

UNIVERSITY OF WOLLONGONG

DOCTORAL THESIS

Thesis Title

Author:

Supervisors:
Dr. Jenny FISHER

*A thesis submitted in fulfillment of the requirements
for the degree of Doctor of Philosophy*

in the

Centre of Atmospheric Chemistry
Chemistry Department

March 15, 2019

Declaration of Authorship

I, Jesse GREENSLADE, declare that this thesis titled, "Thesis Title" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date:

“Thanks to my solid academic training, today I can write hundreds of words on virtually any topic without possessing a shred of information, which is how I got a good job in journalism.”

Dave Barry

UNIVERSITY OF WOLLONGONG

Abstract

Science Medicine And Health
Chemistry Department

Doctor of Philosophy

Thesis Title

by Jesse GREENSLADE

Ozone in the troposphere is a toxic pollutant which causes respiratory and agricultural damage. The two main precursors are biogenic emissions of volatile organic compounds (BVOCs) and transport from the stratosphere. Most tropospheric ozone is formed through chemical reactions involving nitrogen oxides, the hydroxy radical, and volatile organic compounds (such as isoprene). Atmospheric chemistry transport models (CTMs) have uncertain BVOC emissions in Australia, affecting our ability to model and predict ozone production along with other important atmospheric processes. Isoprene, formaldehyde, and ozone in the troposphere are linked by oxidative chemistry and are all important to air quality, climate, and radiation budgets. My thesis has three aims: recalculating formaldehyde amounts over Australia seen by satellite using a global CTM (GEOS-Chem), determination of isoprene emissions using modelled formaldehyde yields along with satellite formaldehyde amounts, and attribution of ozone in the troposphere (from either chemical production following isoprene emissions or transported from the stratosphere). Model and observations are combined for each aim in this thesis.

To quantify isoprene emissions (the dominant BVOC), formaldehyde (one of the main products of isoprene oxidation) is used as a proxy. Formaldehyde observed by satellite over Australia is calculated based on modelled a priori vertical distributions using both CTMs and radiative transfer models. In order to compare satellite formaldehyde products against models or other measurements, corrections are required to remove the influence of the a priori profile. Impacts on formaldehyde levels from anthropogenic and pyrogenic sources are determined and filtered out to determine the biogenic footprint and minimise bias in isoprene emissions quantification.

Isoprene is predominantly emitted by trees and shrubs, and Eucalypts are potentially very high emitters along with being close to population centres in Australia. Subsequent oxidation reactions form formaldehyde which has a sufficiently long lifetime in the atmosphere to establish chemical equilibrium. Using a simple linear model and assuming minimal transport and a chemical steady state allows an estimate of the yield of formaldehyde from isoprene emissions. This yield is modelled over Australia and then applied to the recalculated satellite formaldehyde to create a new estimate of Australian isoprene emissions. This technique is used to improve isoprene emissions estimates without the need for extensive measurement campaigns. Results are compared against existing measurement campaigns for validation and analysis. Isoprene emissions from Australian forests appear substantially lower than previous estimates have predicted.

The second most abundant source of tropospheric ozone is the stratosphere, which occasionally mixes into the troposphere bringing ozone-rich air masses down towards the earth's surface. Analysing the local weather patterns and ozone seasonality, most transport is seen to occur during low pressure frontal weather systems. This work provides a novel technique using a Fourier filter on ozone profiles for estimation and quantification of tropospheric ozone transported from the stratosphere. An estimate encompassing three measurement stations over the Southern Ocean near Australia of about 7.2×10^{17} molec cm⁻² per year is derived.

Overall, this work should improve the knowledge of tropospheric ozone and its precursors for Australia.

Acknowledgements

Thanks to my supervisor Jenny Fisher, the whole Atmospheric Chemistry team at Wollongong who made me feel at home: Clare Murphy, Dagmar, Elise, Joel, Kaitlyn, Max, Nick Deutscher, Nick Jones, Stephen Wilson,

Thanks also to my personal saviour and her noodly appendages.

Contents

Declaration of Authorship	iii
Abstract	viii
Acknowledgements	ix
1 Introduction and Literature Review	1
1.1 The atmosphere	1
1.1.1 Structure	2
1.1.2 Composition and chemistry	3
1.1.3 Radiative Forcing	3
1.2 Ozone	4
1.2.1 Stratospheric ozone	4
1.2.2 Tropospheric ozone	6
1.2.3 Stratosphere to troposphere transport	9
1.2.4 Chemical production	10
1.3 Volatile Organic Compounds	10
1.3.1 Emissions	13
1.3.2 Isoprene	14
1.3.3 Isoprene chemistry	14
1.3.3.1 Oxidation	16
1.3.3.2 High NO _x pathway	16
1.3.3.3 Low NO _x pathway	17
1.3.3.4 Night time processes	18
1.4 Formaldehyde	18
1.4.1 Sources and sinks	18
1.4.2 Measurement techniques	20
1.4.2.1 Satellite measurements	21
1.5 Atmospheric Chemistry Modelling	22
1.5.1 Box models	23
1.5.2 Chemical transport models	23
1.5.3 Emissions	25
1.5.4 Uncertainties	25
1.5.4.1 Emissions Inventories	26
1.5.4.2 Resolution	26
1.5.4.3 Chemistry mechanisms	27
1.5.4.4 Clouds	27
1.5.4.5 Soil Moisture	27
1.6 Australia and the southern hemisphere	28

1.6.1	Ozone	29
1.6.2	VOCs	30
1.6.3	Measurements	31
1.7	Aims	32
2	Data and Modelling	35
2.1	Introduction	35
2.2	Datasets	36
2.2.1	Satellite	37
2.2.1.1	Formaldehyde	37
2.2.1.2	Nitrogen dioxide	38
2.2.1.3	Aerosol optical depth	40
2.2.1.4	Active fires	40
2.2.1.5	Carbon monoxide	40
2.2.1.6	Uncertainties	41
2.2.2	Model datasets	41
2.2.2.1	GEOS-Chem output	41
2.2.2.2	Meteorological reanalysis	42
2.2.2.3	Surface temperatures	42
2.2.3	Campaign datasets	42
2.2.3.1	Measurements of Urban, Marine and Biogenic Air (MUMBA)	45
2.2.3.2	Sydney Particle Studies (SPS1, SPS2)	45
2.2.3.3	Ozonesondes	46
2.2.3.4	Wollongong	47
2.2.3.5	Uncertainties	47
2.3	Satellite formaldehyde	47
2.3.1	Pixel filtering	50
2.3.2	DOAS	51
2.3.3	Air mass factor (AMF)	53
2.3.4	LIDORT	54
2.3.5	Uncertainty	54
2.4	GEOS-Chem	57
2.4.1	Overview	57
2.4.2	Installing and running GEOS-Chem	57
2.4.3	Chemical Mechanism	58
2.4.4	GEOS-Chem isoprene	59
2.4.4.1	Oxidation	60
2.4.4.2	Nitrogen oxide impacts	61
2.4.4.3	OH	61
2.4.5	Emissions from MEGAN	62
2.4.6	Nitrogen oxides	63
2.4.7	GEOS-Chem simulations	64
2.4.7.1	GEOS-Chem outputs	64
2.4.7.2	GEOS-Chem runs	68
2.4.7.3	UCX vs tropchem	69
2.5	Calculating new AMF	70
2.6	Recalculation of OMI HCHO	78

2.6.1	Outline	78
2.6.2	Creating new shape factors	80
2.6.3	Creating new AMF using GEOS-Chem	81
2.6.3.1	Updating shape factors	81
2.6.3.2	Recalculating the AMF using PP code	81
2.6.3.3	Saving the AMF with satellite pixels	82
2.6.4	Vertical columns from AMF	82
2.6.5	Reference sector correction	82
2.6.6	Corrected vertical columns	83
2.6.7	Binning the results daily	86
2.6.8	Difference between original and corrected OMI HCHO columns	87
2.7	Filtering Data	87
2.7.1	Pyrogenic filter	90
2.7.2	Anthropogenic filter	104
2.8	Process schematic	108
2.9	Data Access	108
3	Biogenic Isoprene emissions in Australia	111
3.1	Introduction	111
3.1.1	Aims	112
3.1.2	Existing emissions estimates	112
3.1.3	Top-down isoprene emissions estimates	113
3.1.3.1	Bayesian inversion	113
3.1.3.2	Linear inversion	114
3.2	Methods	115
3.2.1	Outline	115
3.2.2	Masks and reprocessed satellite HCHO	116
3.2.3	GEOS-Chem emissions	118
3.2.4	Relationship between isoprene emissions and formaldehyde	118
3.2.5	Calculation of modelled slope	120
3.2.6	Calculation of Emissions	122
3.2.7	Accounting for smearing	126
3.2.7.1	Calculation of smearing	126
3.2.7.2	Sensitivity to smearing	128
3.2.7.3	Smearing length scale	131
3.2.7.4	NO _x dependence	132
3.2.8	Running GEOS-Chem using a posteriori emissions	132
3.3	Results	136
3.3.1	A posteriori emissions	136
3.3.2	Modelled impacts of reduced isoprene emissions	146
3.3.2.1	Implications for HCHO	146
3.3.2.2	Implications for ozone	149
3.3.2.3	Trends	149
3.3.3	Comparison with in situ measurements	150
3.4	Uncertainty	150
3.4.1	Top down emissions	157
3.4.2	Model Uncertainty	158

3.4.3	Satellite Uncertainty	158
3.4.3.1	Fitting error from the OMI retrieval	160
3.4.3.2	Uncertainty in AMF calculations	160
3.4.3.3	Uncertainty of HCHO background	160
3.4.4	Sensitivity to vertical column recalculation	161
3.4.5	Sensitivity to filtering	161
3.4.6	Summary	161
3.5	Conclusions and implications	161
4	Stratospheric ozone intrusions	163
4.1	Foreword	163
4.2	Introduction	163
4.3	Data and Methods	165
4.3.1	Ozonesonde record in the Southern Ocean	165
4.3.2	Model description	168
4.3.3	Characterisation of STT events and associated fluxes	169
4.3.4	Biomass burning influence	171
4.3.5	Classifying synoptic conditions during STT events	172
4.4	STT event climatologies	172
4.5	Simulated ozone columns	176
4.6	Stratosphere-to-troposphere ozone flux from STT events	180
4.6.1	Method	180
4.6.2	Results	182
4.6.3	Comparison to literature	182
4.7	Sensitivities and limitations	187
4.7.1	Event detection	187
4.7.2	Flux calculations	188
4.8	Conclusions	189
4.9	Contributions and Acknowledgements	190
5	Summary and Concluding remarks	191
5.1	Outcomes	191
5.2	Isoprene emissions	192
5.3	Ozone over Australia	192
5.4	Outputs and future work	194

List of Abbreviations

CTM	Chemical Transport Model
DOAS	Differential Optical Absorption Spectroscopy
HCHO	Formaldehyde
(S,P)OA	(Secondary, Primary) Organic Aerosols
PAN	PeroxyAcetyl Nitrate
PM	Particulate Matter
STT	Stratosphere to Troposphere Transport
(NM)VOC	(Non-Methane) Volatile Organic Compounds

Physical Constants

Speed of Light $c = 2.997\,924\,58 \times 10^8 \text{ m s}^{-1}$ (exact)

List of Symbols

C_x	mixing ratio or mole fraction of gas x	molecules of x per molecule of air
n_x	Number density of gas x	molecules of x per unit volume of air
Ω	Vertical column	molec cm ⁻²

Chapter 1

Introduction and Literature Review

1.1 The atmosphere

The atmosphere is made up of gases held to the earth's surface by gravity. These gases undergo transport on all scales, from barbecue smoke being blown about the garden, to smoke plumes from forest fires travelling across the world and depositing in the Antarctic snow. They take part in innumerable chemical reactions along the way, largely driven by solar input and interactions with each other. Many gases are emitted into the atmosphere by soil, trees, factories, cars, seas and oceans. They are also deposited back to the surface both directly and in rainfall.

