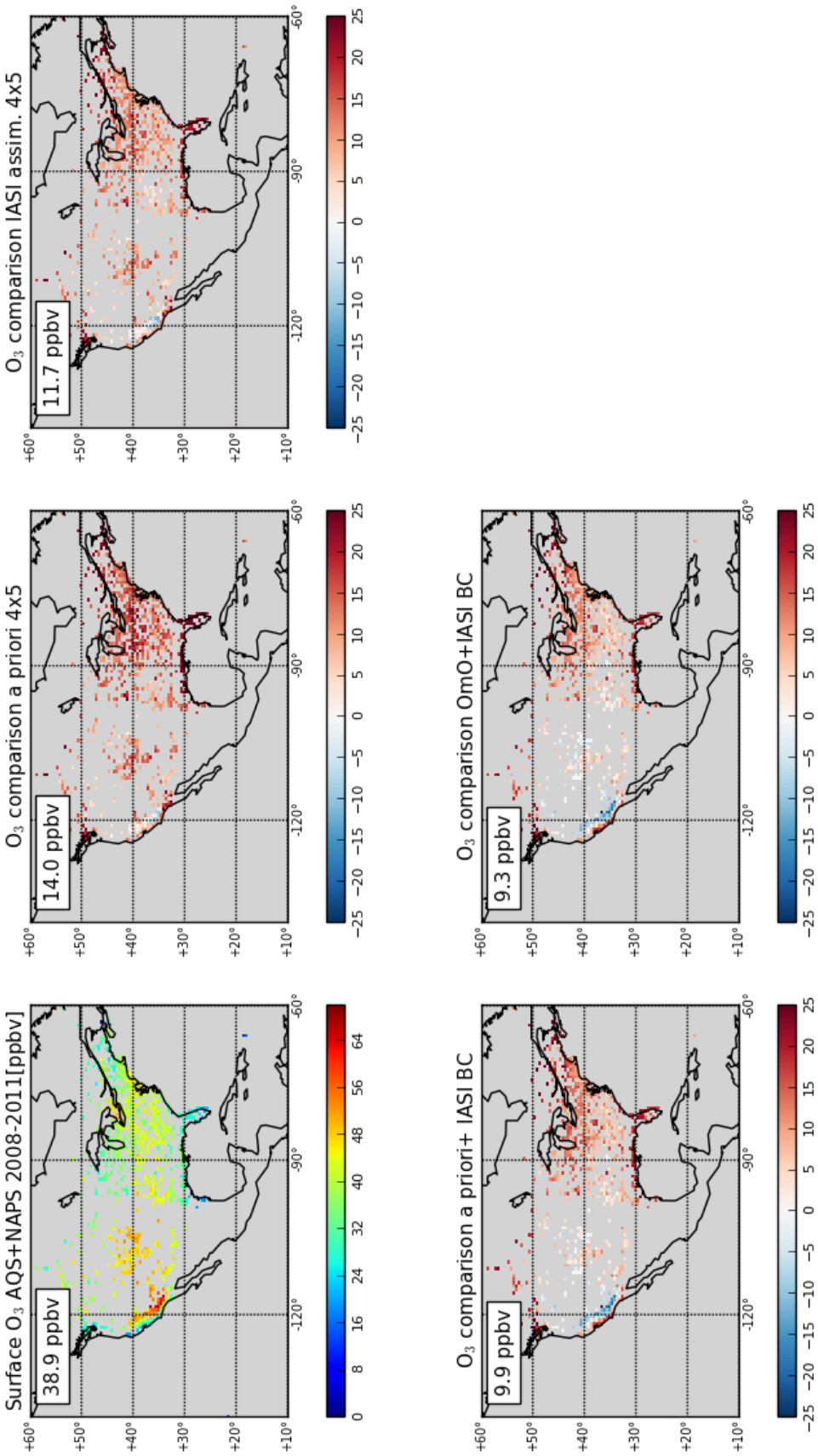

Figure 5.4: Mean summertime AQS and NAPS surface ozone for 2008-2011 (top left) and the absolute mean differences (both in ppbv) between the modelled and AQS and NAPS observations. Shown are the mean differences (in ppbv) for the $4°\times5°$ priori (top middle), the $4°\times5°$ IASI assimilation(top right), the $0.5°\times0.667°$ priori with IASI boundary conditions (bottom left), and the OmO+IASI BC assimilation (bottom right).

high NO_2 concentrations in urban areas such as San Francisco, Los Angeles, Denver, Chicago, and New York as well as in biomass burning regions in Alberta. Many of these locations coincide with high NO_2 tropospheric columns found in Fig. 5.2. This shows that the surface NO_2 observations and the OmO NO_2 tropospheric columns may have consistent information on the NO_x emission hotspots. The a priori NO_2 fields are biased high in the urban areas mentioned above and across the eastern US. There is also significant overestimation in Texas and Louisiana. After the assimilation, the NO_x emission optimization reduces the absolute biases of the surface NO_2 concentrations by 57.1% when compared to AQS and NAPS measurements, with much of the improvement in the eastern US. The mean model biases for the western and eastern US are reduced from 0.7 to -0.3 ppbv and from 0.8 to -0.2 ppbv, respectively. The model also slightly improves the NO_2 fields (by 0.2 ppbv) over the wildfire regions in central Canada. However, the NO_2 biases over rural regions in the Midwest are slightly enhanced. A possible reason for this could be that the representativeness correction factor of 1.7 that was applied to the observations (and which was explained in Section 2.1.2) is too strongly based on urban sampling and, consequently, could result in larger biases in the rural AQS and NAPS observations.

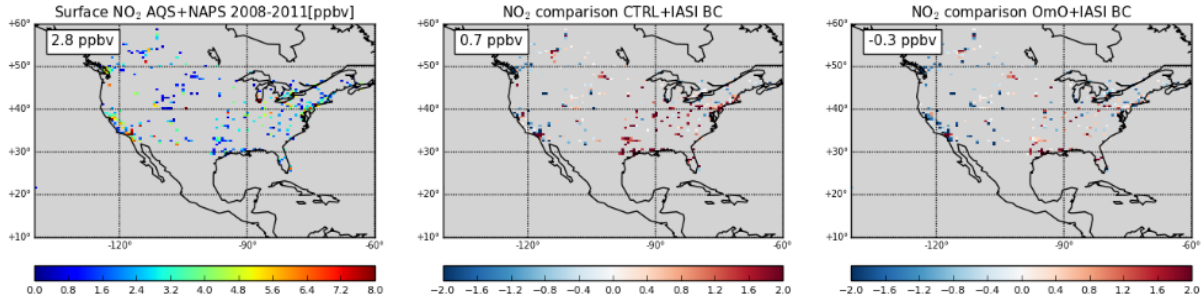

Figure 5.5: Mean AQS and NAPS NO_2 observations for JJA 2008 (left), and mean absolute differences (both in ppbv) between the AQS+NAPS data and the modelled NO_2 for the $0.5° \times 0.667°$ priori simulation (middle) and the OmO+IASI BC assimilation (right).

5.3.6 NO_x emission estimates

Case study: Impact of IASI O_3 assimilation on estimating NO_x emissions for July, 2008

I have compared the bottom-up NO_x emissions with the inferred surface emissions from the OMI and OmO assimilations as well as with the OmO+IASI BC assimilation in

Fig. 5.6. The total a posteriori NO_x emissions for the three experiments are 0.559, 0.52 and 0.547 TgN/month, respectively, compared to the a priori of 0.695 TgN/month. Comparison of the OMI and the OmO experiments indicate larger reductions in emissions across the US and eastern British Columbia in the OmO assimilation. However, as the OmO+IASI BC experiment has effectively reduced the NO_2 biases (shown in Fig. 5.5), its posterior estimates are 5.1% higher than the emission estimates inferred from the OmO assimilation. Unlike the OMI assimilation, for which most of the emission corrections are over urban areas, the OmO and the OmO+IASI BC assimilation produce reductions in both urban and rural areas. Over the eastern US (east of 95°W), the NO_x emissions are 0.281 TgN/month for the OmO+IASI BC assimilation, which is 0.026 TgN/month smaller than the estimates from the OMI assimilation. This is because surface NO_x emissions tend to respond to both the OmO observations and the reduced O_3 near the surface in the eastern US (which was introduced by the assimilated O_3 boundary conditions, shown in 5.1).

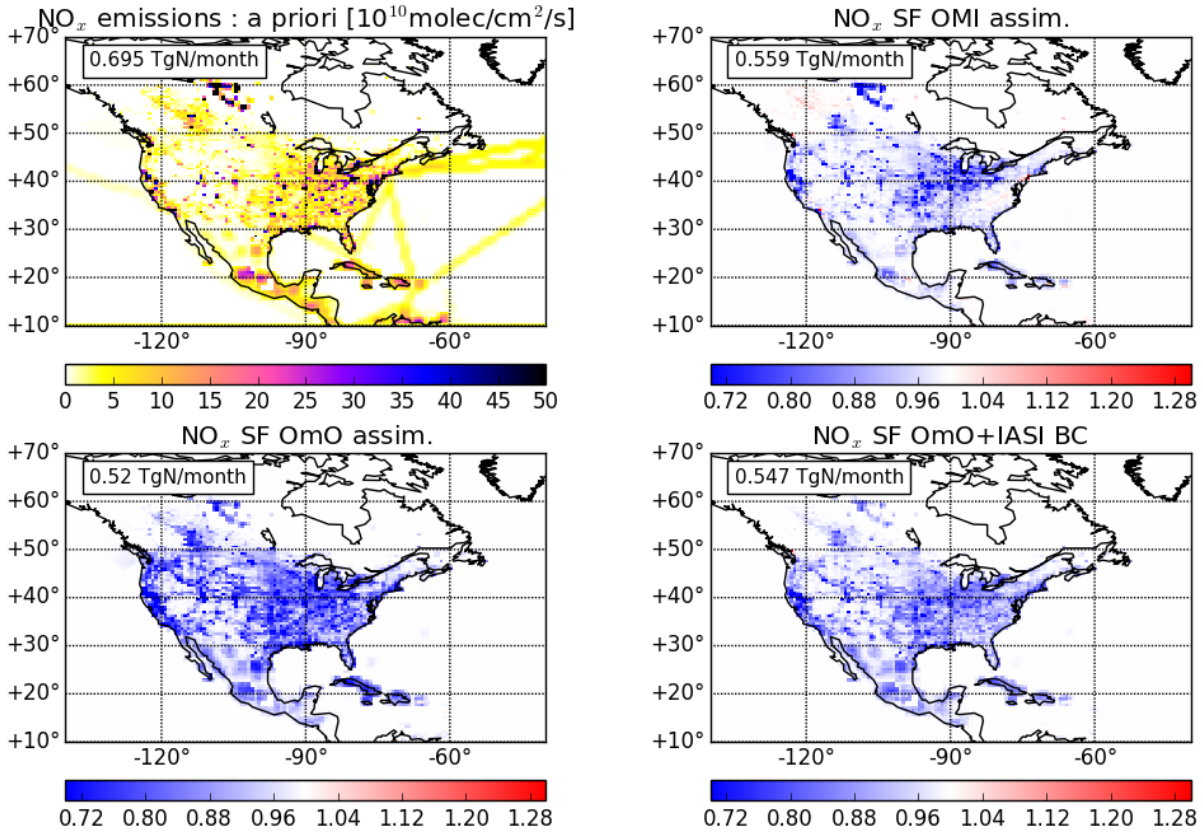

Figure 5.6: The a priori NO_x emissions (in 10^{10} molec/cm^2/s) and the a posteriori NO_x emission scaling factor for OMI, OmO and OmO+IASI BC assimilations for 2008-2011.

Estimated emissions from OmO+IASI BC for JJA 2008-2011

The JJA a priori and a posteriori emissions for 2008-2011 are shown in Figs. 5.7 and 5.8. The four-year mean anthropogenic and biomass burning a priori emissions were 564.8 GgN/month and 55.9 GgN/month, respectively. There are strong emissions (more than 1 GgN/month) from cities on the west coast and east coast of the US, around the Great Lakes, and near Mexico City. In addition, there are large NO_x emissions from power plants and wildfires. For example, there are biomass burning sources in central Canada, which tend to exhibit high year to year variations. Some aircraft and shipping NO_x emissions along the Atlantic are also present. Compared to the a priori emissions, the a posterior emissions have significant reductions over all emission types. Reductions were obtained over remote areas in the southeastern US, the Midwest, and in cities along west and east coast of US. For the OmO+IASI BC assimilation, the anthropogenic and biomass burning emissions were reduced by 44.8% and 26.7%, respectively. In California, reductions of up to 7.7% NO_x were obtained in 2008, with slightly smaller reductions in 2009-2011. In the eastern US, NO_x emissions were reduced by 2.9%, 2.1%, 3.5% and 3.0% in the four years, respectively. The bottom up inventories on Canadian wildfires in 2008, 2010 and 2011 are also overestimated, which are reduced by 1.3%-3.2% after the assimilation.

Table 5.2 shows the JJA mean NO_x emission estimates for North American cities that are large sources of NO_x as well as total North American emissions. Between 2008-2011 the total North American a posteriori emissions were lower than the a priori by 26%, 19%, 36%, 30% in the four years, respectively. In Mexico City, Los Angeles, Houston, Detroit, and New York City, the a posteriori NO_x emissions were 23%, 30%, 56%, 49%, and 41% lower than the priori, respectively, in 2008. The North American a posteriori emissions in 2011 were 13.6% lower than in 2008, which is consistent with the 10.8% reduction between 2008 and 2011 in surface NO_2 concentrations in the AQS observations. Among the four years considered in my analysis, the North American emission estimates in 2010 were the lowest, 6% smaller than NO_x emissions in 2011. The emission reductions relative to the a priori were also the largest in 2010. A possible reason could be the adjoint sensitivity is most effective in responding to the OmO and IASI observations in 2010. Recall that other factors such as meteorology and the chemical coupling of NO_x-related species (e.g. NMVOCs, O_3) could also affect the assimilation. Additional work is needed to determine whether there is increased sensitivity in 2010 due to the observational constraints or the meteorological conditions.