The atmosphere is made up of nitrogen (N_2 : ~ 78%), oxygen (O_2 : ~ 21%), and argon (Ar : ~ 1%), along with water (H_2O) and *trace gases* (those that make up less than 1% of the atmosphere). Water (H_2O) ranges from 0.001 to 1% depending on evaporation and precipitation. Beyond these major constituents the atmosphere has a vast number of trace gases, including carbon dioxide (CO_2 : ~ 0.4%), ozone (O_3 : .000001 to 0.001%), and methane (CH_4 : ~ 0.4%) (BrasseurJacob2017). Trace gases in the atmosphere can have a large impact on conditions for life on earth. They react in complex ways with other elements (anthropogenic and natural), affecting all surface ecosystems upon which life depends.

One important trace gas is ozone (O_3), which affects climate, human health, and ecosystem productivity. This thesis will focus on ozone in the troposphere, which is relatively uncertain over Australia.

Ozone in the lower atmosphere is a serious hazard that causes health problems (Hsieh2013), damages agricultural crops worth billions of dollars (Avnery2011; Yue2017), and increases the rate of climate warming (IPCC_2013_chap8). Around 5 to 20 percent of all air pollution related deaths are due to ozone (Monks2015), roughly .8 million deaths per year (Lelieveld2013). In the short term, ozone concentrations of ~50-60 ppbv over eight hours or ~80 ppbv over one hour are agreed to constitute a human health hazard (Ayers2006; Lelieveld2009). Long term exposure causes problems with crop loss and ecosystem damage (Emberson2003), and concentrations may get worse in the future (Lelieveld2009; Stevenson2013). Further tropospheric ozone enhancements are projected to drive reductions in global crop yields equivalent to losses of up to \$USD₂₀₀₀ 35 billion (equivalent to US dollars in the year 2000) per year by 2030 (Avnery2011), along with detrimental health outcomes equivalent to ~\$USD₂₀₀₀ 11.8 billion per year by 2050 (Selin2009). Recently Yue2017 showed that the net effect of near-surface ozone on is a ~ 14% decrease in net primary productivity in China. They



FIGURE 1.1: Pressure (red) logarithmically decreasing, shown with percentage of atmosphere below at several points. Temperature (green) changes throughout the atmosphere. Figure edited from <https://climate.ncsu.edu/edu/Structure>.

state that reducing this decrease by $\sim 70\%$ before 2030 would require drastic measures.

1.1.1 Structure

Most of the atmosphere ($\sim 85\%$) is within 10 km of the earth's surface. This is due to air pressure, which decreases logarithmically with altitude. Any entity is subjected to the weight of all the air above it, and the density of the atmosphere is driven by this pressure.

The atmosphere extends above the earth's surface to the edges of space. This is split into various layers, defined by the *lapse rate*: the decrease in temperature (T) with increasing altitude (z), or $\frac{-dT}{dz}$. Figure 1.1 shows the pressure and temperature profiles against altitude through the atmosphere. The first layer is the troposphere, which extends to roughly 10 km and is characterised by positive lapse rate (or decreasing temperature with altitude). At the top of the troposphere (the tropopause) the temperature stops decreasing, and then the stratosphere is defined by a negative lapse rate. This is due to UV radiation being absorbed by ozone, and leads to a very vertically stable environment.

In addition to these atmospheric layers, the troposphere can be subset into the *boundary layer* and the *free troposphere*. The *boundary layer* is the lowest layer and involves increased atmospheric mixing due to ground heating and friction effects. It generally extends anywhere from 200 - 1000 m, above which the ground effects have

fewer direct impacts. The *free troposphere* is the remainder of the troposphere and is more affected by transport, both horizontally and from the stratosphere.

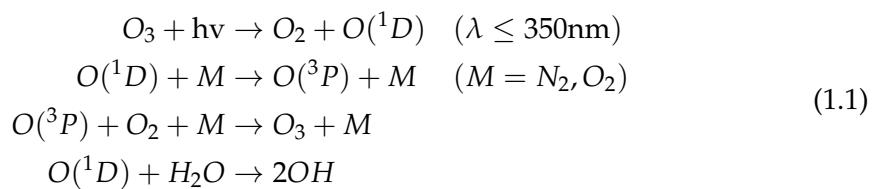
1.1.2 Composition and chemistry

There are a myriad of trace gases in the atmosphere, emitted by plants, animals, earth, and water. These gases react with one another and over time they either deposit back onto the earth or form more stable compounds such as CO₂. Oxidation and photolysis (the process of being broken apart by photons) are the two main processes whereby compounds are broken down in the atmosphere.

OH and HO₂ concentrations largely determine the oxidative capacity of the atmosphere. Concentration of the OH radical drives many processes in the atmosphere, especially during the day when photolysis of ozone produces OH (**Atkinson2000**). OH is a key species which reacts with nearly all the organic compounds in the troposphere, with only a few exceptions (**Atkinson2000**). Over land, isoprene (C₅H₈) and monoterpenes (C₁₀H₁₆) account for 50% and 30% of the OH reactivity respectively (**Fuentes2000**).

Since radicals are involved in all oxidative chemistry in the atmosphere it is important for models to accurately represent them (eg. **Travis2014**). This is difficult as they are coupled with so many other species and measurements of OH are not readily available on a global scale. In the late '90s it was thought that OH radicals were formed exclusively from photolysis of O₃, HONO, HCHO, and other carbonyls (R₂C=O) (**Atkinson2000**). It has been shown since that OH is recycled in various processes. Isoprene (C₅H₈) was thought to be a sink of OH until it was shown by **Paulot2009b** that the radicals are recycled. This recycling process is discussed in more detail in section [1.3.3](#).

Ozone is an important precursor to OH, as excited oxygen atoms (O(¹D)) are created through its photolysis, which then go on to react with water to form OH, as shown in this reaction sequence (**Atkinson2000; AtkinsonArey2003**):



Where $h\nu$ represents radiation and M is an inert molecule. This shows that some of the O(¹D) recycles back to ozone, while some forms OH.

1.1.3 Radiative Forcing

One of the larger uncertainties in atmospheric modelling is how particles in the atmosphere affect radiative forcing. For 12 years it has been understood that aerosols overall cool the atmosphere, with smaller particles having a larger effect as they matching the wavelengths of visible light (**Kanakidou2005**). Aerosol products from gas phase emissions (or the children thereof) play an indirect and complex role in cloud properties, with a net cooling effect (**Kanakidou2005, IPCC_AR5_WG1**).

Transport and indirect effects complicate matters further, with cloud creation and modification of cloud properties being quite difficult to accurately predict. In the third IPCC report ([IPCC2001](#)), the uncertainty involved if OA forcing was a factor of 3 times the estimated effect. This has since been improved however OA and cloud formation still remains a large uncertainty in more recent IPCC reports ([IPCC_Chapter2](#)). Figure [1.2](#) shows the radiative forcing (RF) of various atmospheric constituents, it is clear that OA uncertainty dominates. Figure [1.3](#) shows the same summary updated in chapter 8 of the fifth report, where the SOA uncertainty remains large.

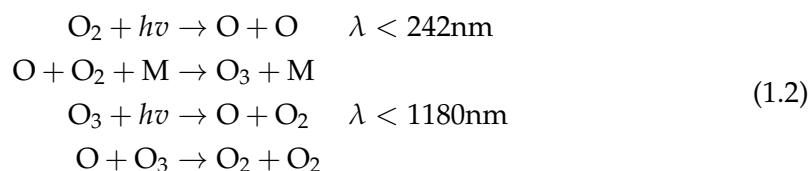
It has been known for quite a while that our understanding of VOC emissions needs to be improved in order to better capture radiative forcing ([Kanakidou2005](#)). VOC emissions affect ozone along with several atmospheric parameters which directly and indirectly alter radiative forcing rates ([Arneth2008](#)) VOCs can lead to changes in cloud formation, as nucleation can arise from the subsequent SOA. [Kanakidou2005](#) concluded that it is very likely that organics contribute to particle growth and formation rates, and that satellite datasets should be used to improve emissions inventories. This is even more important in Australia where VOCs are so poorly represented by contemporary modelling ([Emmerson2016](#)).

1.2 Ozone

Ozone (O_3) is an important greenhouse gas and oxidant. It is mostly located in the stratosphere and prevents much of the shorter wavelength (UV) solar radiation from reaching the earth's surface. Ozone in the troposphere is less beneficial, leading to health issues, radiative forcing ([Stevenson2013](#)), and crop death. Understanding and accurately portraying ozone concentrations in the troposphere is important to allow accurate predictions of future climate. This will become even more important as projections of future climate changes suggest altered vertical mixing rates, ultra violet index and ozone RF ([Hegglin2009](#)).

1.2.1 Stratospheric ozone

In the stratosphere ozone production is driven by the Chapman mechanism, as high energy radiation (with wavelengths $\lambda < 242$ nm) photolyses the molecular oxygen (O_2) in the atmosphere ([BrasseurJacob2017](#)). The Chapman mechanism involves several reactions which lead to rough equilibrium of O , O_2 , O_3 and pressure, as follows:



The high energy photons ($\lambda < 242$ nm) are present from the top of the atmosphere but are mostly removed before reaching the troposphere as their energy is used to split the O_2 molecules. The lifetime of O against loss by O_2 is less than a second in the troposphere, and produced O_3 quickly returns to O and O_2 , as low energy ($\lambda < 1180$ nm) photons and M are abundant. The reduced light penetration towards the surface, in



FIGURE 1.2: The overall radiative forcings and uncertainties of several atmospheric constituents This is an image taken from **IPCC_Chapter2**, found at https://www.ipcc.ch/publications_and_data/ar4/wg1/en/faq-2-1.html.

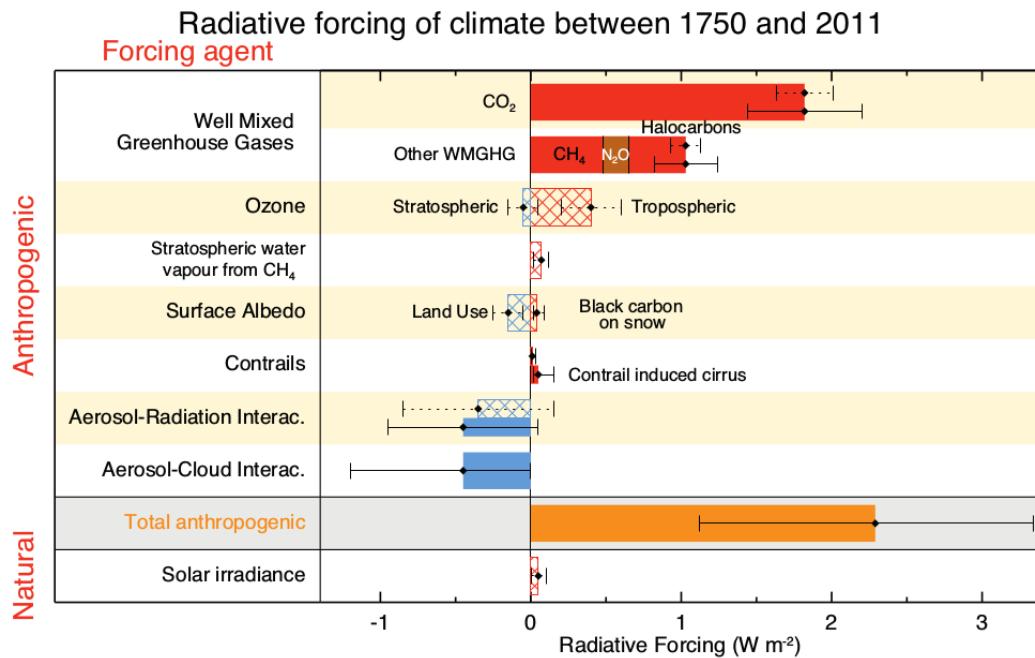


FIGURE 1.3: The overall radiative forcings and uncertainties of several atmospheric constituents This is an image taken from [IPCC_AR5_WG1](#), chapter 8.

addition to the logarithmic increase in atmospheric pressure (which affects M abundance) drives the vertical profile of ozone into what is called the *ozone layer*. This is a layer of relative ozone abundance within the stratosphere. The Chapman mechanism requires radiation so only takes place during the daytime, during the night this process slows to a halt, and the ozone concentrations remain stable unless pollution intrudes ([Jacob_1999_book](#)).

Since the Montreal Protocol on Substances that Deplete the Ozone Layer was established in August 1987, and ratified in August 1989, several satellites and many measurement stations were set up to monitor ozone in the stratosphere. However, in the southern hemisphere there are relatively few records of ozone ([Huang2017](#)). This affects our ability to accurately determine sources of ozone in the troposphere, with current southern hemisphere trends lacking full explanation ([Zeng2017](#)).

1.2.2 Tropospheric ozone

Figure 1.4, copied from [Young2018](#), shows summary of the major processes and emissions affecting tropospheric ozone. This thesis involves improving the highly uncertain natural emissions of volatile organic compounds (VOCs) from Australia, and estimating impacts from STEs.

Generally there are two main drivers of tropospheric ozone concentrations; transport from the stratosphere and chemical production due to emissions of precursors.

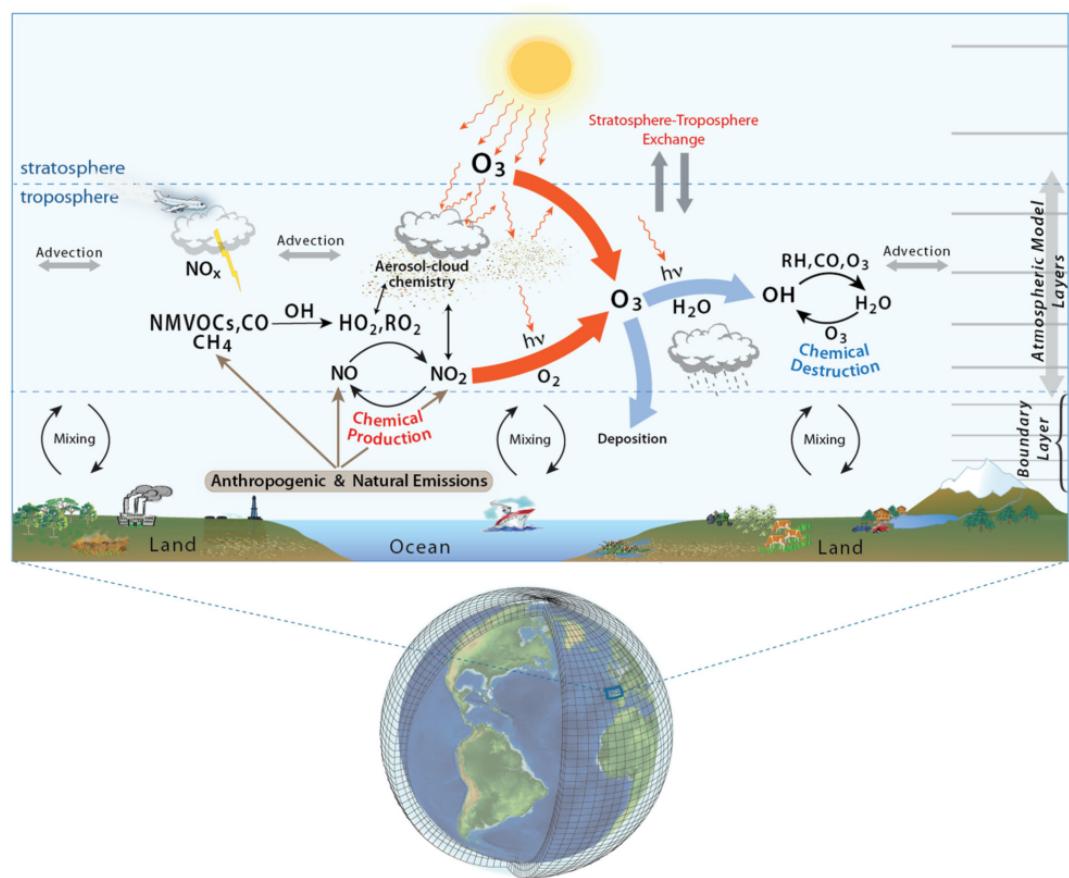


FIGURE 1.4: Tropospheric ozone processes, Figure 1 in Young2018.
DOI: <https://doi.org/10.1525/elementa.265.f1>

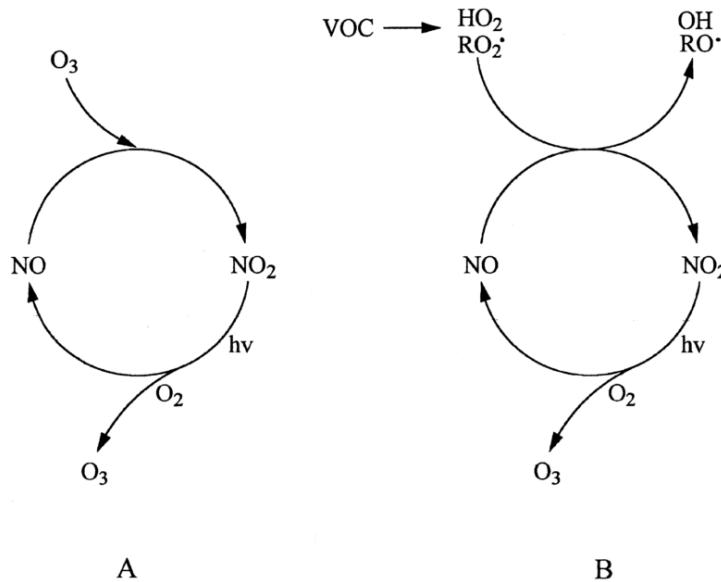
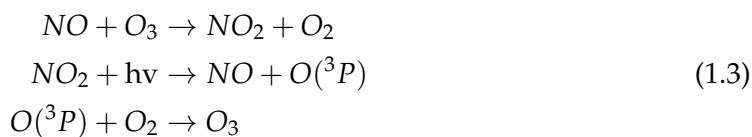


FIGURE 1.5: Figure showing NO, NO_2 , and Ozone photoequilibrium cycle with and without (B, A respectively) influence from VOCs. Figure copied from **Atkinson2000**.

Tropospheric ozone is regulated by NO and NO_2 concentrations, which form an equilibrium (**Cape2008; Young2018**). At small to medium scales, pyrogenic (fire) and anthropogenic (man-made) emissions can be important. Smoke plumes from biomass burning can carry ozone precursors, creating higher ozone concentrations downwind of the plume's source. Emissions of precursors from large cities (primarily traffic and power production) can impact ozone concentrations. These impacts are not always straightforward due to the nonlinear relationship between ozone and its precursors.

NO_X ($\equiv \text{NO}_2 + \text{NO}$) is another important chemical family in the atmosphere which interacts with ozone and regulates the atmospheric oxidative capacity. NO_X or VOC emissions affect the tropospheric ozone equilibrium and can lead to enhanced ozone formation, shown in figure 1.4. NO_X compounds are short lived, with emissions (Power generation and combustion transport) being the main driver of concentrations (**Delmas1997**). NO_X is removed primarily by conversion to nitric acid (HNO_3) followed by wet or dry deposition (**Ayers2006**).

NO_X and O_3 relative concentrations during the day are regulated by the following reactions (**Sillman1999; Atkinson2000**):



This process with and without the influence of VOCs (panel A and B respectively) is summarised in figure 1.5.