To evaluate the emission estimates, my results are compared with the top-down estimates from the TCR-2 NO_x emission reported by *Miyazaki et al.* (2017) for 2008-2011 as

Table 5.2: JJA emission estimates (in GgN/month) for 2008 - 2011 for North American cities that are large sources of NO_x.

	2008		2009		2010		2011	
Regions	A priori	A posteriori	A priori	A posteriori	A priori	A posteriori	A priori	A posteriori
Detriot	2.42	1.24	2.30	1.33	2.24	0.75	2.24	0.72
Houston	3.05	1.35	2.89	1.31	2.69	0.74	2.69	0.77
Los Angeles	3.43	2.40	3.35	2.07	2.91	2.90	2.91	2.67
Mexico City	4.48	3.44	4.48	3.44	4.48	2.08	4.48	3.92
New York City	2.01	1.21	1.85	1.23	1.68	0.95	1.68	1.00
Philadelphia	2.06	1.08	1.85	1.26	1.71	0.84	1.71	0.61
North America	664.15	485.09	580.09	467.75	637.11	408.71	601.49	419.09

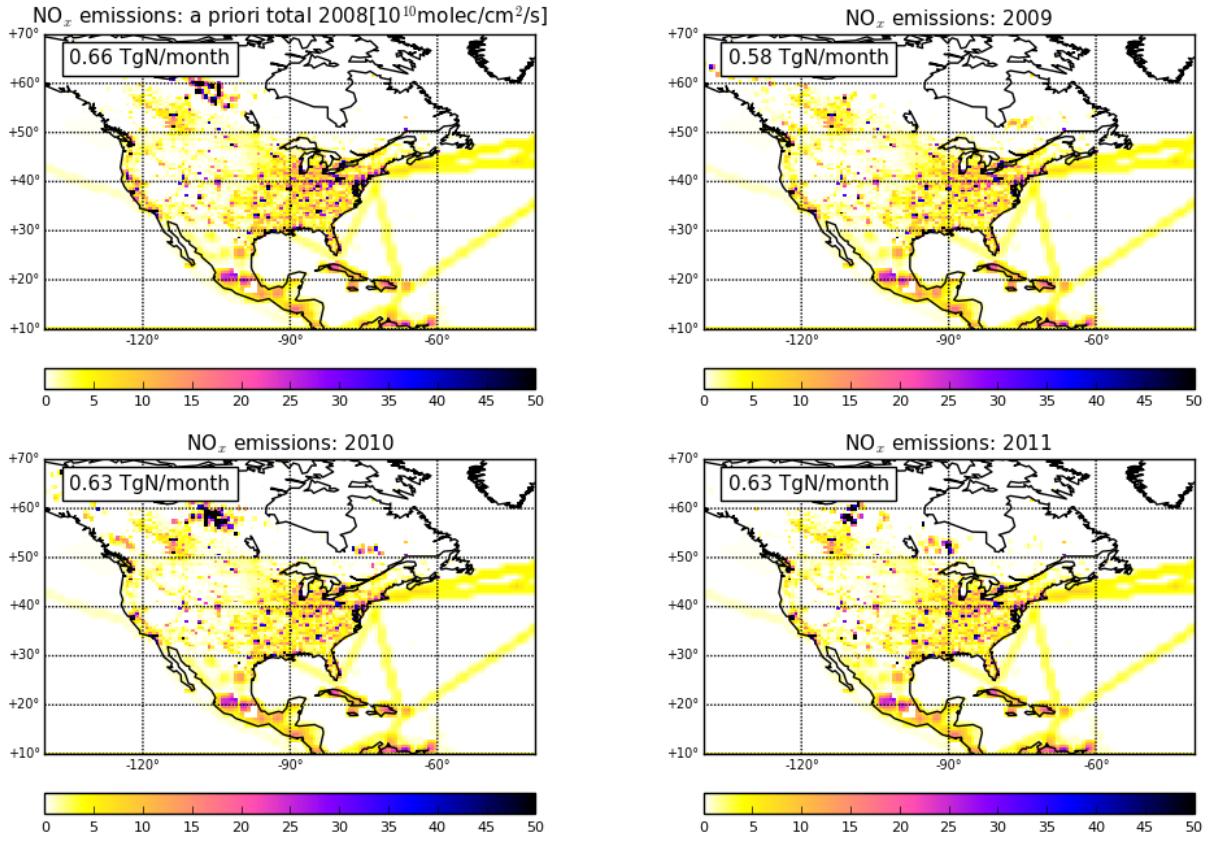

Figure 5.7: The JJA mean prior NO_x emission for 2008-2011 in units of 10^{10} molec cm^{-2} s^{-1}.

well as the emission estimates from the EDGAR-Hemispheric Transport of Air Pollution version 2 (EDGAR-HTAP v2) inventory for 2008 and 2010. My a posteriori estimates of 0.48 and 0.40 TgN/month in 2008 and 2010, respectively, are slightly lower than the *Miyazaki et al.* (2017) estimates of 0.45 and 0.43 TgN/month, but higher than the EDGAR-HTAP estimates of 0.38 and 0.42 TgN/month. For 2009 and 2011, *Miyazaki et al.* (2017) estimated emissions of 0.44 and 0.43 TgN/month, whereas my a posteriori emissions were 0.47 TgN/month and 0.42 TgN/month. *Miyazaki et al.* (2017) also estimated the lowest emissions in 2010. The change of -13.6% in my a posteriori emissions in 2011 relative to 2008 is larger than the -3.5% change reported by *Miyazaki et al.* (2017). However, as discussed above, my estimated reduction in the emissions from 2008 to 2010 is comparable to the -10.8% change in surface NO_2 abundances in the AQS observations. *Jiang et al.* (2018) found that NO_x emissions decreased more slowly after 2011 than predicted by the EPA bottom-up inventory. It would be interesting to extend the analysis beyond 2011 to confirm this slowdown in NO_x emission reduction using OmO data.

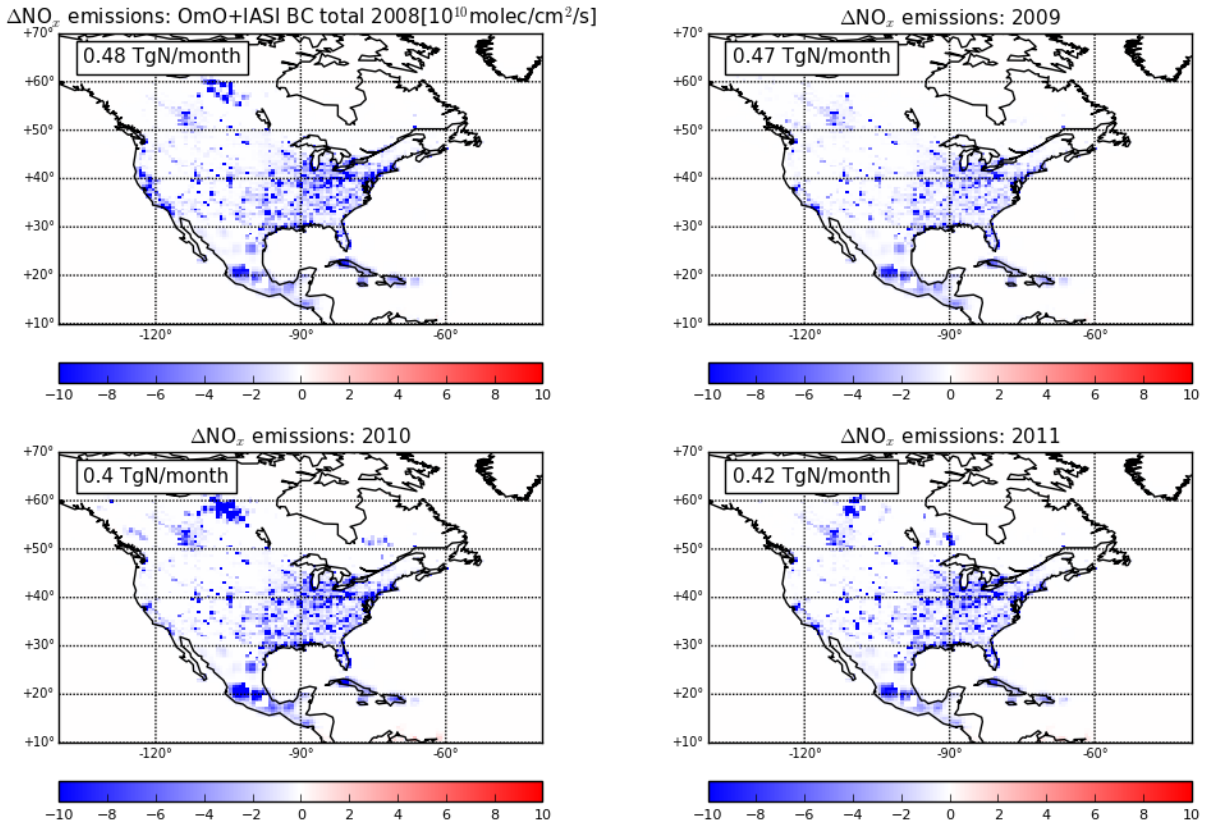

Figure 5.8: The JJA mean posterior NO_x emission changes relative to the a priori for 2008-2011 in units of 10^{10} molec cm^{-2} s^{-1}.

5.4 Discussion

It should be noted that there are several factors that might lead to uncertainties in the a posteriori emission estimates reported here. First, I did not optimize the lightning NO_x emissions my inversion analyses. In the inversions, all of the differences between the modelled NO_2 and the OmO observations were minimized by optimizing only the surface NO_x emissions. In fact, lightning NO_x emissions in summertime also contributed to the total tropospheric column (*Martin et al.*, 2003). However, as discussed in Chapter 3, optimizing surface and lightning NO_x emissions simultaneously using only column NO_2 data is challenging since the inversion cannot distinguish between changes in the column due to changes in lightning NO_x and changes in surface emissions. In Chapters 3 and 4, I had additional chemical constraints (e.g. O_3 and HNO_3) to help discriminate between the two NO_x sources. But in this chapter, because my focus was on assessing the utility of the OmO data for inverse modelling of NO_x emissions, I chose not to conduct a multi-species analysis. Second, the sensitivity of the emissions to the assimilated observations will also

depend on regional meteorology, such as boundary layer mixing, which can impact the vertical profile of NO_2. However, currently this study was unable to quantify this impact on the analysis. Third, although the nested model resolution was much higher than the global model, and that used by *Miyazaki et al.* (2017) and *Jiang et al.* (2018), it was still too coarse for simulating NO_x chemistry. *Valin et al.* (2011) argued that a resolution of 2-12 km is needed to properly capture the nonlinearity in the NO_x chemistry.

5.5 Summary

I estimated surface NO_x emissions in North America by assimilating OmO observations using the GEOS-Chem nested model, constrained with boundary conditions obtained by assimilating IASI O_3 data. Employing the nested-grid model helps to better capture the regional level NO_x emissions. Also, the modelled surface O_3 and NO_2 have significant less bias compared to AQS and NAPS observations when using the nested grid model. The IASI O_3 assimilation produced reductions in O_3 in the middle troposphere over the western and eastern US as well as O_3 increases north of 45°N and south of 30°N. The evaluation of modelled NO_2 with ARCTAS NO_2 observations suggested that the modelled NO_2 in the free troposphere was improved by 5-20 pptv mostly due to the assimilated O_3 fields. In the initial case study for July 2008, the a posterior NO_x emissions inferred from separately assimilating OMI and OmO data were 0.575 and 0.565 TgN/month, respectively, suggesting good agreement between the two data sets on estimating NO_x emissions. The emissions estimated by assimilated by OmO data (OmO+IASI BC) with optimized O_3 boundary conditions were 0.547 TgN/month. Compared with the OmO assimilation, the OmO+IASI BC assimilation produced greater reduced in emissions in rural regions of eastern US as well as in Alberta.

For the 2008-2011 period, the total top-down anthropogenic and biomass burning emissions were reduced by 44.8% and 26.7%, respectively, compared to the a priori. Examination of the cities with large emissions of NO_x showed that the emissions from all of the cities were overestimated in the bottom-up inventory. The most significant reductions were estimated for Houston and Philadelphia, whereas much smaller reductions were obtained for Los Angeles and Mexico City. These top down emission estimates suggest that federal and local emission control policies had a remarkable impact on reducing urban NO_x emissions during the study period. The total North American NO_x emissions were found to be 13% less in 2011 than in 2008. The is consistent with the decreasing trend seen in the surface AQS NO_2 measurements. My results demonstrate that the OmO retrievals can provide reliable constraints on surface NO_x emissions.

Chapter 6

Final Remarks

6.1 Key findings

- Chapter 3
  - Multi-species data assimilation (DA) provides a consistent chemical state over the assimilation window.
  - Multi-species DA provides a better fit to MOPITT CO assimilation due to reductions in the chemical biases in the model. In contrast, multi-species DA fits OMI NO_2 data more poorly than assimilating only NO_2 data because of the strong coupling between NO_2 and O_3.
- Chapter 4
  - Multi-species DA provides a means of indirectly constraining tropospheric OH. The constraints in the extratropics are stronger in summer than in winter, reflecting the stronger chemical coupling in summer.
  - Bottom-up CO emission inventories were underestimated in the northern extratropics in 2016, whereas the bottom-up NO_x emission inventories were overestimated in the northern extratropics.
- Chapter 5
  - Utilizing assimilated O_3 boundary conditions improves the NO_2 fields in the North American middle and upper troposphere in the nested GEOS-Chem simulation.
  - OmO tropospheric NO_2 column retrievals provide constraints on regional NO_x emission estimates that are consistent with those obtained from OMI NO_2 tropospheric NO_2 columns.