1.2.3 Stratosphere to troposphere transport

Historically, ozone transported down from the stratosphere was thought to contribute 10–40 ppb to tropospheric ozone levels, matching tropospheric production (**Atkinson2000; Stohl2003**). The proportion of tropospheric ozone due to transport from the stratosphere was revised down to around 10% over the years as measurement and modelling campaigns improved our understanding of global scale transport, mixing, and chemistry (**Guenther2006; Monks2015**). Intrusions of stratospheric air into the troposphere are often called Stratosphere to Troposphere Transport (STT) events. Although most tropospheric ozone comes from production, STT enhancements of ozone are measurable and can be regionally important (**Jacobson2000; Lelieveld2009; Kuang2017**), and upper tropospheric ozone can be transported long distances (**Cooper2004**). An analysis of the Atmospheric Chemistry and Climate Model Inter-comparison Project (ACCMIP) simulations by **Young2013** found STT is responsible for $540 \pm 140 \text{ Tg yr}^{-1}$, equivalent to $\sim 11\%$ of the tropospheric ozone column (**Monks2015**).

Ozone transported to the troposphere from the stratosphere can occur through diffusion (relatively slowly) or direct mixing (as STT). STT often occur as tongues of stratospheric air descend and get disconnected from the stratosphere, potentially due to low pressure systems and jet streams (**Sprenger2003**). Recently global chemical transport models (CTMs) have been used to trace how much ozone is being transported to the troposphere in this manner. There are a few methods of doing this, such as modeling ozone formed in (and transported from) the stratosphere (**Ojha2016**). Model based estimates require validation against actual measurements, such as those from ozonesondes or satellites. **Hegglin2009** estimate that climate change will lead to increased STT. They posit that this is due to an acceleration in the Brewer Dobson circulation; which is the global scale model of transport of air in the troposphere and stratosphere. They estimate ~ 30 , and $\sim 121 \text{ Tg yr}^{-1}$ increases by 2095 (relative to 1965) in the southern and northern hemispheres respectively, up by 23% globally.

Liu2017 examine southern hemispheric ozone and the processes which control its inter-annual variability (IAV). IAV is the standard deviation of ozone anomalies from the monthly mean. They show that ozone transported from the stratosphere plays a major role in the upper troposphere, especially over the southern Indian ocean during austral winter. STT mostly impacts the upper troposphere, although some areas are impacted right down to the surface. **Kuang2017** found a measurable impact of STT ozone enhancement in the south east US using several different instruments. They also show how ozone depends on both the local topography, weather systems, and trace gases emitted and transported into the region. **Liu2017** examined modelled tropospheric ozone sensitivity to various meteorological parameters. They found tropospheric ozone sensitivity to emissions from South America ($0\text{--}20^\circ\text{S}$, $72.5\text{--}37.5^\circ\text{W}$), southern Africa ($5\text{--}10^\circ\text{S}$, $12\text{--}38^\circ\text{E}$), and South to South east Asia ($70\text{--}125^\circ\text{E}$, $10^\circ\text{S}\text{--}40^\circ\text{N}$). In the US recent work by **Lin2015** suggests that intrusions during spring are increasing surface ozone levels. They recommend improvements to understanding of the frequency and cause of STT are needed effectively implement air quality standards. Recently, modelled ozone concentrations has been shown to be most sensitive to NO_x sources (such as lightning and car exhaust emissions) and isoprene emissions (**Christian2018**).

1.2.4 Chemical production

Ozone produced in the troposphere from precursors and radiation drive ozone levels, especially in the lower (near-surface) troposphere. The main processes involved are shown in figure 1.4, with ozone regulated by reactions 1.3. As discussed above STTs source $\sim 11\%$ of the tropospheric column of ozone, with the remainder produced photochemically (**Monks2015**). A recent summary by **Young2018** estimates ozone production and loss in the troposphere to be $\sim 4900 \text{ Tg yr}^{-1}$, and $\sim 4500 \text{ Tg yr}^{-1}$ respectively. These numbers are at the global scale, and it should be noted that meteorology and topography can play massive roles due to large spatial variability in ozone (**Kuang2017**).

Tropospheric ozone concentrations require climate and ozone precursor emissions; including NO, NO₂, CO, and VOCs such as HCHO (**Atkinson2000**; **Young2013**; **Marvin2017**). Ozone predictions are uncertain and changing climate affects transport, deposition, destruction, and plant based precursor emissions. All of these processes are tightly coupled and difficult to accurately model, as they depend on uncertain assumptions such as CO₂ dependency (**Young2013**). Even with all the work done over the prior decades there remain large uncertainties about ozone precursors in the troposphere (**Mazzuca2016**).

Ozone is formed in the troposphere through oxidation of VOCs (described in Section 1.3) in the presence of NO_X. Net formation or loss of O₃ is determined by interactions between VOCs, NO_X, and HO_X, and is a complicated system of positive and negative feedbacks (**Atkinson2000**). Figure 1.6 shows an example of this non-linear relationship between NO_X, VOCs, and ozone production as modelled in **Mazzuca2016**. This non-linear relationship is examined in more detail in the following section (1.3). Recently the relationship has been examined on the intradiel timescale showing that ozone production can be more or less sensitive to VOCs at different hours (**Mazzuca2016**). This shows how important it is to correctly determine the precursors concentrations in order to estimate ozone levels and production.

Tropospheric ozone is lost via chemical destruction and dry deposition, estimated to be $4700 \pm 700 \text{ Tg yr}^{-1}$ and $1000 \pm 200 \text{ Tg yr}^{-1}$, respectively (**Stevenson2006**; **Young2018**). The main loss channel is through equation 1.1, where photolysis and collisions (increasing with pressure) create OH from the O₃.

1.3 Volatile Organic Compounds

The least well understood precursors to tropospheric ozone production belong to a class of organic compounds. Organic compounds are members of a large class of chemicals whose molecules contain carbon, with the exception of a few compounds such as carbides, carbonates, and simple oxides of carbon and cyanide. Organic compounds can be categorised based on their vapour pressure, which is the tendency of a liquid or solid to vaporise. Compounds with high vapour pressures at standard temperature are classed as volatile, evaporating at low temperatures. Plants contain tens of thousands of organic compounds, with fewer than 40 having high enough volatility to be emitted (**Guenther2000**). Gas phase emissions with higher vapour pressures



FIGURE 1.6: Ozone production rate dependent on NO_x and VOC concentrations (Mazzuca2016).

can be oxidised into lower vapour pressure products which will partition between gas and particle phase, often called semi or non-volatile.

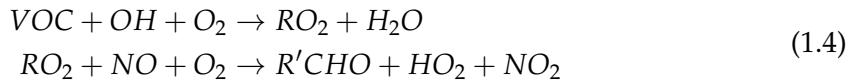
Atmospheric organic compounds are legion and differ by orders of magnitude with respect to their fundamental properties, such as volatility, reactivity, and cloud droplet formation propensity, etc. Volatile organic compounds (VOCs) have vapour pressure greater than 10^{-5} atm, and are mostly generated naturally by plants, which emit around 1000 Tg yr^{-1} (Guenther1995; Glasius2016). Due to their high volatility these compounds generally exist in the gas phase. Organic compounds with a lower volatility are classed as semi-volatile (SVOCs: vapour pressure between 10^{-5} and 10^{-11} atm) are found in both gas and particle phase depending on temperature and pressure. Organic compounds with even lower vapour pressure are generally found in the particle phase in aerosol particulate matter (Glasius2016). Understanding the drivers of trends in biogenic VOC emissions (BVOCs) is required in order to estimate future carbon fluxes, changes in the water cycle, ozone production, air quality, and other climate responses (Yue2015). In the last 20 years anthropogenic emissions of VOCs have been increasing while biogenic VOC emissions have decreased, due to rapid economic growth and lower annual temperatures (Stavrakou2014; Kwon2017).

Methane (CH_4) is one of the more abundant VOCs, however it is often classified separately and compared against non-methane VOCs (NMVOCs). NMVOCs include alkanes, alkenes, and aromatic hydrocarbons, with isoprene (an alkene) being the most abundant (Guenther1995). Methane is relatively long lived (years) and is well mixed in the atmosphere while other VOC levels are spatially diverse due to their shorter lifetimes. In this thesis I work towards a better understanding of the isoprene emissions coming from Australia.

VOCs are an important driver of atmospheric processes, especially near forests.

VOCs are broken down into HCHO, O₃, CO₂ and many other species, mainly through oxidation by OH. VOC emissions result in radical cycling, acid deposition, production of tropospheric ozone, and secondary organic aerosols (SOAs) (**Atkinson2000; Kanakidou2005**). VOC emissions affect surface pollution levels, potentially enhancing particulate matter (PM) and ozone levels. A regional-model study in Europe (**Aksyoglu2017**) has also shown VOCs impact secondary inorganic aerosol concentrations. These have impacts on climate (through radiative forcing) and air quality (from ozone and SOA enhancements), affecting both human health and crop yields (**IPCC_Chapter2; Avnery2011; Lelieveld2015**).

Ozone in rural areas is often higher than in populous cities, due to titration (removal) of ozone by NO in polluted areas (**Cooper2014; Monks2015**). In areas with high VOC concentrations, ozone production may be enhanced through the following reaction sequence (**Sillman1999**):



with R and R' representing organic species. The reactions of VOCs or CO with OH convert NO to NO₂, which leads to ozone formation as NO₂ production in reaction 1 of [1.3](#) is bypassed.

One aspect associated with VOC emissions is the production of aerosols. Aerosols are suspended particulates and liquid compounds in the atmosphere, often called particulate matter (PM). PM in the atmosphere is a major problem, causing an estimated 2-3 million deaths annually (**Hoek2013; Krewski2009; Silva2013; Lelieveld2015**). Fine particulate matter (PM_{2.5}) penetrates deep into the lungs and is detrimental to human health. Some PM comes from small organic aerosols (OA) emitted in the particulate phase and referred to as primary OA (POA).

A substantial amount of PM is due to gaseous organic compounds transforming in the troposphere leading to what is known as secondary OA (SOA) (**Kroll2008**). Formation of SOA is generally due to VOC oxidation and subsequent reactions, while removal from the atmosphere is largely due to wet or dry deposition, and cloud scavenging (**Kanakidou2005**). It can be difficult to attribute the formation of SOA, in part due to the complex relationship between NO_x, OH, O₃, and the uncertainty surrounding precursor emissions. Most of the tropospheric SOA comes from biogenic precursors, the evidence for this has grown over the last two decades (**Guenther1995; Kanakidou2005; Guenther2012**). Improved concentration estimates of these precursors requires a better understanding of their emissions, which is one of the foci in this thesis.

Photolysis and oxidation of many VOCs initially form alkyl radicals (R'). VOCs are removed mainly by photolysis and oxidation, but also by wet and dry deposition, reaction with NO₃, and ozonolysis (at night time or in polluted areas) (**AtkinsonArey2003; Brown2009**). The process of deposition only accounts for a small fraction of the VOC loss, with the possible exception of the long lived methane compound (**AtkinsonArey2003**).

1.3.1 Emissions

VOC emissions are classified as either anthropogenic, biogenic (BVOC), or pyrogenic. Global VOC levels are estimated at 85 %, 13 %, and 3 % from biogenic, anthropogenic, and pyrogenic sources respectively (**Kanakidou2005; Kefauver2014**). Methane makes up a third of atmospheric VOCs and is relatively ubiquitous due to its longer lifetime, non-methane VOCs (NMVOC) are often grouped together. Due to the lack of in-situ ground based measurements, estimates of VOC emissions are uncertain, with large scale extrapolation required (**Millet2006**). The ocean also plays a role in VOC emissions, with the Oceanic Niño Index (ONI) showing positive VOC emission anomalies associated with neighbouring countries (**Stavrakou2014**).

The main non-methane BVOC emissions are isoprene (44%) and monoterpenes (11%) (**Guenther2000; Kefauver2014**). There are ten times the mass of NMVOCs from natural sources as there are from anthropogenic sources (**Guenther2006; Kanakidou2005; Millet2006**). Major emitters are broadleafs (notably Eucalyptus), and shrubs (**Guenther2006; Arneth2008; Niinemets2010; Monks2015**). NMVOC emissions are a byproduct of photosynthesis, and also a response to environmental conditions such as drought or sunlight. Emissions are affected by many factors including temperature, atmospheric CO₂, soil moisture, drought stress, etc. Land use changes can drastically affect isoprene sources, for instance in the tropics where large scale deforestation has converted forest into crop lands (**Kanakidou2005**). In this thesis I focus on emissions of isoprene in Australia.

Globally around 710 - 1150 Tg C yr⁻¹ of BVOCs are emitted (**Guenther1995; Lathiere2006; Guenther2012; Lathiere2016**). 90% of these emissions come from plants and trees, with the most dominant species being isoprene (C₅H₈) (~ 50%), monoterpenes (C₁₀H₁₆), methanol (CH₃OH), ethanol (C₂H₆O), acetdehyde (CH₃CHO), acetone ((CH₃)₂CO), ethene (C₂H₄) and propene (C₃H₆) (together making up ~ 30%) (**Guenther2012**). Many of these estimates come from MEGAN, a bottom-up biogenic emissions model which is highly sensitive to several parameters including soil moisture and plant functional type. MEGAN “is a modelling framework for estimating fluxes of biogenic compounds between terrestrial ecosystems and the atmosphere to account for the major known processes controlling biogenic emissions” (**Guenther2012**). It allows parameterisation of various BVOC emissions, with descriptions given in **Guenther2012**. MEGAN has recently been analysed using 30 years of meteorological reanalysis information by **Sindelarova2014**. They estimate emissions of BVOCs to be 760 Tg C yr⁻¹, 70% (532 Tg C yr⁻¹) of which is isoprene. This is similar to isoprene emission estimates from MEGAN itself, of 400-600 Tg C yr⁻¹ (**Guenther2006**). Another model (ORCHIDEE, with inputs similar to MEGAN) estimates 752 ± 16 Tg C yr⁻¹, sensitive to terrestrial vegetation variations (**Lathiere2006**). MEGAN emissions estimates are termed bottom-up, as opposed to top-down which are derived from satellite measurements of the products of various VOCs. Using GOME satellite HCHO and a Bayesian inversion technique to derive isoprene emissions, **Shim2005** estimated global isoprene emissions to be ~ 566 Tg C yr⁻¹. This estimate decreases simulated OH concentrations by ~ 10%, to 9.5e5 molec cm⁻³.

1.3.2 Isoprene

Isoprene, or 2-methylbuta-1,3-diene, is a VOC with the chemical formula C₅H₈. It is of major importance to the atmosphere, as it is involved in various processes which alter the oxidative capacity of the atmosphere. Isoprene affects NO_x and HO_x cycling, see for example formulae 1.1, 1.3. In the presence of NO_x, isoprene forms tropospheric ozone and SOAs (**Wagner2002; Millet2006**). It has a short lifetime during the day, roughly an hour due to OH oxidation (**AtkinsonArey2003**).

Measurements of isoprene are often uncertain or difficult to make accurately. Chamber experiments are used to determine how isoprene behaves once it is emitted into the atmosphere, however reaction rates may be unsuitable to the natural atmosphere which is often very different (**Kanakidou2005; Nguyen2014**). Improving chamber study methods could improve understanding of ambient atmospheric oxidation mechanisms of isoprene (and other organic hydrocarbons), which could reduce some of the high uncertainties involved with VOC chemistry (**Nguyen2014**). Uncertainties in measurements exist both structurally (between different techniques) per measurement due to the difficulty of detecting isoprene and its high reactivity.

Guenther1995, and subsequent updates (**Guenther2000; Guenther2006; Guenther2012**), have been used ubiquitously by the atmospheric community as a global estimate of isoprene emissions, at roughly 500-600 Tg yr⁻¹, emitted mostly during the day. Recently an estimate of global isoprene emissions, of around 465 Tg C yr⁻¹, has been made using a completely different model (**Messina2016**). The global emission factors used to derive both these estimates are based on modelling emissions from different plant species (phenotypes), and relatively few Australian species are used when forming in these estimates. This leads to increased uncertainty for Australian emissions estimates. Due to the highly reactive nature of isoprene, modelling is sensitive to uncertainties, for example the diurnal pattern of isoprene emissions affects modelled ground level ozone (**Hewitt2011; Fan2004**).

Isoprene emissions estimates are still fairly uncertain, as global measurements are difficult and regional emissions and chemistry can be very different. The global uncertainty of isoprene emission was estimated to be a factor of 2 to 5 (250-750 Tg yr⁻¹) (**Kanakidou2005**). Improvements over the years have been incremental, and generally localised to regions of particular interest for air quality such as China and the USA (**Guenther2012; Jiang2018**). The lack of accuracy in BVOC emissions measurements (in general) prevents accurate determinations of the sources and distribution of pollutants including ozone and organic aerosols.

1.3.3 Isoprene chemistry

Isoprene forms many products with various lifetimes, here I will present an overview of some important mechanisms and products. Isoprene is emitted and enters the atmosphere in the gas phase, where it reacts quickly with OH and other radicals. One common compound which is produced by these reactions is HCHO, which is easier to measure and often used to estimate how much isoprene is being emitted. Alkenes (VOCs with double bonded carbon, such as isoprene) react with OH, ozone, or NO₃, leading to organic peroxy radicals (ROO[·]). These go on to form many products and

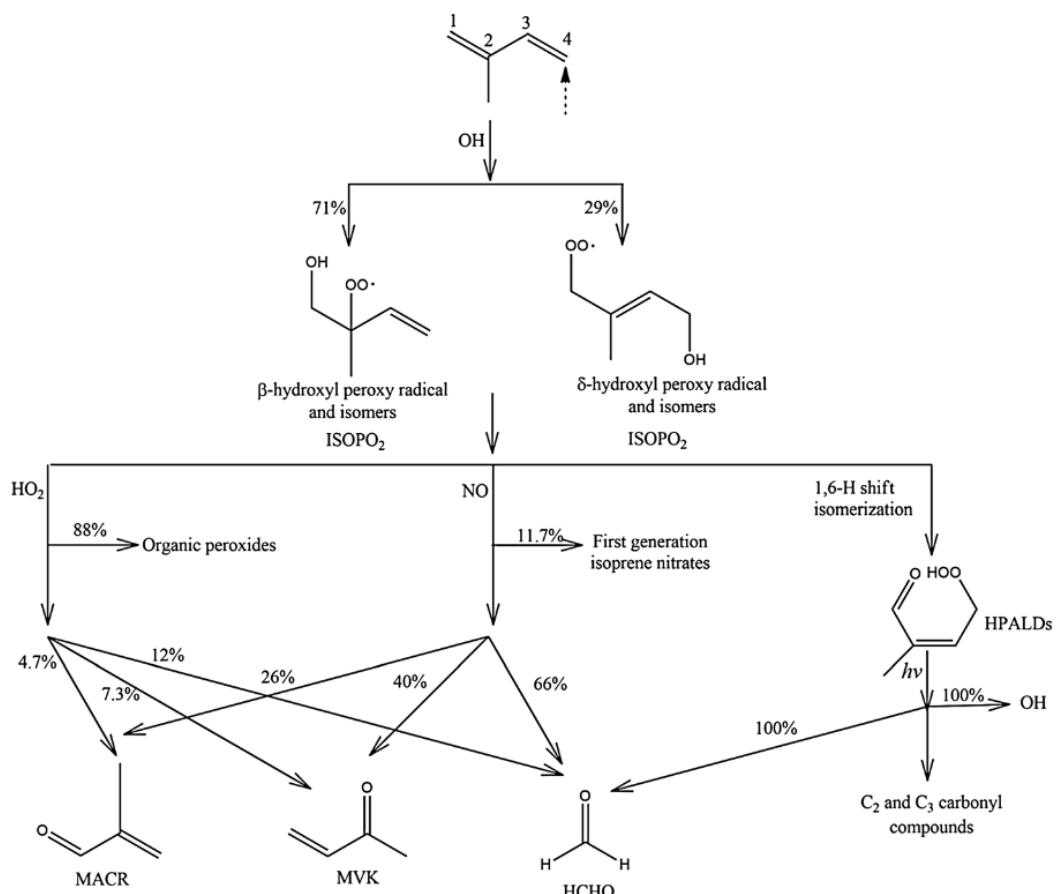
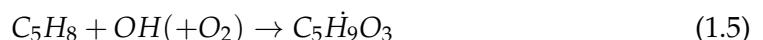


FIGURE 1.7: Isoprene products following oxidation by OH, figure from Mao2013

lead to (amongst other things) aerosol, formaldehyde, and ozone formation, depending on sunlight and NO_x concentrations (Atkinson2000). Reactions with NO can lead to ozone production within environments rich in isoprene or other NMVOCs (Patchen2007; AtkinsonArey2003).

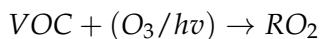
Figure 1.7 shows the first stages of oxidation of isoprene by OH. Isoprene reactions are important to understand due to their impacts on air quality, ozone, and physical properties in the lower troposphere. The primary first step for atmospheric isoprene is photooxidation, reacting with OH to form isoprene hydroperoxy radicals (ISOPOO - a subset of ROO)) (Wolfe2016; Marvin2017; Patchen2017). This is largely split into two types of ISOPOO, based on which carbon the OH adds to (see figure 1.7).