6.2 Conclusion

The focus of the thesis was to implement a multi-species chemical data assimilation framework to constrain tropospheric O_3-HO_x-NO_x-hydrocarbon chemistry, with the goal of obtaining improved top-down CO and NO_x emission estimates. In Chapter 3, I conducted a case study for November 2009 and July 2010 in which I assimilated MOPITT CO, TES and OSIRIS O_3, and OMI NO_2 observations to optimized CO and NO_x emissions as well as O_3 initial conditions. Assimilation of TES and OSIRIS O_3 observations reduced the modelled O_3 bias when evaluated against HIPPO O_3 aircraft data. Between the two O_3 data sets, TES O_3 retrievals were more impactful in improving O_3 within the tropical middle and lower troposphere, whereas the OSIRIS O_3 retrievals were more effective at improving O_3 in the UTLS. I found that when combining TES and OSIRIS data, the dominant contribution to the cost function minimization came from the OSIRIS O_3 data due to the strong adjoint sensitivity in the UTLS where the modelled O_3 biases were most pronounced. In the assimilation using MOPITT CO, TES O_3, OMI NO_2 and OSIRIS O_3, the mean posteriori bias was -1.9 ppbv, 33% less than the a priori biases. Neither TES nor OSIRIS data resulted in changes in surface O_3 in the MSA framework. Instead, surface O_3 in the MSA was improved through the chemical responses to the adjusted CO and NO_x emissions, demonstrating the benefit of the MSA framework.

I found that assimilation of MOPITT data in the the MSA approach provided a better fit to the MOPITT data than assimilating only MOPITT CO. In the MOPITT-only assimilation the model is incapable of adjusting the surface emissions to fully minimize the discrepancies between the model and the observations that arise from errors in the surface emissions as well as in the model chemistry. However, because the MSA is able to reduce the chemical biases (e.g., the OH biases), the assimilation is able to better fit the MOPITT data. In contrast, I found that the NO_2-only assimilation provided a better fit to the OMI NO_2 data. This is because the NO_x emission optimization responded to not only the NO_2 observations but also to the O_3 observations in the MSA, as a result of the strong NO_x-O_3 chemical coupling in the model. The value of the MSA approach is that it exploits the chemical coupling in the model to produce a modelled state that best matches all of the available observations. It allows the observations to impact the unobserved species in the model according to the chemical mechanism. As such, I also found that impact of the MSA on tropospheric OH in the extratropics was larger July 2010 than in November 2010, reflecting the stronger chemical coupling in the chemistry in summer than in winter. I also investigated the impact of each type of observations to the assimilated results. I found that if the adjoint forcing from different species cannot

contribute comparably to the total cost function, the observations with the maximum adjoint sensitivity (UTLS O_3 in my case) would dominate the assimilation framework which leads to all O_3-related tracers responding to O_3 changes rather than responding to their corresponding observations. Such sensitivity among different species would also vary by seasons. The strength of chemical coupling reflects the photochemical lifetime of the species and has an impact on the analysis. For instance, the tropospheric OH in July 2010 over northern extratropics was more sensitive to assimilated observations than in November 2009.

In Chapter 4, I extended the multi-species data assimilation framework used in Chapter 3 to include observations of HNO_3 and HCHO. Observations of CO from MOPITT CO, NO_2 and HCHO from OMI, O_3 from IASI and OSIRIS, and HNO_3 from MLS were assimilated throughout 2016 and the first two months of 2017. Evaluation of the assimilated CO with TCCON data showed that the assimilation significantly reduced the model underestimation of CO abundances in the northern hemisphere. For example, over North America the mean a priori XCO was 69 ppbv, whereas for TCCON it was 86 ppbv. The MOPITT-only assimilation and the MSA increased the mean XCO to 83 ppbv and 84 ppbv, respectively. Similarly, over Europe the mean a posteriori XCO was 89 ppbv for both the MOPITT-only assimilation and the MSA, which matched the TCCON measurements. However, I found that there large residual biases in CO relative to the ATom-1 and ATom-2 observations over the Atlantic and Pacific oceans. Evaluation of the O_3 against ATom and SHADOZ O_3 data showed that the MSA reduced the positive bias in tropical O_3 in the middle and upper troposphere. Unlike Chapter 4 in which the OSIRIS O_3 data dominated the cost function minimization, the IASI data a greater impact in the assimilation than OSIRIS. This is likely due to the much greater density of IASI data than OSIRIS data.

A key objective of the work was to improve tropospheric OH. I found that the global mean tropospheric OH for the MSA was reduced from 13.0 to 12.0×10^5molec/cm^3, which is closer to the climatological mean of 11.6×10^5molec/cm^3 from *Spivakovsky et al.* (2000). The study estimated a mean methane lifetime of 7.8 years with the a priori model, which increased to 8.68 years with the MSA a posteriori fields. The reductions in OH in the northern extratropics were mostly driven by the chemical feedbacks associated with changes in CO and NO_2 in the assimilation. In the tropics, the OH reductions were mostly driven by changes in O_3 in the assimilation. Both hemispheres experience OH reductions such that the OH inter-hemispheric ratio did not change significantly as a result of the assimilation. The a prior NH/SH ratio was 1.21 and the a posteriori was 1.23. Both the a prior and the a posterior is still different from the *Spivakovsky et al.*

(2000)'s climatology. Furthermore, *Müller et al.* (2018) have argued that the ratio should be as low as 0.85, which would be inconsistent with the constraints obtained in the MSA.

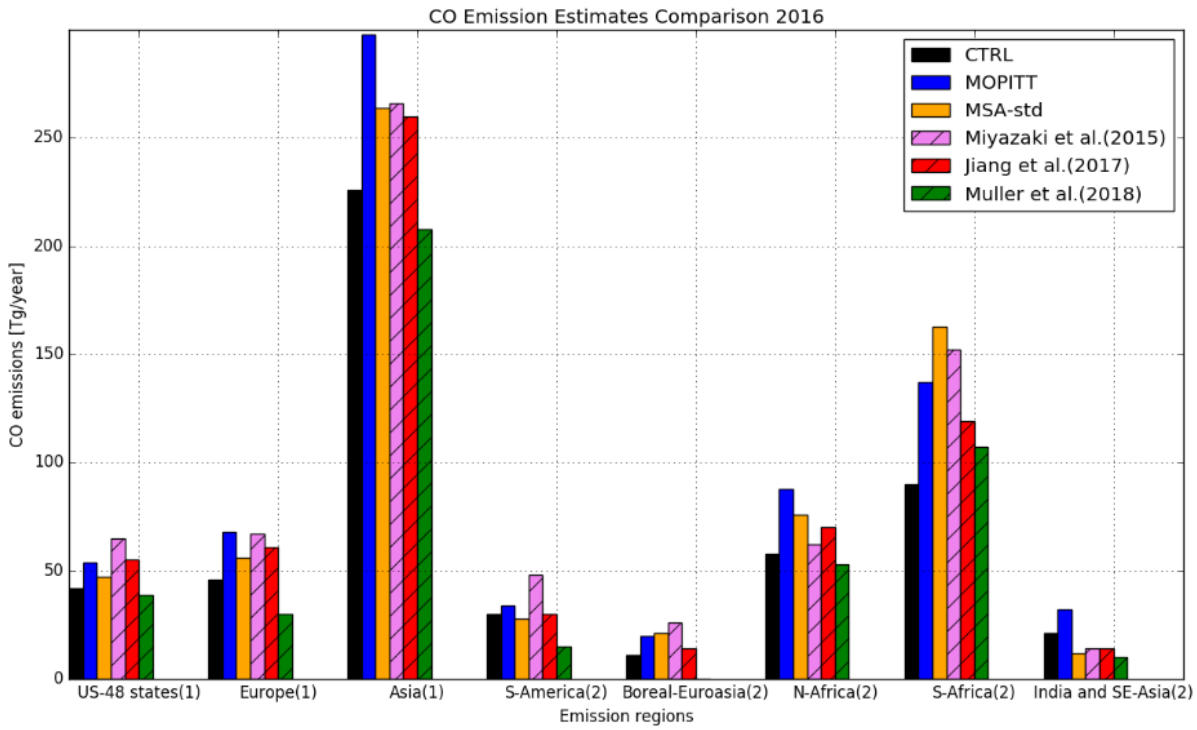

Figure 6.1: Annual CO emissions (in TgCO/month) for CTRL, MOPITT assimilation, and MSA-std (in solid bars), compared to CO emission estimates from TCR-2 data based on *Miyazaki et al.* (2015), *Jiang et al.* (2017) and *Müller et al.* (2018) (in dashed bars). The index in brackets suggests emission types (1: fossil fuel emissions, 2: biomass burning emissions)

As shown in Fig. 5.1, the global CO emission estimates from the a priori and the standard MSA assimilation were 280 and 417 TgCO/year, respectively. The a priori inventories were underestimated in all major emission source regions, except for India and Southeast Asia. Regionally, the MSA emission estimates were consistent with the estimates from *Jiang et al.* (2017), who assimilated MOPITT into the GEOS-Chem model using a two-step approach to reduce the impact of discrepancies in OH on their analysis. They assimilated MOPITT data over the oceans to provide boundary conditions for the continental regions, and then assimilated the data over the continents to quantify the CO sources, with the assumption that the MOPITT data over the continents are sampling fresh CO emissions with little chemical aging due to reaction with OH. My results are also generally consistent with the emission estimates from *Miyazaki et al.* (2015), who used an ensemble Kalman filter to conduct a MSA. Overall, the top down CO emission estimates in my study have shown good agreement with recent CO inverse modelling

studies that tried to mitigate biases in OH (e.g., *Miyazaki et al.*, 2015; *Jiang et al.*, 2017; *Müller et al.*, 2018) .

In Chapter 5 I investigated the potential of using OmO NO_2 retrievals to quantify North American NO_x emissions using the nested version of the GEOS-Chem model. I also assimilated IASI O_3 data to produce optimized boundary conditions for the regional model and found that the O_3 assimilation resulted in large reductions in O_3 abundances in the middle troposphere over the eastern US, and increases in O_3 north of 50°N and south of 25°N. I found that assimilating the OmO data with the optimized boundary conditions for O_3 resulted in improvements in modelled NO_2 in the middle and upper troposphere, relative to ARCTAS aircraft data. For the case study (July 2008), the largest discrepancies between the OmO and the OmO+IASI BC experiments occurred in rural regions. In the eastern US, the emission estimates from OmO+IASI BC assimilation are 5.3% higher than from the OmO assimilation. This is consistent with the 6% difference found in the surface O_3 abundance between the two experiments, suggesting that the additional NO_x emission adjustment in the OmO+IASI BC assimilation was related to surface O_3 changes from the IASI O_3 assimilation.

I quantified summertime (JJA) NO_x emissions for 2008-2011 and found that the bottom-up inventories overestimated surface NO_x emissions by 27-36%. The assimilation suggested significant reductions in emissions from major urban regions, such as Los Angeles, Houston, and New York, that are large sources of NOx. These emission statistics suggested the bottom up inventories were overestimating North American surface NO_x emissions. The inferred emissions in 2008 and 2010 were within the range between the top-down estimates from *Miyazaki et al.* (2017) and bottom-up estimates from EDGAR-HTAP v2. I also found that the total North American emissions in 2011 were 13.6% lower than the emission in 2008, which was similar to the reduction in surface NO_2 abundances measured by the AQS network.

Overall, the results in Chapters 3 and 4 revealed that the MSA approach can mitigate model biases in tropospheric CO, OH, NO_2 and O_3 modelled states, which lead to improved regional CO and NO_x emission estimates. The MSA approach is a powerful means of integrating the growing suite of space-based atmospheric composition measurements to obtain more reliable air quality simulations. The results in Chapter 5 showed that the new OmO retrievals can provide constraints on regional NO_x emissions in an inverse modelling context. The OmO data represent a data-driven means of separating the stratospheric contribution to the total NO_2 columns without relying on models or ad hoc assumptions about the variability of the stratospheric NO_2 columns. The utility of the OmO data shown here highlights the complementarity of space-based limb profile data

with nadir observations to better quantify the changing composition of the troposphere.

6.3 Thesis implications and suggested future work

6.3.1 Implications on CTMs and MSA

As mentioned in Chapters 3 and 4, the modelled global mean tropospheric OH was biased high. The MSA using 4D-var was able to reduce the bias in global mean OH, but could not effectively adjust the hemispheric ratio of OH. Since neither *Miyazaki et al.* (2015) nor my study was able to produce a NH/SH ratio close to unity, it is possible that there are still unknown mechanisms that are neither recognized by the laboratory-based studies nor implemented in atmospheric models that play an important role in OH chemistry. It is especially difficult to see how it would be possible for the models to achieve the low NH/SH ratio of 0.85 that was suggested by *Müller et al.* (2018). Validation against independent observations in Chapter 4 revealed some remaining biases in CO and O_3 within the tropics and subtropics of the SH. These biases are due to either ineffectiveness of the current optimization scheme or biases in model transport (convective and interhemispheric transport). The weak-constraint 4D-Var scheme could be used in the MSA approach to mitigate for these biases. However, the weak-constraint scheme will further increase the complexity of the MSA framework, which will also increase the computational cost.

In Chapter 4, the assimilated HCHO fields did not lead to consistent changes in isoprene emissions between the HCHO-only assimilation and the MSA due to biases in the isoprene oxidation scheme in the model. Recently, an adjoint of the updated isoprene chemistry was implemented in GEOS-Chem by *Kaiser et al.* (2018). However, it was a non-standard implementation, which I could not use for the analyses in this thesis. A key task for future work would be to port the updates from *Kaiser et al.* (2018) into the standard version of the adjoint model and repeat the analysis conducted here, assimilating O_3, NO_2, HCHO, and CO data. Furthermore, I believe that this future analysis should focus on North America using the nested model, building on the work done in this Chapter 5. A North American focus would be ideal because the southeastern United States is associated with high NMVOC emissions and has been extensively studied, consequently there is a wealth of independent observations available to evaluate the assimilation. The work in Chapter 4 highlighted the issues with the old isoprene oxidation scheme. This future analysis will help evaluate the new isoprene oxidation scheme in the context of the multi-species constraint on the chemistry.

When observations of various species were ingested in the assimilation the total cost function tends to be dominated by certain species whose observations exhibit high accuracy or high adjoint sensitivity. This could lead to unrealistic adjustments of some species due to chemical feedbacks linking those species to the observations. Indeed, the resulting changes in the species could be inconsistent with the individual observations of those species. To mitigate this problem requires appropriately balancing the contributions of the different observations in the cost function or imposing a more sophisticated background error covariance. However, achieving these is not trivial. In my current framework, I used an ad hoc approach of scaling the various terms in the cost function to produce CO and NO_x emission estimates that are consistent with the estimates obtained from the single instrument assimilation runs with an idealized error covariance. However, it is desirable to obtain an approach that does not require ad hoc assumptions. One possible approach would be iteratively minimize the cost function according to the lifetime of the species in the state vector. In this framework, I would first optimize the shorter-lived species (e.g. NO_2 and HCHO) and then use that information in the optimization of the longer-lived species (e.g., CO and O_3). Thus, each iteration of the current MSA would involve two sub-loops, in which I separately adjust the shorter-lived and longer-lived species. This approach would increase the computation cost of the assimilation, but it may alleviate the need for arbitrary scaling of the cost function. This approach would also allow us to use shorter assimilation windows for the shorter-lived species, which would help improve the quality of the assimilation. Another possible approach is to enable state uncertainties using initialized ensemble states. Future MSA systems should use advanced data assimilation techniques such as a hybrid 4D-EnVar to allow the propagation of the background error covariance in order to account for the spatial and temporal correlations in the MSA system.

6.3.2 4D-Var vs EnkF data assimilation using GEOS-Chem on modeled NO_x and O_3 abundances

The multi-species data assimilation in Chapters 3 and 4 showed the capability of the approach for improving the modelled troposphere by taking advantage of a range of observations of various chemical species. *Miyazaki et al.* (2012a) were the first to conduct a multi-species assimilation using a suite of satellite observations. They used an EnKF data assimilation scheme with the CHASER-DAS CTM. Recently, Kazuyuki Miyazaki has developed an LETKF data assimilation framework for the GEOS-Chem model, which enables direct comparisons for the LETKF and 4D-Var data assimilation schemes. The

EnKF data assimilation scheme updates the state vector in a sequential manner within a much shorter assimilation window (e.g., 90 min in *Miyazaki et al.* (2012a)). This imposes a strong localization in space and time on the modelled state. In contrast, the 4D-Var assimilation scheme acts as a smoother, which minimize the cost function to better match all of the observations in space and time over the assimilation window. In the future, it would be useful to conduct a comparison between the LETKF and 4D-Var assimilation schemes to determine the strengths and limitations of the two approaches for multi-species data assimilation. Key areas of focus should be on 1) the impact of the length of the assimilation window in the 4D-Var scheme, 2) the impact of the localization length scale in the EnKF, and 3) the impact of the choice of the a priori error covariance matrices in the 4D-Var scheme.

6.3.3 Future MSA in the geostationary satellite era

In the next decades, multiple geostationary satellite (GEO) instruments such as TEMPO for North America, Sentinel-4 for Europe, and the Geostationary Environmental Monitoring Spectrometer (GEMS) for Asia will be available to provide O_3, NO_2, and HCHO measurements with increased spatial and temporal sampling. These GEO instruments will revolutionize the global monitoring system for air quality. They are capable of providing more continuous daytime observations with spatial resolution less than 5 km in contrast to at best twice per day for low Earth orbiting satellites. The MSA approach offers the best means of exploiting these data and integrating them with data from the existing low Earth orbiting satellites to provide vastly improved constraints on air quality. The MSA could also be operated with much-shorter assimilation windows, which allows more emphasis on assimilating short lifetime species on regional and urban scales. It would be interesting to start conducting OSSEs for TEMPO, Sentinel-4, and GEMS to assess the utility of data from each mission when they are integrated in a multi-species data assimilation approach. In particular, it would be interesting to quantify how the observational coverage and measurement precision impacts the ability to constrain tropospheric CO-OH-NO_x-O_3 chemistry.

Bibliography

Adams, C., E. N. Normand, C. A. McLinden, A. E. Bourassa, N. D. Lloyd, D. A. Degenstein, N. A. Krotkov, M. Belmonte Rivas, K. F. Boersma, and H. Eskes (2016), Limb-nadir matching using non-coincident NO_2 observations: proof of concept and the OMI-minus-OSIRIS prototype product, *Atmospheric Measurement Techniques*, *9*(8), 4103–4122, doi:10.5194/amt-9-4103-2016.

Alvarado, M. J., V. H. Payne, K. E. Cady-Pereira, J. D. Hegarty, S. S. Kulawik, K. J. Wecht, J. R. Worden, J. V. Pittman, and S. C. Wofsy (2015), Impacts of updated spectroscopy on thermal infrared retrievals of methane evaluated with HIPPO data, *Atmospheric Measurement Techniques*, *8*(2), 965–985, doi:10.5194/amt-8-965-2015.

Archibald, A. T., A. S. Petit, C. J. Percival, J. N. Harvey, and D. E. Shallcross (2009), On the importance of the reaction between OH and RO_2 radicals, *Atmospheric Science Letters*, *10*(2), 102–108, doi:10.1002/asl.216.

Arellano, A. F., P. S. Kasibhatla, L. Giglio, G. R. van der Werf, and J. T. Randerson (2004), Top-down estimates of global CO sources using MOPITT measurements, *Geophysical Research Letters*, *31*(1), doi:10.1029/2003GL018609, l01104.

Arellano, A. F., P. S. Kasibhatla, L. Giglio, G. R. van der Werf, J. T. Randerson, and G. J. Collatz (2006), Time-dependent inversion estimates of global biomass-burning co emissions using Measurement of Pollution in the Troposphere (MOPITT) measurements, *Journal of Geophysical Research: Atmospheres*, *111*(D9), doi:10.1029/2005JD006613.

ATom Science Team (2016), Atomospheric Tomography (ATom) [image], *Oak Ridge National Laboratory DAAC, Accessed: March, 2018*.

Barkley, M. P., I. D. Smedt, M. Van Roozendael, T. P. Kurosu, K. Chance, A. Arneth, D. Hagberg, A. Guenther, F. Paulot, E. Marais, and J. Mao (2013), Top-down isoprene emissions over tropical South America inferred from SCIAMACHY and OMI

formaldehyde columns, *Journal of Geophysical Research: Atmospheres*, *118*(12), 6849–6868, doi:10.1002/jgrd.50552.

Barré, J., V.-H. Peuch, J.-L. Attié, L. El Amraoui, W. A. Lahoz, B. Josse, M. Claeyman, and P. Nédélec (2012), Stratosphere-troposphere ozone exchange from high resolution MLS ozone analyses, *Atmospheric Chemistry and Physics*, *12*(14), 6129–6144, doi:10.5194/acp-12-6129-2012.

Beer, R., T. A. Glavich, and D. M. Rider (2001), Tropospheric emission spectrometer for the Earth Observing System's Aura satellite, *Appl. Opt.*, *40*(15), 2356–2367, doi:10.1364/AO.40.002356.

Boersma, K. F., D. J. Jacob, H. J. Eskes, R. W. Pinder, J. Wang, and R. J. van der A (2008), Intercomparison of SCIAMACHY and OMI tropospheric NO_2 columns: Observing the diurnal evolution of chemistry and emissions from space, *Journal of Geophysical Research: Atmospheres*, *113*(D16), doi:10.1029/2007JD008816, d16S26.

Bourassa, A. E., L. A. Rieger, N. D. Lloyd, and D. A. Degenstein (2012), Odin-OSIRIS stratospheric aerosol data product and SAGE III intercomparison, *Atmospheric Chemistry and Physics*, *12*(1), 605–614, doi:10.5194/acp-12-605-2012.

Bousquet, P., D. A. Hauglustaine, P. Peylin, C. Carouge, and P. Ciais (2005), Two decades of OH variability as inferred by an inversion of atmospheric transport and chemistry of methyl chloroform, *Atmospheric Chemistry and Physics*, *5*(10), 2635–2656, doi:10.5194/acp-5-2635-2005.

Bousserez, N. (2014), Space-based retrieval of NO_2 over biomass burning regions: quantifying and reducing uncertainties, *Atmospheric Measurement Techniques*, *7*(10), 3431–3444, doi:10.5194/amt-7-3431-2014.

Bowman, K. W., C. D. Rodgers, S. S. Kulawik, J. Worden, E. Sarkissian, G. Osterman, T. Steck, M. Lou, A. Eldering, M. Shephard, et al. (2006), Tropospheric emission spectrometer: Retrieval method and error analysis, *IEEE Transactions on Geoscience and Remote Sensing*, *44*(5), 1297–1307.

Boxe, C. S., J. R. Worden, K. W. Bowman, S. S. Kulawik, J. L. Neu, W. C. Ford, G. B. Osterman, R. L. Herman, A. Eldering, D. W. Tarasick, A. M. Thompson, D. C. Doughty, M. R. Hoffmann, and S. J. Oltmans (2010), Validation of northern latitude

Tropospheric Emission Spectrometer stare ozone profiles with ARC-IONS sondes during ARCTAS: sensitivity, bias and error analysis, *Atmospheric Chemistry and Physics*, *10*(20), 9901–9914, doi:10.5194/acp-10-9901-2010.

Boynard, A., D. Hurtmans, M. E. Koukouli, F. Goutail, J. Bureau, S. Safieddine, C. Lerot, J. Hadji-Lazaro, C. Wespes, J.-P. Pommereau, A. Pazmino, I. Zyrichidou, D. Balis, A. Barbe, S. N. Mikhailenko, D. Loyola, P. Valks, M. Van Roozendael, P.-F. Coheur, and C. Clerbaux (2016), Seven years of IASI ozone retrievals from FORLI: validation with independent total column and vertical profile measurements, *Atmospheric Measurement Techniques*, *9*(9), 4327–4353, doi:10.5194/amt-9-4327-2016.

Brasseur, G. P., and D. J. Jacob (2017), *Modeling of Atmospheric Chemistry*, Cambridge University Press.

Brohede, S., C. A. McLinden, J. Urban, C. S. Haley, A. I. Jonsson, and D. Murtagh (2008), Odin stratospheric proxy NO_y measurements and climatology, *Atmospheric Chemistry and Physics*, *8*(19), 5731–5754, doi:10.5194/acp-8-5731-2008.

Buchholz, R. R., M. N. Deeter, H. M. Worden, J. Gille, D. P. Edwards, J. W. Hannigan, N. B. Jones, C. Paton-Walsh, D. W. T. Griffith, D. Smale, J. Robinson, K. Strong, S. Conway, R. Sussmann, F. Hase, T. Blumenstock, E. Mahieu, and B. Langerock (2017), Validation of MOPITT carbon monoxide using ground-based Fourier transform infrared spectrometer data from NDACC, *Atmospheric Measurement Techniques*, *10*(5), 1927–1956, doi:10.5194/amt-10-1927-2017.

Bucsela, E. J., E. A. Celarier, M. O. Wenig, J. F. Gleason, J. P. Veefkind, K. F. Boersma, and E. J. Brinksma (2006), Algorithm for NO_2 vertical column retrieval from the Ozone Monitoring Instrument, *IEEE Transactions on Geoscience and remote sensing*, *44*(5), 1245–1258.

Burkholder, J., S. Sander, J. Abbatt, J. Barker, R. Huie, C. Kolb, M. Kurylo, V. Orkin, D. Wilmouth, and P. Wine (2015), Chemical kinetics and photochemical data for use in atmospheric studies: evaluation number 18, *Tech. rep.*, Pasadena, CA: Jet Propulsion Laboratory, National Aeronautics and Space .

Byrd, R., P. Lu, J. Nocedal, and C. Zhu (1995), A limited memory algorithm for bound constrained optimization, *SIAM Journal on Scientific Computing*, *16*(5), 1190–1208, doi:10.1137/0916069.

Canty, T. P., L. Hembeck, T. P. Vinciguerra, D. C. Anderson, D. L. Goldberg, S. F. Carpenter, D. J. Allen, C. P. Loughner, R. J. Salawitch, and R. R. Dickerson (2015), Ozone and NO_x chemistry in the eastern US: evaluation of CMAQ/CB05 with satellite (OMI) data, *Atmospheric Chemistry and Physics*, *15*(19), 10,965–10,982, doi:10.5194/acp-15-10965-2015.

Chai, T., G. R. Carmichael, A. Sandu, Y. Tang, and D. N. Daescu (2006), Chemical data assimilation of transport and chemical evolution over the Pacific (TRACE-P) aircraft measurements, *Journal of Geophysical Research: Atmospheres*, *111*(D2), doi:10.1029/2005JD005883, d02301.

Chai, T., H.-C. Kim, P. Lee, D. Tong, L. Pan, Y. Tang, J. Huang, J. McQueen, M. Tsidulko, and I. Stajner (2013), Evaluation of the United States National Air Quality Forecast Capability experimental real-time predictions in 2010 using Air Quality System ozone and NO_2 measurements, *Geoscientific Model Development*, *6*(5), 1831–1850, doi:10.5194/gmd-6-1831-2013.