The many children processes and products which begin with isoprene oxidation are often called the isoprene (photochemical) cascade (Crounse2012; Paulot2012; Wolfe2016).

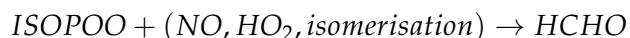
1.3.3.1 Oxidation

The primary sink for isoprene is oxidation by OH. First isoprene has its double bond replaced by OH, as summarised by the equation from **Patchen2007**: $R-CH=CH-R' + OH \longrightarrow R-CH(OH)CH-R'$ where R and R' represent hydrocarbons. Ozonolysis and photolysis are lesser oxidation pathways for volatile alkenes, involving the splitting of carbon chains by ozone molecules or photons respectively:



(**Nguyen2016; Wolfe2016**). Ozonolysis also leads to HCHO, with yields depending on subsequent reactions.

After oxidation by OH, the adducted OH then reacts with O_2 to produce ISOPOO, which can be any of six different isomers (**Patchen2007**). ISOPOO reacts with HO_2 or NO, producing stable products (often called oxidised VOCs or OVOCs). One important product produced (with varying yields) through each oxidation pathway is HCHO:



During the day HCHO has a lifetime of 1-2 hrs, while ROO lasts ~ 100 s, making reaction **1.5** a rate limiting factor in HCHO production (**Wolfe2016**). ISOPOO also can isomerise and produce HPALDS (see figure **1.7**), which also leads to HCHO. At higher NO mixing ratios (at least a few hundred pptv), ROO react mostly with NO. At low NO (less than 50 pptv), ROO is more likely to either isomerise, or react with HO_2 , or another ROO .

There is uncertainty about which pathways are most important following ISOPOO production, affecting predictions by atmospheric models (**Nguyen2014**). This limits understanding of the relative importance of some chemical processes, such as auto-oxidation (of ISOPOO and other ROO) (**Crounse2013**). The reaction pathways depend on local concentrations of NO_x : the high and low NO_x pathways are dominated by NO and HO_2 reactions respectively. HO_2 reactions predominantly produce hydroxyhydroperoxides (ISOPOOH), while NO reactions produce isoprene nitrates (ISOPN) (**Crounse2006**). If measured, first generation ISOPN and ISOPOOH products can be used to determine the portion of isoprene oxidation following each pathway (**Yu2016**). Globally around one third of ISOPOO react with HO_2 , and two thirds react with NO (**Paulot2009b**). Most of these reaction pathways produce HCHO, however this along with methyl vinyl ketone (MVK), and methacrolein (MACR) are formed at different yields between the two pathways (**Marais2012; Liu2016a; Wolfe2016**).

1.3.3.2 High NO_x pathway

In the presence of NO_x , ISOPOO reacts with NO and forms ISOPN, which affect levels of both HO_x (H, OH, peroxy radicals) and NO_x . ISOPN generally act as a sink of HO_x , and can be a sink or reservoir for NO_x (**Mao2013**). A portion of the ISOPN are recycled back to NO_x , serving as a reservoir of nitrogen and allow its transport to the boundary layer of remote regions (**Patchen2007; Paulot2009a; Yu2016**). The nitrates can also build up in the winter, when removal processes are not as dominant

(**Lelieveld2009**). Reactions of OH with NO_x are the main radical sink in high-NO_x systems (**Wolfe2012**).

First generation ISOPN produce MVK($\sim 40\%$), MACR($\sim 26\%$), and HCHO($\sim 60\%$) at higher yields than is produced by ISOOPOH (**Liu2013; Mao2013**). The MVK and MACR products form additional HCHO within a few hours due to oxidation by OH (**Palmer2006**). Under high NO_x conditions there is a higher and faster yield of HCHO, with most of the ultimate HCHO production occurring within one day (**Palmer2006**).

1.3.3.3 Low NO_x pathway

In low NO_x environments, ISOOPOH is formed in yields $> 70\%$, while MACR, MVK, and HCHO are formed at $\sim 5\%$, $\sim 7\%$, and $\sim 12\%$ respectively (**Paulot2009b; Mao2013**). This ISOOPOH mostly reacts with OH to form IEPOX while regenerating OH (**Mao2013**). This pathway has lower and slower ultimate yields of HCHO from isoprene emissions when compared to the high-NO_x pathway (**Palmer2006**).

Isoprene oxidation and subsequent reactions are less well understood when lower concentrations of NO are present in the atmosphere. It was thought that in low NO environments, like those far from anthropogenic pollution and fires, oxidation of isoprene would create ISOOPOH and reduce local concentrations of OH and HO₂ (**Guenther2000; Paulot2009b**). However this reduction was not seen in measurements and HO_x levels have been shown to be largely unaffected by isoprene concentrations (**Paulot2009b**). HO_x is recycled through dihydroxyperoxides (IEPOX), formed from ISOOPOH oxidation, and some HO_x is produced in the formation of MACR and MVK (**Paulot2009b**). **Paulot2009b** estimated that $95 \pm 45 \text{ Tg yr}^{-1}$ of IEPOX was being created in the atmosphere, which (at the time) was not modelled by CTMs. **Peeters2010** suggested that the work of **Paulot2009b** only partially bridges the gap between clean air OH concentration measurements and models. They suggested four new mechanisms for OH recycling in these pristine conditions. These can be summarised as OH regenerating reactions which occur during photolysis of hydroperoxy-methyl-buteneals (HPALDs), and resulting photolabile peroxy-acid-aldehydes (PACALDs). These reactions are highly non-linear and subject to large uncertainty, however they were shown to improve modeled HO_x concentrations against several campaigns. **Peeters2010** showed that HO₂ is produced at near unity yields following isoprene oxidation initiated by OH. Their results were backed up by observations of OH recycling observed in low NO conditions (**Crounse2012**).

Uncertainties and bias from measurements have made it more difficult to understand what happens in low NO_x conditions as many observations of OH were still quite under-predicted in models (**Mao2012**). Due to OVOC interference, measurements in low NO_x environments can lead to massively overestimated MVK and MACR yields (**Nguyen2014**). **Nguyen2014** show preliminary estimates of low-NO yields of MVK and MACR to be $6 \pm 3\%$ and $4 \pm 2\%$ respectively, consistent with **Liu2013**, but only when cold-trapping methods are employed. **Mao2012** showed that many instruments were generating OH internally, creating anomalous VOC readings due to within-instrument oxidation.

Improved understanding of both the chemistry and instrument sensitivities has helped closed the gap between model predictions and detected concentrations of VOCs

and OH (**Mao2012**). But even with the recent boom in analysis, uncertainties remain in isoprene oxidation mechanisms. Examples (taken from **Nguyen2014**) include isoprene nitrate yields, which range from 4-15% (**Paulot2009a**), 90% disagreements in MACR and MVK yields (**Liu2013**), various possible sources for SOA (**Chan2010; Surratt2010; Lin2013**), unknown HPALD fates, incomplete O₂ incorporation (**Peeters2009; Crounse2013**), and under-characterised RO₂ lifetime impacts (**Wolfe2012**).

1.3.3.4 Night time processes

At night when OH concentrations have dropped, isoprene can remain in the atmosphere. Typically less than half of this night time isoprene is removed through ozonolysis (**AtkinsonArey2003**), however, in polluted areas where high levels of NO_X exist, isoprene is consumed by nitrate radicals (NO₃), which joins to one of the double bonds and produces organic nitrates in high yield (65% to 85%) (**Mao2013**). NO₃ are largely formed through ozone reactions, as in equation 1.3. A build up of NO₃ radicals can be seen at night, when photolysis is not removing them (**Atkinson2000; Brown2009**).

In areas with high NO_X levels, greater than 20% of the isoprene emitted late in the day ends up being oxidised by the NO₃ radical overnight (**Brown2009**). At night isoprene affects on both NO_X concentrations and ozone levels, and can form harmful organic nitrates and SOAs (**Brown2009; Mao2013**). These nitrates go on to produce further SOAs, largely due to NO₃ reacting with first generation isoprene oxidation products (**Rollins2009**). The night-time concentrations of OH and ozone also have a complex effect on NO_X removal in high latitude winters, when photolysis and NO reactions are reduced (**Ayers2006**).

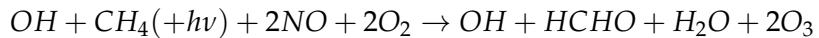
1.4 Formaldehyde

Formaldehyde (HCHO), aka methanal, methyl aldehyde, or methylene oxide, is of the aldehyde family. HCHO is an OVOC which is toxic, allergenic, and a potential carcinogen. In this thesis HCHO is used to estimate isoprene emissions over Australia. One of the major products of isoprene chemistry is HCHO. HCHO is important both for its own atmospheric impacts, and as a proxy for determination of isoprene emissions. Given a modelled yield of HCHO from isoprene, it is possible to work backwards from measured HCHO concentrations to determine the isoprene emissions. HCHO production does depend on NO_X concentrations, as it affects the yield from isoprene oxidation. HCHO yield is higher in the high-NO_X pathway (compared to the low-NO_X pathway) from isoprene reactions (**Marais2012**). HCHO measurements are often used as a check on how well isoprene reactions are simulated, as HCHO levels depend on initial VOCs and oxidants (which can be prescribed) (**Marvin2017**).

1.4.1 Sources and sinks

Background levels of HCHO in the atmosphere are driven by the oxidation of methane (CH₄) by the hydroxyl radical (OH), which produces $\sim 970 \text{ Tg yr}^{-1}$ (**FortemsCheiney2012**). **Atkinson2000** summarised the background formation of HCHO with the following

reaction:



which shows that photolysis and oxidation of methane forms HCHO and ozone in a process that regenerates the OH radicals. CH₄ concentrations are relatively well constrained in models, with the ACCMIP comparison showing only $\sim 3\%$ inter-quartile range (**Young2013**). There is a complex relationship between VOCs, HO_X, and NO_X: with higher levels of NO_X increasing the rate at which VOCs are converted into HCHO (**Wolfe2016**).

Within the continental boundary layer (CBL), HCHO is enhanced above background HCHO levels, due to NMVOC emissions reacting with OH radicals in the presence of NO_X (**Wagner2002; Millet2006; Kefauver2014**). The total contribution from NMVOC oxidation is $\sim 358 \text{ Tg yr}^{-1}$ (**FortemsCheiney2012**). Enhancements to regional and continental HCHO are largely driven by isoprene emissions (**Guenther1995; Palmer2003; Shim2005; Kefauver2014**). This is true except near fires or anthropogenic sources of HCHO and precursors (**Guenther1995; Kefauver2014; Wolfe2016**). Biomass burning (BB) can be a source of HCHO, and various other pollutants, precursors, and aerosols (**Guenther1995; Andreae2001**). Additionally HCHO is emitted into the atmosphere directly through fossil fuel combustion, natural gas flaring, ethanol refining, and agricultural activity (**Wolfe2016**).