Chance, K. (1998), Analysis of BrO measurements from the Global Ozone Monitoring Experiment, *Geophysical Research Letters*, *25*(17), 3335–3338, doi:10.1029/98GL52359.

Chance, K. (2007), OMI/Aura Formaldehyde (HCHO) Total Column 1-orbit L2 Swath 13×24 km V003 [Dataset], *Goddard Earth Sciences Data and Information Services Center (GES DISC), Accessed: January, 2017*, doi:10.5067/Aura/OMI/DATA2015.

Clerbaux, C., A. Boynard, L. Clarisse, M. George, J. Hadji-Lazaro, H. Herbin, D. Hurtmans, M. Pommier, A. Razavi, S. Turquety, C. Wespes, and P.-F. Coheur (2009), Monitoring of atmospheric composition using the thermal infrared IASI/MetOp sounder, *Atmospheric Chemistry and Physics*, *9*(16), 6041–6054, doi:10.5194/acp-9-6041-2009.

Cooper, O. R., D. Parrish, J. Ziemke, M. Cupeiro, I. Galbally, S. Gilge, L. Horowitz, N. Jensen, J.-F. Lamarque, V. Naik, et al. (2014), Global distribution and trends of tropospheric ozone: An observation-based review, *Elementa science of the Anthropocene*, *2*, doi:http://doi.org/10.12952/journal.elementa.000029, p.000029.

Dalsøren, S. B., I. S. A. Isaksen, L. iI, and A. Richter (2009), Effect of emission changes in Southeast Asia on global hydroxyl and methane lifetime, *Tellus B*, *61*(4), 588–601, doi:10.1111/j.1600-0889.2009.00429.x.

Deeter, M. N., H. M. Worden, J. C. Gille, D. P. Edwards, D. Mao, and J. R. Drummond (2011), MOPITT multispectral CO retrievals: Origins and effects of geophys-

ical radiance errors, *Journal of Geophysical Research: Atmospheres*, *116*(D15), doi: 10.1029/2011JD015703, d15303.

Deeter, M. N., D. P. Edwards, G. L. Francis, J. C. Gille, S. Martínez-Alonso, H. M. Worden, and C. Sweeney (2017), A climate-scale satellite record for carbon monoxide: the MOPITT Version 7 product, *Atmospheric Measurement Techniques*, *10*(7), 2533–2555, doi:10.5194/amt-10-2533-2017.

Degenstein, D. A., A. E. Bourassa, C. Z. Roth, and E. J. Llewellyn (2009), Limb scatter ozone retrieval from 10 to 60 km using a multiplicative algebraic reconstruction technique, *Atmospheric Chemistry and Physics*, *9*(17), 6521–6529, doi:10.5194/acp-9-6521-2009.

Deng, F., D. B. A. Jones, T. W. Walker, M. Keller, K. W. Bowman, D. K. Henze, R. Nassar, E. A. Kort, S. C. Wofsy, K. A. Walker, A. E. Bourassa, and D. A. Degenstein (2015), Sensitivity analysis of the potential impact of discrepancies in stratosphere-troposphere exchange on inferred sources and sinks of CO_2, *Atmospheric Chemistry and Physics*, *15*(20), 11,773–11,788, doi:10.5194/acp-15-11773-2015.

Dobber, M. R., R. J. Dirksen, P. F. Levelt, G. H. J. van den Oord, R. H. M. Voors, Q. Kleipool, G. Jaross, M. Kowalewski, E. Hilsenrath, G. W. Leppelmeier, Johan de Vries, W. Dierssen, and N. C. Rozemeijer (2006), Ozone monitoring instrument calibration, *IEEE Transactions on Geoscience and Remote Sensing*, *44*(5), 1209–1238.

Duncan, B. N., J. A. Logan, I. Bey, I. A. Megretskaia, R. M. Yantosca, P. C. Novelli, N. B. Jones, and C. P. Rinsland (2007), Global budget of CO, 19881997: Source estimates and validation with a global model, *Journal of Geophysical Research: Atmospheres*, *112*(D22), doi:10.1029/2007JD008459, d22301.

Eibern, H., and H. Schmidt (1999), A four-dimensional variational chemistry data assimilation scheme for Eulerian chemistry transport modeling, *Journal of Geophysical Research: Atmospheres*, *104*(D15), 18,583–18,598, doi:10.1029/1999JD900280.

Emili, E., B. Barret, S. Massart, E. Le Flochmoen, A. Piacentini, L. El Amraoui, O. Pannekoucke, and D. Cariolle (2014), Combined assimilation of IASI and MLS observations to constrain tropospheric and stratospheric ozone in a global chemical transport model, *Atmospheric Chemistry and Physics*, *14*(1), 177–198, doi:10.5194/acp-14-177-2014.

EOS MLS Science Team (2011), MLS/Aura Level 2 Nitric Acid (HNO_3) Mixing Ratio

V003 [Dataset], *Goddard Earth Sciences Data and Information Services Center (GES DISC), Accessed: March, 2017*, doi:10.5067/Aura/MLS/ML2HNO3.003.

Fortems-Cheiney, A., F. Chevallier, I. Pison, P. Bousquet, S. Szopa, M. N. Deeter, and C. Clerbaux (2011), Ten years of CO emissions as seen from Measurements of Pollution in the Troposphere (MOPITT), *Journal of Geophysical Research: Atmospheres*, *116*(D5), doi:10.1029/2010JD014416, d05304.

Gaudel, A., O. Cooper, G. Ancellet, B. Barret, A. Boynard, J. Burrows, C. Clerbaux, P.-F. Coheur, J. Cuesta, E. Cuevas Agulló, et al. (2018), Tropospheric Ozone Assessment Report: Present-day distribution and trends of tropospheric ozone relevant to climate and global atmospheric chemistry model evaluation, *Elementa: Science of the Anthropocene*, *6*(1), doi:10.1525/elementa.291.

Giglio, L., J. T. Randerson, G. R. van der Werf, P. S. Kasibhatla, G. J. Collatz, D. C. Morton, and R. S. DeFries (2010), Assessing variability and long-term trends in burned area by merging multiple satellite fire products, *Biogeosciences*, *7*(3), 1171–1186, doi:10.5194/bg-7-1171-2010.

Giglio, L., J. T. Randerson, and G. R. van der Werf (2013), Analysis of daily, monthly, and annual burned area using the fourth-generation global fire emissions database (GFED4), *Journal of Geophysical Research: Biogeosciences*, *118*(1), 317–328, doi:10.1002/jgrg.20042.

Goldan, P. D., W. C. Kuster, D. L. Albritton, F. C. Fehsenfeld, P. S. Connell, R. B. Norton, and B. J. Huebert (1983), Calibration and tests of the filter-collection method for measuring clean-air, ambient levels of nitric acid, *Atmospheric Environment (1967)*, *17*(7), 1355–1364, doi:https://doi.org/10.1016/0004-6981(83)90410-9.

González Abad, G., X. Liu, K. Chance, H. Wang, T. P. Kurosu, and R. Suleiman (2015), Updated Smithsonian Astrophysical Observatory Ozone Monitoring Instrument (SAO OMI) formaldehyde retrieval, *Atmospheric Measurement Techniques*, *8*(1), 19–32, doi:10.5194/amt-8-19-2015.

Gu, D., Y. Wang, C. Smeltzer, and K. F. Boersma (2014), Anthropogenic emissions of NO_x over China: Reconciling the difference of inverse modeling results using GOME-2 and OMI measurements, *Journal of Geophysical Research: Atmospheres*, *119*(12), 7732–7740, doi:10.1002/2014JD021644, 2014JD021644.

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, doi:10.5194/acp-6-3181-2006.

Haan, D., P. Martinerie, and D. Raynaud (1996), Ice core data of atmospheric carbon monoxide over Antarctica and Greenland during the last 200 years, *Geophysical Research Letters*, *23*(17), 2235–2238, doi:10.1029/96GL02137.

Heald, C. L., D. J. Jacob, D. B. A. Jones, P. I. Palmer, J. A. Logan, D. G. Streets, G. W. Sachse, J. C. Gille, R. N. Hoffman, and T. Nehrkorn (2004), Comparative inverse analysis of satellite (MOPITT) and aircraft (TRACE-P) observations to estimate Asian sources of carbon monoxide, *Journal of Geophysical Research: Atmospheres*, *109*(D23), doi:10.1029/2004JD005185, d23306.

Heland, J., H. Schlager, A. Richter, and J. P. Burrows (2002), First comparison of tropospheric NO_2 column densities retrieved from GOME measurements and in situ aircraft profile measurements, *Geophysical Research Letters*, *29*(20), 44–1–44–4, doi:10.1029/2002GL015528, 1983.

Henze, D. K., A. Hakami, and J. H. Seinfeld (2007), Development of the adjoint of GEOS-Chem, *Atmospheric Chemistry and Physics*, *7*(9), 2413–2433, doi:10.5194/acp-7-2413-2007.

Herman, R., and B. Osterman (2014), Earth Observing System (EOS) Tropospheric Emission Spectrometer (TES) Data Validation Report (Version F07-10 data).

Hooghiemstra, P. B., M. C. Krol, P. Bergamaschi, A. T. J. de Laat, G. R. van der Werf, P. C. Novelli, M. N. Deeter, I. Aben, and T. Röckmann (2012), Comparing optimized CO emission estimates using MOPITT or NOAA surface network observations, *Journal of Geophysical Research: Atmospheres*, *117*(D6), doi:10.1029/2011JD017043, d06309.

Hudman, R. C., L. T. Murray, D. J. Jacob, D. B. Millet, S. Turquety, S. Wu, D. R. Blake, A. H. Goldstein, J. Holloway, and G. W. Sachse (2008a), Biogenic versus anthropogenic sources of CO in the United States, *Geophysical Research Letters*, *35*(4), doi:10.1029/2007GL032393, l04801.

Hudman, R. C., L. T. Murray, D. J. Jacob, D. B. Millet, S. Turquety, S. Wu, D. R. Blake, A. H. Goldstein, J. Holloway, and G. W. Sachse (2008b), Biogenic versus an-

thropogenic sources of CO in the United States, *Geophysical Research Letters*, *35*(4), doi:10.1029/2007GL032393, l04801.

Huijnen, V., M. J. Wooster, J. W. Kaiser, D. L. Gaveau, J. Flemming, M. Parrington, A. Inness, D. Murdiyarso, B. Main, and M. van Weele (2016), Fire carbon emissions over maritime southeast Asia in 2015 largest since 1997, *Nature: Scientific reports*, *6*, 26,886.

Hurtmans, D., P.-F. Coheur, C. Wespes, L. Clarisse, O. Scharf, C. Clerbaux, J. Hadji-Lazaro, M. George, and S. Turquety (2012), FORLI radiative transfer and retrieval code for IASI, *Journal of Quantitative Spectroscopy and Radiative Transfer*, *113*(11), 1391–1408, doi:http://dx.doi.org/10.1016/j.jqsrt.2012.02.036, three Leaders in Spectroscopy.

IASI Science Team (2017), IASI-FORLI Level 2 ozone (O_3), version 20151001 [Dataset], *French Atmosphere Infrastructure AERIS Data Center-ESPRI (former Ether), Accessed: April, 2015*.

Inness, A., A.-M. Blechschmidt, I. Bouarar, S. Chabrillat, M. Crepulja, R. J. Engelen, H. Eskes, J. Flemming, A. Gaudel, F. Hendrick, V. Huijnen, L. Jones, J. Kapsomenakis, E. Katragkou, A. Keppens, B. Langerock, M. de Mazière, D. Melas, M. Parrington, V. H. Peuch, M. Razinger, A. Richter, M. G. Schultz, M. Suttie, V. Thouret, M. Vrekoussis, A. Wagner, and C. Zerefos (2015), Data assimilation of satellite-retrieved ozone, carbon monoxide and nitrogen dioxide with ECMWF's composition-IFS, *Atmospheric Chemistry and Physics*, *15*(9), 5275–5303, doi:10.5194/acp-15-5275-2015.

Jacob, D. (1999), *Introduction to atmospheric chemistry*, Princeton University Press.

Jacob, D. J., J. H. Crawford, H. Maring, A. D. Clarke, J. E. Dibb, L. K. Emmons, R. A. Ferrare, C. A. Hostetler, P. B. Russell, H. B. Singh, A. M. Thompson, G. E. Shaw, E. McCauley, J. R. Pederson, and J. A. Fisher (2010), The Arctic Research of the Composition of the Troposphere from Aircraft and Satellites (ARCTAS) mission: design, execution, and first results, *Atmospheric Chemistry and Physics*, *10*(11), 5191–5212, doi:10.5194/acp-10-5191-2010.

Jaeglé, L., L. Steinberger, R. V. Martin, and K. Chance (2005), Global partitioning of NO_x sources using satellite observations: Relative roles of fossil fuel combustion, biomass burning and soil emissions, *Faraday discussions*, *130*, 407–423.

Jiang, Z., D. B. A. Jones, H. M. Worden, M. N. Deeter, D. K. Henze, J. Worden, K. W. Bowman, C. A. M. Brenninkmeijer, and T. J. Schuck (2013), Impact of model errors in convective transport on CO source estimates inferred from MOPITT CO retrievals, *Journal of Geophysical Research: Atmospheres*, *118*(4), 2073–2083, doi: 10.1002/jgrd.50216.

Jiang, Z., D. B. A. Jones, H. M. Worden, and D. K. Henze (2015a), Sensitivity of top-down CO source estimates to the modeled vertical structure in atmospheric CO, *Atmospheric Chemistry and Physics*, *15*(3), 1521–1537, doi:10.5194/acp-15-1521-2015.

Jiang, Z., D. B. A. Jones, J. Worden, H. M. Worden, D. K. Henze, and Y. X. Wang (2015b), Regional data assimilation of multi-spectral MOPITT observations of CO over North America, *Atmospheric Chemistry and Physics*, *15*(12), 6801–6814, doi: 10.5194/acp-15-6801-2015.

Jiang, Z., J. R. Worden, H. Worden, M. Deeter, D. B. A. Jones, A. F. Arellano, and D. K. Henze (2017), A 15-year record of CO emissions constrained by MOPITT CO observations, *Atmospheric Chemistry and Physics*, *17*(7), 4565–4583, doi:10.5194/acp-17-4565-2017.

Jiang, Z., B. C. McDonald, H. Worden, J. R. Worden, K. Miyazaki, Z. Qu, D. K. Henze, D. B. A. Jones, A. F. Arellano, E. V. Fischer, L. Zhu, and K. F. Boersma (2018), Unexpected slowdown of US pollutant emission reduction in the past decade, *Proceedings of the National Academy of Sciences*, *115*(20), 5099–5104, doi:10.1073/pnas.1801191115.

Jones, D. B. A., K. W. Bowman, J. A. Logan, C. L. Heald, J. Liu, M. Luo, J. Worden, and J. Drummond (2009), The zonal structure of tropical O_3 and CO as observed by the Tropospheric Emission Spectrometer in november 2004 - part 1: Inverse modeling of CO emissions, *Atmospheric Chemistry and Physics*, *9*(11), 3547–3562, doi:10.5194/acp-9-3547-2009.

Jonsson, A. I., J. de Grandpr, V. I. Fomichev, J. C. McConnell, and S. R. Beagley (2004), Doubled CO_2-induced cooling in the middle atmosphere: Photochemical analysis of the ozone radiative feedback, *Journal of Geophysical Research: Atmospheres*, *109*(D24), doi:10.1029/2004JD005093.

Jourdain, L., H. M. Worden, J. R. Worden, K. Bowman, Q. Li, A. Eldering, S. S. Kulawik, G. Osterman, K. F. Boersma, B. Fisher, C. P. Rinsland, R. Beer, and M. Gunson (2007), Tropospheric vertical distribution of tropical Atlantic ozone observed by TES

during the northern African biomass burning season, *Geophysical Research Letters*, *34*(4), doi:10.1029/2006GL028284, l04810.

Kaiser, J., D. J. Jacob, L. Zhu, K. R. Travis, J. A. Fisher, G. González Abad, L. Zhang, X. Zhang, A. Fried, J. D. Crounse, J. M. St. Clair, and A. Wisthaler (2018), High-resolution inversion of OMI formaldehyde columns to quantify isoprene emission on ecosystem-relevant scales: application to the southeast US, *Atmospheric Chemistry and Physics*, *18*(8), 5483–5497, doi:10.5194/acp-18-5483-2018.

Kang, D., V. P. Aneja, R. Mathur, and J. D. Ray (2004), Observed and modeled VOC chemistry under high VOC/NOx conditions in the Southeast United States national parks, *Atmospheric Environment*, *38*(29), 4969–4974, doi: https://doi.org/10.1016/j.atmosenv.2004.05.054.

Kar, J., D. B. A. Jones, J. R. Drummond, J. L. Atti, J. Liu, J. Zou, F. Nichitiu, M. D. Seymour, D. P. Edwards, M. N. Deeter, J. C. Gille, and A. Richter (2008), Measurement of low-altitude CO over the Indian subcontinent by MOPITT, *Journal of Geophysical Research: Atmospheres*, *113*(D16), doi:10.1029/2007JD009362.

Keller, M. (2014), Mitigating model error in CO emission estimation, *Ph.D. thesis*.

Kharol, S., R. Martin, S. Philip, B. Boys, L. Lamsal, M. Jerrett, M. Brauer, D. Crouse, C. McLinden, and R. Burnett (2015), Assessment of the magnitude and recent trends in satellite-derived ground-level nitrogen dioxide over North America, *Atmospheric Environment*, *118*, 236 – 245, doi:https://doi.org/10.1016/j.atmosenv.2015.08.011.

Kleinman, L., Y.-N. Lee, S. R. Springston, L. Nunnermacker, X. Zhou, R. Brown, K. Hallock, P. Klotz, D. Leahy, J. H. Lee, et al. (1994), Ozone formation at a rural site in the southeastern United States, *Journal of Geophysical Research: Atmospheres*, *99*(D2), 3469–3482.

Klonecki, A., M. Pommier, C. Clerbaux, G. Ancellet, J.-P. Cammas, P.-F. Coheur, A. Cozic, G. S. Diskin, J. Hadji-Lazaro, D. A. Hauglustaine, D. Hurtmans, B. Khattatov, J.-F. Lamarque, K. S. Law, P. Nedelec, J.-D. Paris, J. R. Podolske, P. Prunet, H. Schlager, S. Szopa, and S. Turquety (2012), Assimilation of IASI satellite CO fields into a global chemistry transport model for validation against aircraft measurements, *Atmospheric Chemistry and Physics*, *12*(10), 4493–4512, doi:10.5194/acp-12-4493-2012.

Kopacz, M., D. J. Jacob, J. A. Fisher, J. A. Logan, L. Zhang, I. A. Megretskaia, R. M. Yantosca, K. Singh, D. K. Henze, J. P. Burrows, M. Buchwitz, I. Khlystova, W. W.

McMillan, J. C. Gille, D. P. Edwards, A. Eldering, V. Thouret, and P. Nedelec (2010), Global estimates of CO sources with high resolution by adjoint inversion of multiple satellite datasets (MOPITT, AIRS, SCIAMACHY, TES), *Atmospheric Chemistry and Physics*, *10*(3), 855–876, doi:10.5194/acp-10-855-2010.

Krol, M., P. J. van Leeuwen, and J. Lelieveld (1998), Global OH trend inferred from methylchloroform measurements, *Journal of Geophysical Research: Atmospheres*, *103*(D9), 10,697–10,711, doi:10.1029/98JD00459.

Krotkov, N., L. Lamsal, S. Marchenko, E. Celarier, E. Bucsela, W. Swartz, and P. Veefkind (2018), OMI/Aura Nitrogen Dioxide (NO_2) Total and Tropospheric Column 1-orbit L2 Swath 13x24 km V003 [Dataset], *Goddard Earth Sciences Data and Information Services Center (GES DISC)*, doi:10.5067/Aura/OMI/DATA2017.

Krotkov, N. A., L. N. Lamsal, E. A. Celarier, W. H. Swartz, S. V. Marchenko, E. J. Bucsela, K. L. Chan, M. Wenig, and M. Zara (2017), The version 3 OMI NO_2 standard product, *Atmospheric Measurement Techniques*, *10*(9), 3133–3149, doi:10.5194/amt-10-3133-2017.

Kuhn, G. M., H., and V. Etyemezian (2003), *Big Bend Regional Aerosol and Visibility Observational (BRAVO) Study Emissions Inventory, Report prepared for BRAVO Steering Committee*, Desert Research Institute, Las Vegas, Nevada.

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), doi:10.1029/2007JD009235, d16308.

Lelieveld, J., S. Gromov, A. Pozzer, and D. Taraborrelli (2016), Global tropospheric hydroxyl distribution, budget and reactivity, *Atmospheric Chemistry and Physics*, *16*(19), 12,477–12,493, doi:10.5194/acp-16-12477-2016.

Lin, M., L. W. Horowitz, R. Payton, A. M. Fiore, and G. Tonnesen (2017), Us surface ozone trends and extremes from 1980 to 2014: quantifying the roles of rising Asian emissions, domestic controls, wildfires, and climate, *Atmospheric Chemistry and Physics*, *17*(4), 2943–2970, doi:10.5194/acp-17-2943-2017.

Lin, S.-J., and R. B. Rood (1996), Multidimensional flux-form semi-lagrangian transport schemes, *Monthly Weather Review*, *124*(9), 2046–2070, doi:10.1175/1520-0493(1996)124¡2046:MFFSLT¿2.0.CO;2.

Liu, H., D. J. Jacob, I. Bey, and R. M. Yantosca (2001), Constraints from 210Pb and 7Be on wet deposition and transport in a global three-dimensional chemical tracer model driven by assimilated meteorological fields, *Journal of Geophysical Research: Atmospheres*, *106*(D11), 12,109–12,128, doi:10.1029/2000JD900839.

Liu, X., K. Chance, and T. P. Kurosu (2007), Improved ozone profile retrievals from GOME data with degradation correction in reflectance, *Atmospheric Chemistry and Physics*, *7*(6), 1575–1583, doi:10.5194/acp-7-1575-2007.

Livesey, N. J., W. G. Read, L. Froidevaux, A. Lambert, G. L. Manney, H. C. Pumphrey, M. L. Santee, M. J. Schwartz, S. Wang, R. E. Cofield, D. T. Cuddy, R. A. Fuller, R. F. Jarnot, J. H. Jiang, B. W. Knosp, P. C. Stek, P. A. Wagner, and D. L. Wu (2011), Aura microwave limb sounder (MLS) version 3.3 level 2 data quality and description document, version 3.3 x-1.0, *Tech. Rep. JPL D-33509*.

Llewellyn, E. J., N. D. Lloyd, D. A. Degenstein, R. L. Gattinger, S. V. Petelina, A. E. Bourassa, J. T. Wiensz, E. V. Ivanov, I. C. McDade, B. H. Solheim, J. C. McConnell, C. S. Haley, C. von Savigny, C. E. Sioris, C. A. McLinden, E. Griffioen, J. Kaminski, W. F. Evans, E. Puckrin, K. Strong, V. Wehrle, R. H. Hum, D. J. Kendall, J. Matsushita, D. P. Murtagh, S. Brohede, J. Stegman, G. Witt, G. Barnes, W. F. Payne, L. Pich, K. Smith, G. Warshaw, D. L. Deslauniers, P. Marchand, E. H. Richardson, R. A. King, I. Wevers, W. McCreath, E. Kyrl, L. Oikarinen, G. W. Leppelmeier, H. Auvinen, G. Mgie, A. Hauchecorne, F. Lefvre, J. de La Ne, P. Ricaud, U. Frisk, F. Sjoberg, F. von Schele, and L. Nordh (2004), The OSIRIS instrument on the Odin spacecraft, *Canadian Journal of Physics*, *82*(6), 411–422, doi:10.1139/p04-005.

Mao, J., F. Paulot, D. J. Jacob, R. C. Cohen, J. D. Crounse, P. O. Wennberg, C. A. Keller, R. C. Hudman, M. P. Barkley, and L. W. Horowitz (2013), Ozone and organic nitrates over the eastern United States: Sensitivity to isoprene chemistry, *Journal of Geophysical Research: Atmospheres*, *118*(19), 11,256–11,268, doi:10.1002/jgrd.50817.

Marais, E. A., D. J. Jacob, T. P. Kurosu, K. Chance, J. G. Murphy, C. Reeves, G. Mills, S. Casadio, D. B. Millet, M. P. Barkley, F. Paulot, and J. Mao (2012), Isoprene emissions in Africa inferred from OMI observations of formaldehyde columns, *Atmospheric Chemistry and Physics*, *12*(14), 6219–6235, doi:10.5194/acp-12-6219-2012.

Martin, R. V., D. J. Jacob, K. Chance, T. P. Kurosu, P. I. Palmer, and M. J. Evans (2003), Global inventory of nitrogen oxide emissions constrained by space-based ob-

servations of NO_2 columns, *Journal of Geophysical Research: Atmospheres*, *108*(D17), doi:10.1029/2003JD003453, 4537.

Martin, R. V., B. Sauvage, I. Folkins, C. E. Sioris, C. Boone, P. Bernath, and J. Ziemke (2007), Space-based constraints on the production of nitric oxide by lightning, *Journal of Geophysical Research: Atmospheres*, *112*(D9), doi:10.1029/2006JD007831, d09309.

Massart, S., A. Piacentini, and O. Pannekoucke (2012), Importance of using ensemble estimated background error covariances for the quality of atmospheric ozone analyses, *Quarterly Journal of the Royal Meteorological Society*, *138*(665), 889–905, doi:10.1002/qj.971.

McDonald, B. C., S. A. McKeen, Y. Y. Cui, R. Ahmadov, S.-W. Kim, G. J. Frost, I. B. Pollack, J. Peischl, T. B. Ryerson, J. S. Holloway, M. Graus, C. Warneke, J. B. Gilman, J. A. de Gouw, J. Kaiser, F. N. Keutsch, T. F. Hanisco, G. M. Wolfe, and M. Trainer (2018), Modeling Ozone in the Eastern U.S. using a Fuel-Based mobile source emissions inventory, *Environmental Science & Technology*, *52*(13), 7360–7370, doi:10.1021/acs.est.8b00778, pMID: 29870662.

McLinden, C. A., S. C. Olsen, B. Hannegan, O. Wild, M. J. Prather, and J. Sundet (2000), Stratospheric ozone in 3-D models: A simple chemistry and the cross-tropopause flux, *Journal of Geophysical Research: Atmospheres*, *105*(D11), 14,653–14,665, doi:10.1029/2000JD900124.

McPeters, R. D., G. J. Labow, and J. A. Logan (2007), Ozone climatological profiles for satellite retrieval algorithms, *Journal of Geophysical Research: Atmospheres*, *112*(D5), doi:10.1029/2005JD006823.

Millet, D. B., D. J. Jacob, S. Turquety, R. C. Hudman, S. Wu, A. Fried, J. Walega, B. G. Heikes, D. R. Blake, H. B. Singh, B. E. Anderson, and A. D. Clarke (2006), Formaldehyde distribution over North America: Implications for satellite retrievals of formaldehyde columns and isoprene emission, *Journal of Geophysical Research: Atmospheres*, *111*(D24), doi:10.1029/2005JD006853.