Other terpenoids (monoterpenes, sesquiterpenes, etc.) can also produce HCHO, although generally to a lesser extent than isoprene, methane and biomass burning (**Guenther2012**). Many of the HCHO yields from terpenoids are estimated through chamber studies which examine molecular mass and charge after mixing the compound of choice into a known volume of air (**Nguyen2014**). These conditions generally do not match those of the real world, where ambient air will have a cocktail of these compounds and other reactants. One issue with chamber studies is the difficulty they have trying to accurately reproduce ambient outside air, which limits the scope to which the studies may be applied (**Nguyen2014**).

Anthropogenic sources of HCHO are largely negligible, however in very large cities or by using oversampling techniques an anthropogenic signal can be found (**Millet2008; Zhu2014**). If the population centres and industrial districts are large enough they can emit huge amounts of VOCs into the atmosphere (**Fu2007**), leading to increased surface ozone levels (**Zhu2014**) In Australia this is not yet a major issue, however anthropogenic sources of pollution can be detected (see section ??).

In the past, HCHO levels were underestimated by models, often with large discrepancies, due to the poor understanding of methyl peroxy radical (CH₃OO) chemistry (**Wagner2002**). Nowadays HCHO concentrations are better understood, however precursor emissions are one of the main unknowns (**Emmerson2016; Marvin2017**). **Marvin2017** found that discrepancies in modelled HCHO concentrations are primarily due to second and later generation isoprene oxidation chemistry.

HCHO has two major sinks, reactions with OH (oxidation), and photolysis (adding up to $\sim 1210 \text{ Tg yr}^{-1}$) (**Levy1972; Crutzen1999; Wagner2002; FortemsCheiney2012;**

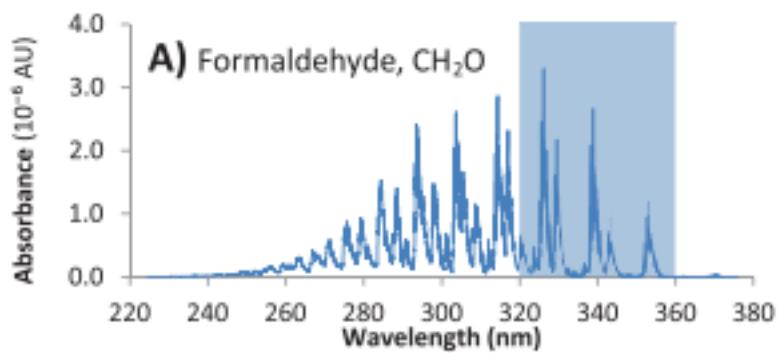
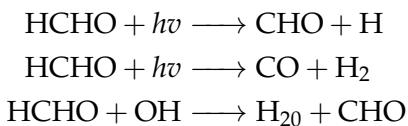


FIGURE 1.8: HCHO spectrum, with a typical band of wavelengths used for DOAS path measurements. This is a portion of an image from [Davenport2015](#).

[Kefauver2014](#)) with reactions as follows ([Ayers1997](#)):



These reactions lead to a daytime lifetime of a few hours ([Atkinson2000](#); [Millet2006](#)). Both these loss processes (photolysis, oxidation) form CO and hydroperoxyl radicals (HO₂), and have global significance to radiative forcing and oxidative capacity ([Franco2015](#)). The other major sinks are wet and dry deposition, although these are not as significant ($\sim 32 \text{ Tg yr}^{-1}$) ([Atkinson2000](#); [FortemsCheiney2012](#)).

1.4.2 Measurement techniques

There are a few ways to measure HCHO, including Fourier Transform Infra-Red (FTIR) Spectrometry and Differential Optical Absorption Spectroscopy (DOAS). FTIR examines the Fourier transform of a measured spectrum in order to detect things which affect that spectrum. DOAS methods are based on light interference and absorption through air masses.

The DOAS technique takes advantage of the optically thin nature of HCHO in order to linearise the radiance differential through air masses with and without HCHO, using the Beer-Lambert intensity law. This method is used both on the ground, and from space, globally for HCHO detection ([Guenther1995](#); [Abad2015](#); [Davenport2015](#)). As a trace gas HCHO interferes with light over a few wavelength bands, which allows instruments to detect concentrations between a known light source and a detector. Figure 1.8 shows the interference spectrum of HCHO along with a typical band used to examine interference in the DOAS technique. One difficulty is that this interference is relatively small (HCHO is optically thin) and other compounds absorb light at similar wavelengths ([Davenport2015](#)).

FTIR and DOAS measurements have a range of uncertainties, including systematic and random measurement errors and uncertain a priori shape factors and water profiles (eg: **Franco2015**). Other types of measurement involve directly measuring the air, and determining chemical compounds through their physical properties such as by mass spectrometry analysis of mass to charge ratios (m/z) of ionised air masses. Two examples of this include proton transfer reaction mass spectrometers (PTR-MS), and gas chromatography mass spectrometers (GC-MS). These instruments can be used to determine gas phase evolution of isoprene and monoterpene products such as HCHO (**Lee2006a; Nguyen2014; Wolfe2016; Lerner2017**).

Other measurement techniques include chromatographic and fluorimetric methods, both of which differ widely from each other and the spectroscopic methods (**Hak2005**). **Hak2005** examine a single air mass with 8 instruments using the four techniques (MAX-DOAS, FTIR, chromatographic, and fluorimetric), and show that reasonable agreements can be achieved. Generally the measurements were close, the five Hantzsch instruments agreeing to within 11% (after removing two potentially faulty measurements), although different calibration standards were used. Titration for the different calibration solutions could not be resolved, which may account for absolute offsets up to 30%. These differences and non-uniformities between measurements (even among identical instruments) are part of the reason HCHO does not have a consistent network for global measurements like those for greenhouse gases or ozone (**FortemsCheiney2012**).

1.4.2.1 Satellite measurements

Satellites remotely sense atmospheric HCHO through irradiance measurements of solar light which has reflected off the earth's surface. These irradiances are affected by gases which exist along the reflected path of light between the detector, earth, and sun. The irradiance is then used to estimate how much of a particular gas exists along this path, which gives us an estimate which is called a slant column (SC). The retrieved SC of a particular gas (or species) can be transformed into a vertical column (VC) by scaling the path length in conjunction with accounting for the trace gas' light scattering properties. The scaling coefficient created to transform from SC to VC is called the Air Mass Factor (AMF).

Several satellites provide long term trace gas observations with near complete global coverage, including the ERS-2 launched in April 1995 which houses the GOME ultraviolet and visible (UV-Vis) spectrometer, the AURA launched in July 2004 which houses the OMI UV-Vis spectrometer, the MetOp-A and B launched in October 2006 and September 2012 respectively both housing a GOME-2 UV-Vis spectrometer. These satellites are on Low Earth Orbit (LEO) trajectories and overpass any area up to once per day. Satellites use DOAS techniques with radiative transfer calculations on solar radiation absorption spectra to measure column HCHO . An example of a spectrum retrieved from the GOME-2 instrument is given in figure 1.9.

Satellite based chemical concentrations often require both remote and in situ measurements combined with modelled data for validation (**Marais2014**). There is less information available from satellite measurements at higher latitudes due to increased error in measurements over the more slanted column paths (**DeSmedt2015**). Validation is important due to the various uncertainties in the satellite remote sensing

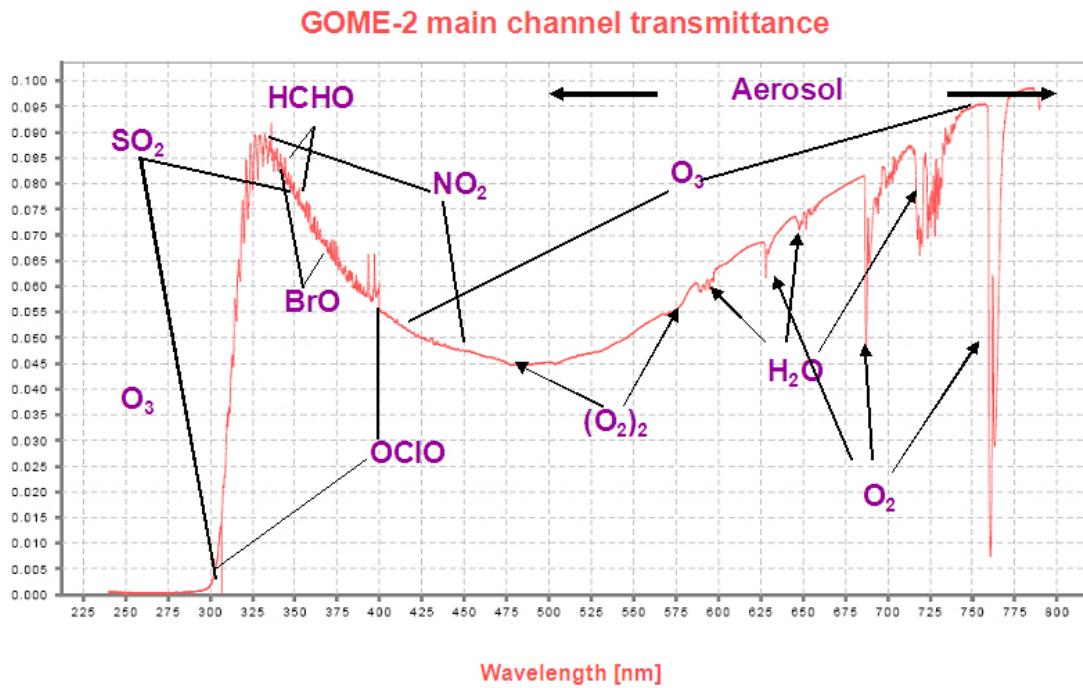


FIGURE 1.9: An example spectrum showing interferences used for species concentration measurements by GOME-2. Image by EUMET-SAT and ESA ([GOME2Image](#)).

process. One example is [Zhu2016](#), who used SEAC⁴RS aircraft HCHO measurements over the southeastern US as model validation. They showed that a bias in the assumed OMI shape factor leads to a bias between satellite and SEAC⁴RS measurements. Different instruments will give different results (called structural bias or uncertainty) due mostly to a priori assumptions and different measurement techniques ([Lorente2017](#)).