Miyazaki, K., H. J. Eskes, K. Sudo, M. Takigawa, M. van Weele, and K. F. Boersma (2012a), Simultaneous assimilation of satellite NO_2, O_3, CO, and HNO_3 data for the analysis of tropospheric chemical composition and emissions, *Atmospheric Chemistry and Physics*, *12*(20), 9545–9579, doi:10.5194/acp-12-9545-2012.

Miyazaki, K., H. J. Eskes, and K. Sudo (2012b), Global NO_x emission estimates derived from an assimilation of OMI tropospheric NO_2 columns, *Atmospheric Chemistry and Physics*, *12*(5), 2263–2288, doi:10.5194/acp-12-2263-2012.

Miyazaki, K., H. J. Eskes, K. Sudo, and C. Zhang (2014), Global lightning NO_x production estimated by an assimilation of multiple satellite data sets, *Atmospheric Chemistry and Physics*, *14*(7), 3277–3305, doi:10.5194/acp-14-3277-2014.

Miyazaki, K., H. J. Eskes, and K. Sudo (2015), A tropospheric chemistry reanalysis for the years 2005 - 2012 based on an assimilation of OMI, MLS, TES, and MOPITT satellite data, *Atmospheric Chemistry and Physics*, *15*(14), 8315–8348, doi:10.5194/acp-15-8315-2015.

Miyazaki, K., H. Eskes, K. Sudo, K. F. Boersma, K. Bowman, and Y. Kanaya (2017), Decadal changes in global surface NO_x emissions from multi-constituent satellite data assimilation, *Atmospheric Chemistry and Physics*, *17*(2), 807–837, doi:10.5194/acp-17-807-2017.

Montzka, S. A., C. M. Spivakovsky, J. H. Butler, J. W. Elkins, L. T. Lock, and D. J. Mondeel (2000), New observational constraints for atmospheric hydroxyl on global and hemispheric scales, *Science*, *288*(5465), 500–503, doi:10.1126/science.288.5465.500.

Montzka, S. A., M. Krol, E. Dlugokencky, B. Hall, P. Jöckel, and J. Lelieveld (2011), Small interannual variability of global atmospheric hydroxyl, *Science*, *331*(6013), 67–69, doi:10.1126/science.1197640.

Müller, J.-F., and T. Stavrakou (2005), Inversion of CO and NO_x emissions using the adjoint of the IMAGES model, *Atmospheric Chemistry and Physics*, *5*(5), 1157–1186, doi:10.5194/acp-5-1157-2005.

Müller, J.-F., T. Stavrakou, M. Bauwens, M. George, D. Hurtmans, P.-F. Coheur, C. Clerbaux, and C. Sweeney (2018), Top-down CO emissions based on IASI observations and hemispheric constraints on OH levels, *Geophysical Research Letters*, *45*(3), 1621–1629, doi:10.1002/2017GL076697.

Murray, L. T., D. J. Jacob, J. A. Logan, R. C. Hudman, and W. J. Koshak (2012), Optimized regional and interannual variability of lightning in a global chemical transport model constrained by LIS/OTD satellite data, *Journal of Geophysical Research: Atmospheres*, *117*(D20), doi:10.1029/2012JD017934, d20307.

Naik, V., A. Voulgarakis, A. M. Fiore, L. W. Horowitz, J.-F. Lamarque, M. Lin, M. J. Prather, P. J. Young, D. Bergmann, P. J. Cameron-Smith, I. Cionni, W. J. Collins, S. B. Dalsøren, R. Doherty, V. Eyring, G. Faluvegi, G. A. Folberth, B. Josse, Y. H. Lee, I. A. MacKenzie, T. Nagashima, T. P. C. van Noije, D. A. Plummer, M. Righi, S. T. Rumbold, R. Skeie, D. T. Shindell, D. S. Stevenson, S. Strode, K. Sudo, S. Szopa, and G. Zeng (2013), Preindustrial to present-day changes in tropospheric hydroxyl radical and methane lifetime from the Atmospheric Chemistry and Climate Model Intercomparison Project (ACCMIP), *Atmospheric Chemistry and Physics*, *13*(10), 5277–5298, doi:10.5194/acp-13-5277-2013.

Nassar, R., J. A. Logan, H. M. Worden, I. A. Megretskaia, K. W. Bowman, G. B. Osterman, A. M. Thompson, D. W. Tarasick, S. Austin, H. Claude, M. K. Dubey, W. K. Hocking, B. J. Johnson, E. Joseph, J. Merrill, G. A. Morris, M. Newchurch, S. J. Oltmans, F. Posny, F. J. Schmidlin, H. Vömel, D. N. Whiteman, and J. C. Witte (2008), Validation of tropospheric emission spectrometer (TES) nadir ozone profiles using ozonesonde measurements, *Journal of Geophysical Research: Atmospheres*, *113*(D15), doi:10.1029/2007JD008819, d15S17.

Nassar, R., J. A. Logan, I. A. Megretskaia, L. T. Murray, L. Zhang, and D. B. A. Jones (2009), Analysis of tropical tropospheric ozone, carbon monoxide, and water vapor during the 2006 El Niño using TES observations and the GEOS-Chem model, *Journal of Geophysical Research: Atmospheres*, *114*(D17), doi:10.1029/2009JD011760.

NCAR: ACOM Science Team (2016), MOPITT V7 TIR/NIR Daytime and Nighttime CO Retrievals–Quick look images [image], *NASA Langley Atmospheric Science Data Center-DAAC, Accessed: September 2013*, doi:10.5067/terra/mopitt/mop02j_l2.007.

Nicely, J. M., R. J. Salawitch, T. Canty, D. C. Anderson, S. R. Arnold, M. P. Chipperfield, L. K. Emmons, J. Flemming, V. Huijnen, D. E. Kinnison, J.-F. Lamarque, J. Mao, S. A. Monks, S. D. Steenrod, S. Tilmes, and S. Turquety (2017), Quantifying the causes of differences in tropospheric OH within global models, *Journal of Geophysical Research: Atmospheres*, *122*(3), 1983–2007, doi:10.1002/2016JD026239.

Olivier, J., and J. Berdowski (2001), *Global emissions sources and sinks, in: Berdowski, J., Guicherit, R. and B.J. Heij (eds.) "The Climate System"*, 33-78 pp., A.A. Balkema Publishers/Swets & Zeitlinger Publishers, Lisse, The Netherlands.

Palmer, P. I., D. J. Jacob, D. B. A. Jones, C. L. Heald, R. M. Yantosca, J. A. Logan, G. W. Sachse, and D. G. Streets (2003a), Inverting for emissions of carbon monoxide

from Asia using aircraft observations over the western Pacific, *Journal of Geophysical Research: Atmospheres*, *108*(D21), doi:10.1029/2003JD003397, 8828.

Palmer, P. I., D. J. Jacob, A. M. Fiore, R. V. Martin, K. Chance, and T. P. Kurosu (2003b), Mapping isoprene emissions over North America using formaldehyde column observations from space, *Journal of Geophysical Research: Atmospheres*, *108*(D6), doi:10.1029/2002JD002153, 4180.

Parrington, M., D. B. A. Jones, K. W. Bowman, L. W. Horowitz, A. M. Thompson, D. W. Tarasick, and J. C. Witte (2008), Estimating the summertime tropospheric ozone distribution over North America through assimilation of observations from the Tropospheric Emission Spectrometer, *Journal of Geophysical Research: Atmospheres*, *113*(D18), doi:10.1029/2007JD009341, d18307.

Parrington, M., P. I. Palmer, A. C. Lewis, J. D. Lee, A. R. Rickard, P. Di Carlo, J. W. Taylor, J. R. Hopkins, S. Punjabi, D. E. Oram, G. Forster, E. Aruffo, S. J. Moller, S. J.-B. Bauguitte, J. D. Allan, H. Coe, and R. J. Leigh (2013), Ozone photochemistry in boreal biomass burning plumes, *Atmospheric Chemistry and Physics*, *13*(15), 7321–7341, doi:10.5194/acp-13-7321-2013.

Patra, P., M. Krol, S. Montzka, T. Arnold, E. L. Atlas, B. Lintner, B. Stephens, B. Xiang, J. Elkins, P. Fraser, et al. (2014), Observational evidence for interhemispheric hydroxyl-radical parity, *Nature*, *513*(7517), 219.

Peiro, H., E. Emili, D. Cariolle, B. Barret, and E. Le Flochmoën (2018), Multi-year assimilation of IASI and MLS ozone retrievals: variability of tropospheric ozone over the tropics in response to ENSO, *Atmospheric Chemistry and Physics*, *18*(9), 6939–6958, doi:10.5194/acp-18-6939-2018.

Pétron, G., C. Granier, B. Khattatov, J.-F. Lamarque, V. Yudin, J.-F. Mller, and J. Gille (2002), Inverse modeling of carbon monoxide surface emissions using Climate Monitoring and Diagnostics Laboratory network observations, *Journal of Geophysical Research: Atmospheres*, *107*(D24), ACH 10–1–ACH 10–23, doi:10.1029/2001JD001305, 4761.

Pison, I., P. Bousquet, F. Chevallier, S. Szopa, and D. Hauglustaine (2009), Multi-species inversion of H_4, CO and H_2 emissions from surface measurements, *Atmospheric Chemistry and Physics*, *9*(14), 5281–5297, doi:10.5194/acp-9-5281-2009.

Prather, M. J., and D. Ehhalt (2001), Climate Change 2001: The Science of Climate Change, *Intergovernmental panel on climate change (IPCC)*.

Price, C., and D. Rind (1992), A simple lightning parameterization for calculating global lightning distributions, *Journal of Geophysical Research: Atmospheres*, *97*(D9), 9919–9933, doi:10.1029/92JD00719.

Prinn, R., D. Cunnold, R. Rasmussen, P. Simmonds, F. Alyea, A. Crawford, P. Fraser, and R. Rosen (1987), Atmospheric trends in methylchloroform and the global average for the hydroxyl radical, *Science*, *238*(4829), 945–950, doi:10.1126/science.238.4829.945.

Qu, Z., D. K. Henze, S. L. Capps, Y. Wang, X. Xu, J. Wang, and M. Keller (2017), Monthly top-down NO_x emissions for China (2005-2012): A hybrid inversion method and trend analysis, *Journal of Geophysical Research: Atmospheres*, *122*(8), 4600–4625, doi:10.1002/2016JD025852, 2016JD025852.

Reidmiller, D. R., A. M. Fiore, D. A. Jaffe, D. Bergmann, C. Cuvelier, F. J. Dentener, B. N. Duncan, G. Folberth, M. Gauss, S. Gong, P. Hess, J. E. Jonson, T. Keating, A. Lupu, E. Marmer, R. Park, M. G. Schultz, D. T. Shindell, S. Szopa, M. G. Vivanco, O. Wild, and A. Zuber (2009), The influence of foreign vs. North American emissions on surface ozone in the us, *Atmospheric Chemistry and Physics*, *9*(14), 5027–5042, doi:10.5194/acp-9-5027-2009.

Rodgers, C. D. (2000), *Inverse methods for atmospheric sounding: theory and practice*, vol. 2, World scientific.

Rohrer, F., K. Lu, A. Hofzumahaus, B. Bohn, T. Brauers, C.-C. Chang, H. Fuchs, R. Häseler, F. Holland, M. Hu, et al. (2014), Maximum efficiency in the hydroxyl-radical-based self-cleansing of the troposphere, *Nature Geoscience*, *7*(8), 559.

Roth, C., D. Degenstein, A. Bourassa, and E. Llewellyn (2007), The retrieval of vertical profiles of the ozone number density using Chappuis band absorption information and a multiplicative algebraic reconstruction technique, *Canadian Journal of Physics*, *85*(11), 1225–1243.

Sander, S., R. Friedl, J. Barker, D. Golden, M. Kurylo, P. Wine, J. Abbatt, J. Burkholder, C. Kolb, G. Moortgat, et al. (2003), Chemical kinetics and photochemical data for use in atmospheric studies, evaluation number 14, *JPL Publ., 02*, *25*, 334.

Santee, M. L., G. L. Manney, N. J. Livesey, M. J. Schwartz, J. L. Neu, and W. G. Read (2017), A comprehensive overview of the climatological composition of the Asian summer monsoon anticyclone based on 10 years of Aura Microwave Limb Sounder

measurements, *Journal of Geophysical Research: Atmospheres*, *122*(10), 5491–5514, doi:10.1002/2016JD026408.

Shindell, D. T., G. Faluvegi, and N. Bell (2003), Preindustrial-to-present-day radiative forcing by tropospheric ozone from improved simulations with the GISS chemistry-climate GCM, *Atmospheric Chemistry and Physics*, *3*(5), 1675–1702, doi:10.5194/acp-3-1675-2003.

Shindell, D. T., G. Faluvegi, D. S. Stevenson, M. C. Krol, L. K. Emmons, J.-F. Lamarque, G. Ptron, F. J. Dentener, K. Ellingsen, M. G. Schultz, O. Wild, M. Amann, C. S. Atherton, D. J. Bergmann, I. Bey, T. Butler, J. Cofala, W. J. Collins, R. G. Derwent, R. M. Doherty, J. Drevet, H. J. Eskes, A. M. Fiore, M. Gauss, D. A. Hauglustaine, L. W. Horowitz, I. S. A. Isaksen, M. G. Lawrence, V. Montanaro, J.-F. Mller, G. Pitari, M. J. Prather, J. A. Pyle, S. Rast, J. M. Rodriguez, M. G. Sanderson, N. H. Savage, S. E. Strahan, K. Sudo, S. Szopa, N. Unger, T. P. C. van Noije, and G. Zeng (2006), Multimodel simulations of carbon monoxide: Comparison with observations and projected near-future changes, *Journal of Geophysical Research: Atmospheres*, *111*(D19), doi:10.1029/2006JD007100, d19306.

Sigrist, M. W., J. D. Winefordner, I. Kolthoff, et al. (1994), *Air monitoring by spectroscopic techniques*, vol. 127, 27-84 pp., John Wiley & Sons.

Skachko, S., R. Ménard, Q. Errera, Y. Christophe, and S. Chabrillat (2016), EnKF and 4D-Var data assimilation with chemical transport model BASCOE (version 05.06), *Geoscientific Model Development*, *9*(8), 2893–2908, doi:10.5194/gmd-9-2893-2016.

Spivakovsky, C. M., J. A. Logan, S. A. Montzka, Y. J. Balkanski, M. Foreman-Fowler, D. B. A. Jones, L. W. Horowitz, A. C. Fusco, C. A. M. Brenninkmeijer, M. J. Prather, S. C. Wofsy, and M. B. McElroy (2000), Three-dimensional climatological distribution of tropospheric OH: Update and evaluation, *Journal of Geophysical Research: Atmospheres*, *105*(D7), 8931–8980, doi:10.1029/1999JD901006.

Stanevich, I. (2018), Variational data assimilation of satellite remote sensing observations for improving methane simulations in chemical transport models, *Ph.D. thesis*.

Stavrakou, T., and J.-F. Müller (2006), Grid-based versus big region approach for inverting CO emissions using Measurement of Pollution in the Troposphere (MOPITT) data, *Journal of Geophysical Research: Atmospheres*, *111*(D15), doi:10.1029/2005JD006896, d15304.

Stavrakou, T., J.-F. Mller, K. F. Boersma, I. De Smedt, and R. J. van der A (2008), Assessing the distribution and growth rates of NO_x emission sources by inverting a 10-year record of NO_2 satellite columns, *Geophysical Research Letters*, *35*(10), doi: 10.1029/2008GL033521.

Streets, D. G., Q. Zhang, L. Wang, K. He, J. Hao, Y. Wu, Y. Tang, and G. R. Carmichael (2006), Revisiting China's CO emissions after the Transport and Chemical Evolution over the Pacific (TRACE-P) mission: Synthesis of inventories, atmospheric modeling, and observations, *Journal of Geophysical Research: Atmospheres*, *111*(D14), doi: 10.1029/2006JD007118, d14306.

Strode, S. A., B. N. Duncan, E. A. Yegorova, J. Kouatchou, J. R. Ziemke, and A. R. Douglass (2015), Implications of carbon monoxide bias for methane lifetime and atmospheric composition in chemistry climate models, *Atmospheric Chemistry and Physics*, *15*(20), 11,789–11,805, doi:10.5194/acp-15-11789-2015.

Tellmann, S., V. Rozanov, M. Weber, and J. Burrows (2004), Improvements in the tropical ozone profile retrieval from GOME-UV/Vis nadir spectra, *Advances in Space Research*, *34*(4), 739 – 743, doi:https://doi.org/10.1016/j.asr.2003.05.056, trace Constituents in the Troposphere and Lower Stratosphere.

TES Science Team (2008), TES/Aura Level 2 ozone (O_3) nadir, version 4 [Dataset], *NASA Langley Research Center's Atmospheric Science Data Center-ASDC, Accessed: September, 2013*, doi:10.5067/AURA/TES/TESO3LN_L2.004.

Thompson, A. M., J. C. Witte, H. G. J. Smit, S. J. Oltmans, B. J. Johnson, V. W. J. H. Kirchhoff, and F. J. Schmidlin (2007), Southern Hemisphere Additional ozonesondes (SHADOZ) 19982004 tropical ozone climatology: 3. instrumentation, station-to-station variability, and evaluation with simulated flight profiles, *Journal of Geophysical Research: Atmospheres*, *112*(D3), doi:10.1029/2005JD007042, d03304.

Travis, K. R., D. J. Jacob, J. A. Fisher, P. S. Kim, E. A. Marais, L. Zhu, K. Yu, C. C. Miller, R. M. Yantosca, M. P. Sulprizio, A. M. Thompson, P. O. Wennberg, J. D. Crounse, J. M. St. Clair, R. C. Cohen, J. L. Laughner, J. E. Dibb, S. R. Hall, K. Ullmann, G. M. Wolfe, I. B. Pollack, J. Peischl, J. A. Neuman, and X. Zhou (2016), Why do models overestimate surface ozone in the Southeast United States?, *Atmospheric Chemistry and Physics*, *16*(21), 13,561–13,577, doi:10.5194/acp-16-13561-2016.

Trémolet, Y. (2006), Accounting for an imperfect model in 4D-Var, *Quarterly Journal of the Royal Meteorological Society*, *132*(621), 2483–2504, doi:10.1256/qj.05.224.

Valin, L. C., A. R. Russell, R. C. Hudman, and R. C. Cohen (2011), Effects of model resolution on the interpretation of satellite NO_2 observations, *Atmospheric Chemistry and Physics*, *11*(22), 11,647–11,655, doi:10.5194/acp-11-11647-2011.

van der A, R. J., H. J. Eskes, K. F. Boersma, T. P. C. van Noije, M. Van Roozendael, I. De Smedt, D. H. M. U. Peters, and E. W. Meijer (2008), Trends, seasonal variability and dominant NO_x source derived from a ten year record of NO_2 measured from space, *Journal of Geophysical Research: Atmospheres*, *113*(D4), doi:10.1029/2007JD009021, d04302.

van der Werf, G. R., J. T. Randerson, L. Giglio, G. J. Collatz, P. S. Kasibhatla, and A. F. Arellano Jr. (2006), Interannual variability in global biomass burning emissions from 1997 to 2004, *Atmospheric Chemistry and Physics*, *6*(11), 3423–3441, doi:10.5194/acp-6-3423-2006.

Verstraeten, W. W., K. F. Boersma, J. Zörner, M. A. F. Allaart, K. W. Bowman, and J. R. Worden (2013), Validation of six years of TES tropospheric ozone retrievals with ozonesonde measurements: implications for spatial patterns and temporal stability in the bias, *Atmospheric Measurement Techniques*, *6*(5), 1413–1423, doi:10.5194/amt-6-1413-2013.

Voulgarakis, A., V. Naik, J.-F. Lamarque, D. T. Shindell, P. J. Young, M. J. Prather, O. Wild, R. D. Field, D. Bergmann, P. Cameron-Smith, I. Cionni, W. J. Collins, S. B. Dalsøren, R. M. Doherty, V. Eyring, G. Faluvegi, G. A. Folberth, L. W. Horowitz, B. Josse, I. A. MacKenzie, T. Nagashima, D. A. Plummer, M. Righi, S. T. Rumbold, D. S. Stevenson, S. A. Strode, K. Sudo, S. Szopa, and G. Zeng (2013), Analysis of present day and future OH and methane lifetime in the ACCMIP simulations, *Atmospheric Chemistry and Physics*, *13*(5), 2563–2587, doi:10.5194/acp-13-2563-2013.

Wang, B., and Z. Chen (2013), An intercomparison of satellite-derived ground-level NO2 concentrations with GMSMB modeling results and in-situ measurements A North American study, *Environmental Pollution*, *181*, 172 – 181, doi: https://doi.org/10.1016/j.envpol.2013.06.037.

Wang, Y., D. J. Jacob, and J. A. Logan (1998a), Global simulation of tropospheric O_3-NO_x -hydrocarbon chemistry: 1. Model formulation, *Journal of Geophysical Research: Atmospheres*, *103*(D9), 10,713–10,725, doi:10.1029/98JD00158.

Wang, Y., D. J. Jacob, and J. A. Logan (1998b), Global simulation of tropospheric O_3-NO_x -hydrocarbon chemistry: 3. Origin of tropospheric ozone and effects of non-

methane hydrocarbons, *Journal of Geophysical Research: Atmospheres*, *103*(D9), 10,757–10,767, doi:10.1029/98JD00156.

Wild, O. (2007), Modelling the global tropospheric ozone budget: exploring the variability in current models, *Atmospheric Chemistry and Physics*, *7*(10), 2643–2660, doi: 10.5194/acp-7-2643-2007.

Witte, J. C., A. M. Thompson, H. G. J. Smit, M. Fujiwara, F. Posny, G. J. R. Coetzee, E. T. Northam, B. J. Johnson, C. W. Sterling, M. Mohamad, S.-Y. Ogino, A. Jordan, and F. R. da Silva (2017), First reprocessing of Southern Hemisphere ADditional OZonesondes (SHADOZ) profile records (19982015): 1. Methodology and evaluation, *Journal of Geophysical Research: Atmospheres*, *122*(12), 6611–6636, doi: 10.1002/2016JD026403.

Wofsy, S. C., the HIPPO Science Team, C. Modellers, and S. Teams (2011), HIAPER Pole-to-Pole Observations (HIPPO): fine-grained, global-scale measurements of climatically important atmospheric gases and aerosols, *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, *369*(1943), 2073–2086, doi:10.1098/rsta.2010.0313.

Worden, H. M., J. A. Logan, J. R. Worden, R. Beer, K. Bowman, S. A. Clough, A. Eldering, B. M. Fisher, M. R. Gunson, R. L. Herman, S. S. Kulawik, M. C. Lampel, M. Luo, I. A. Megretskaia, G. B. Osterman, and M. W. Shephard (2007), Comparisons of tropospheric emission spectrometer (TES) ozone profiles to ozonesondes: Methods and initial results, *Journal of Geophysical Research: Atmospheres*, *112*(D3), doi:10.1029/2006JD007258, d03309.

Worden, H. M., M. N. Deeter, D. P. Edwards, J. C. Gille, J. R. Drummond, and P. Nédélec (2010), Observations of near-surface carbon monoxide from space using MOPITT multispectral retrievals, *Journal of Geophysical Research: Atmospheres*, *115*(D18), doi:10.1029/2010JD014242, d18314.

Wunch, D., G. C. Toon, P. O. Wennberg, S. C. Wofsy, B. B. Stephens, M. L. Fischer, O. Uchino, J. B. Abshire, P. Bernath, S. C. Biraud, J.-F. L. Blavier, C. Boone, K. P. Bowman, E. V. Browell, T. Campos, B. J. Connor, B. C. Daube, N. M. Deutscher, M. Diao, J. W. Elkins, C. Gerbig, E. Gottlieb, D. W. T. Griffith, D. F. Hurst, R. Jiménez, G. Keppel-Aleks, E. A. Kort, R. Macatangay, T. Machida, H. Matsueda, F. Moore, I. Morino, S. Park, J. Robinson, C. M. Roehl, Y. Sawa, V. Sherlock, C. Sweeney, T. Tanaka, and M. A. Zondlo (2010), Calibration of the Total Car-

bon Column Observing Network using aircraft profile data, *Atmospheric Measurement Techniques*, *3*(5), 1351–1362, doi:10.5194/amt-3-1351-2010.

Yan, Y.-Y., J.-T. Lin, Y. Kuang, D. Yang, and L. Zhang (2014), Tropospheric carbon monoxide over the Pacific during HIPPO: two-way coupled simulation of GEOS-Chem and its multiple nested models, *Atmospheric Chemistry and Physics*, *14*(23), 12,649–12,663, doi:10.5194/acp-14-12649-2014.

Yantosca, B. (2018), Mean OH concentration, *GEOS-Chem community wiki*.

Zhang, L., D. J. Jacob, K. F. Boersma, D. A. Jaffe, J. R. Olson, K. W. Bowman, J. R. Worden, A. M. Thompson, M. A. Avery, R. C. Cohen, J. E. Dibb, F. M. Flock, H. E. Fuelberg, L. G. Huey, W. W. McMillan, H. B. Singh, and A. J. Weinheimer (2008a), Transpacific transport of ozone pollution and the effect of recent Asian emission increases on air quality in North America: an integrated analysis using satellite, aircraft, ozonesonde, and surface observations, *Atmospheric Chemistry and Physics*, *8*(20), 6117–6136, doi:10.5194/acp-8-6117-2008.

Zhang, L., E. M. Constantinescu, A. Sandu, Y. Tang, T. Chai, G. R. Carmichael, D. Byun, and E. Olaguer (2008b), An adjoint sensitivity analysis and 4D-Var data assimilation study of Texas air quality, *Atmospheric Environment*, *42*, 5787–5804, doi:10.1016/j.atmosenv.2008.03.048.

Zhang, R., W. Lei, X. Tie, and P. Hess (2004), Industrial emissions cause extreme urban ozone diurnal variability, *Proceedings of the National Academy of Sciences of the United States of America*, *101*(17), 6346–6350.

Zhang, Y., D. J. Jacob, J. D. Maasakkers, M. P. Sulprizio, J.-X. Sheng, R. Gautam, and J. Worden (2018), Monitoring global tropospheric OH concentrations using satellite observations of atmospheric methane, *Atmospheric Chemistry and Physics*, *18*(21), 15,959–15,973, doi:10.5194/acp-18-15959-2018.

Zheng, F., T. Yu, T. Cheng, X. Gu, and H. Guo (2014), Intercomparison of tropospheric nitrogen dioxide retrieved from Ozone Monitoring Instrument over China, *Atmospheric Pollution Research*, *5*(4), 686–695, doi:https://doi.org/10.5094/APR.2014.078.

Zhu, L., L. J. Mickley, D. J. Jacob, E. A. Marais, J. Sheng, L. Hu, G. Abad, and K. Chance (2017), Long-term (2005-2014) trends in formaldehyde (HCHO) columns across North America as seen by the OMI satellite instrument: Evidence of changing

emissions of volatile organic compounds, *Geophysical Research Letters*, *44*(13), 7079–7086, doi:10.1002/2017GL073859.