In conjunction with atmospheric chemistry and radiative models, satellite measurements quantify the abundance of HCHO in the atmosphere. Isoprene is hard to measure directly due to its short lifetime and weak spectral absorption, instead HCHO is often used as a proxy ([Millet2006](#); [Fu2007](#); [Dufour2009](#); [Marais2012](#); [bauwens2013satellite](#); [Kefauver2014](#); [Bauwens2016](#); [Surl2018](#)). This leads to a method of isoprene emissions estimation termed top-down (as opposed to bottom-up estimates). The existence of satellite data covering remote areas provides an opportunity to improve VOC emissions estimates leading to more robust models of global climate and chemistry. Satellite data gives us another way to estimate large scale isoprene emissions, and their subsequent chemistry. This method is described in detail in section ??.

1.5 Atmospheric Chemistry Modelling

Models can fill the gaps (both spatial and temporal) in measurement records, and can help us improve our understanding of the natural world. They are used to examine future outcomes resulting from changing our emissions, from small to large scales. They can be used to increase measurement accuracy (for instance in satellite

measurements) and determine where we lack information, while also checking the performance of new instruments. Precisely representing various chemicals and reactions in the atmosphere allows efficient mitigation of pollution, since we can compare scenarios against one another. Models can always be expanded to include new compounds or processes, however validation is always necessary. Currently they require improved isoprene emissions and subsequent chemistry understanding for effective air quality determination (**Marvin2017**).

1.5.1 Box models

Box models are much smaller scale than global CTMs, examining one uniform environment with many parametrisations such as transport and emissions. Box models can be used to check chemical mechanisms in specific scenarios, such as high or low NO_x environments. For example: **Marvin2017** use a box model matching conditions in southeast USA to evaluate isoprene mechanisms from several models. A box model involves modelling chemistry in a singular set of conditions without transport or any spatial gradients.

By allowing for interactions between boxes this concept can be extended to multiple-box models. These are simply multiple instances of single boxes with the addition of transport between them, which requires meteorological fields such as wind velocities and turbulence. The meteorology fields can be modelled, and/or input as parameters.

1.5.2 Chemical transport models

Chemical transport models (CTMs) provide a simulation of chemical densities and transport over time, through the atmosphere. They require many inputs (such as wind velocities) in order to accurately represent scenarios or regions on earth. Models of emissions are often used as drivers for atmospheric chemistry models, which require initial and boundary conditions in order to run. Chemistry in the atmosphere is a complex system of coupled reactions and dynamics, which can be solved using numerical partial differential equation solvers.

CTMs simulate production, loss, and transport of chemical species. This is generally calculated using one or both of the Eulerian (box) or Lagrangian (puff) frames of reference. Eulerian models use examine equations and transport within and between volumes in a spatial coordinate systems, while Lagrangian models look at behaviour within a potentially changing frame of reference (for example within a cloud). CTMs normally solve continuity equations simultaneously for many coupled species. The continuity equations describe transport of a conserved quantity such as mass or energy, which, solved together with production and loss of a chemical can provide detailed simulations of natural processes.

The general continuity equation links a quantity of a substance (q) to the field in which it flows and can be described by the formula:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot j = \sigma$$

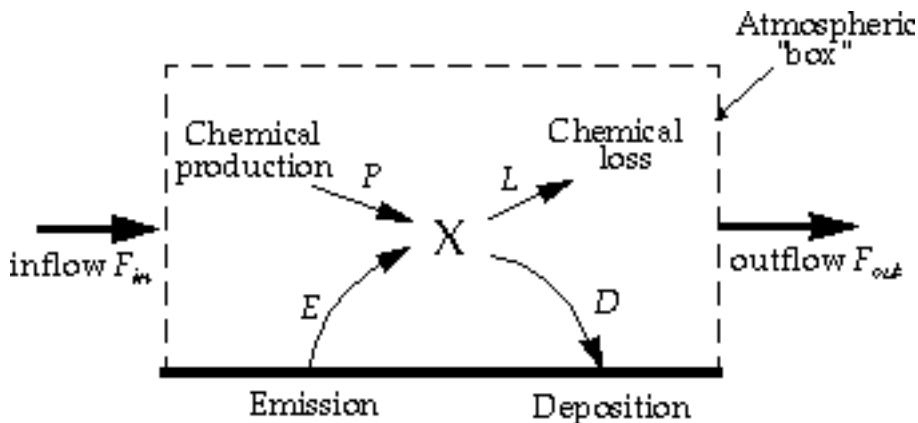


FIGURE 1.10: Standard box model parameters, image taken from [Jacob_1999_book](#).

where ρ is density of q in the field, t is time, ∇ is divergence, j is the flux (q per unit area per unit time entering or leaving the field), and σ is the generation or loss of q per unit volume per unit time.

The type of model best suited to modelling the entire earth uses the Eulerian frame of reference, where the atmosphere is broken up into 3-D boxes with densities and transport calculated and stored for sequential steps in time at each location. The mass balance equation must be satisfied in any realistic long term model and is as follows:

$$\begin{aligned}\frac{dm}{dt} &= \sum \text{sources} - \sum \text{sinks} \\ &= F_{in} + E + P - F_{out} - L - D\end{aligned}$$

where m is mass of a chemical, E and D are emission and deposition, P and L are production and loss, and F is chemical transport in and out, as shown in figure 1.10. Many chemical species interact with each other through production and loss. Any large chemical model will solve this mass balance equation over highly coupled arrays of partial differential equations, which becomes computation time expensive as complexity increases.

Contemporary models generally use mathematical differential solving tools of various complexity (often called chemical mechanisms) to solve chemical equations in order to predict chemical species evolutions over time. Different solvers may be slower or faster and more suited to particular situations based on the stability of the equations and systems involved, and chemical mechanisms may vary in how many reactions and chemicals are listed and grouped together. For example: Since $[O] \ll [O_3]$ the chemical family O_X ($O_X \equiv O + O_3$) can be used to simplify chemistry simulations and approximate O_3 concentrations ([BrasseurJacob2017](#)). Different chemical mechanisms may find different solutions to the same problems, due to how the numerical solvers are implemented, which can affect model output ([Zhang2012](#)).

1.5.3 Emissions

There are two commonly used ways of estimating isoprene emissions, top-down or bottom-up. Bottom-up emission estimates generally model the flora which emit isoprene, along with the rates of emissions and things which affect these rates. The general formula governing modelled emissions E for a species i (from **BrasseurJacob2017**) is as follows:

$$E_i = A \times EF_i \times S_i$$

with A the activity rate (eg. how many trees in an area), EF_i being the emission factors (eg. isoprene emitted per tree per year), and S_i is a scaling factor accounting for meteorology and other effects not included in A or F (eg. seasonal temperature).

Isoprene is emitted by trees or shrubs, depending on several parameters such as leaf area index (LAI), plant functional type (PFT), and light density fraction (LDF). Models use these properties of the emitters in order to estimate how much isoprene is being produced (**Guenther1995; Guenther2006**). Understanding how much isoprene is emitted, when and by what, is complicated. One frequently used bottom up emissions model is the Model of Emissions of Gases and Aerosols from Nature (MEGAN, **Guenther1995**). MEGAN has estimated ~ 1150 Tg C yr $^{-1}$ BVOC emissions globally, of which $\sim 465\text{--}500$ Tg C yr $^{-1}$ is isoprene (**Guenther2006; Messina2016**). Since little data exists with which to verify these bottom-up emission inventories, they can be uncertain on a large scale.

Bottom up models of VOC emissions are sensitive to parameters. For example **Stavrakou2014** examined modelled Asian emissions and altered model parameters for temperature, plant type emission factors, incoming solar radiation (insolation) intensity, land use changes, and palm tree forest expansion. Changes were constrained by a network of radiation measurements and some experiments with south east Asian forest emissions - and led to reduction in a priori isoprene emissions by a factor of two over the region in 2005. Sensitivity to these factors is pervasive in bottom up emissions models (**Marais2014; Miller2014; Messina2016**). One of the important uncertainties seen in MEGAN is the isoprene emissions due to PFTs. If one plant species is emitting heavily near a measuring instrument, possible overestimations may occur due to extrapolation over the entire forest. Global emissions inventories like MEGAN often have large areas based on extrapolations which introduces uncertainties (**Miller2014**). Current emissions estimates require more validation against observations, and recently a comparison of two major VOC models (MEGAN and ORCHIDEE) was undertaken by **Messina2016** reiterating this requirement. In their work they examine model sensitivities and show that the most important parameters are LAI, EF, PFT, and LDF. There is high uncertainty in LAI and EF, which require more or improved measurements at the global scale, as well as more PFTs and improved LDF parameterisation (**Messina2016**).

1.5.4 Uncertainties

Here I will attempt to list and partially explain the major uncertainties models have in relation to VOCs, and ozone. Atmospheric chemical models by necessity require various simplifications of real world processes, and also utilise information which may be itself uncertain or extrapolated. Uncertainty is introduced through both of these

channels as well as through computational limitations and non-linear non-continuous system solution approximations.

1.5.4.1 Emissions Inventories

Using different emissions inventories in a CTM can have large impacts on the simulation. Natural (biogenic or pyrogenic) and human driven (anthropogenic) emissions often drive a large fraction of atmospheric oxidation and radical chemistry, especially in the continental boundary layer. Emissions inventories have been found to be generally OK at larger (regional to global) scales, as long as they are derived from accurate input measurements (**Zeng2015**). Modelled ozone concentrations has been found to be most sensitive to isoprene emissions and NO_x sources, both of which have uncertainty factors of ~ 2 (**Christian2017**).

Many estimates of isoprene emission are based on a few algorithms which can depend greatly on input parameters (**Arneth2008; Niinemets2010**). **Arneth2008** argue that this monopoly of emissions estimates may be leading us to an incorrect understanding of isoprene chemistry. **Yue2015** have shown that this is still a problem by looking at land carbon fluxes and modelling the sensitivity to VOC emissions estimates using two independent models of VOC emission. One model is photosynthesis based and estimates isoprene emissions using electron transfer energies and leaf physiology (**Niinemets1999**), while the other (MEGAN) uses the light and canopy temperature (**Guenther1995; Arneth2007**). Both are sensitive to light and temperature parameterisations.

The concentration of NO_x is an important factor in determining the yield of HCHO and ozone from BVOCs. **Travis2016** show how modelled surface ozone is overestimated due to high estimates of NO_x emissions, which affect oxidative capacity and VOC reactions in the US. NO_x and isoprene emissions are shown to be the most significant sources of uncertainty for ozone concentrations near the surface in GOES-Chem over the US, while isoprene derived products and lightning NO_x drives uncertainty in the upper atmosphere (**Christian2017**).

1.5.4.2 Resolution

Atmospheric chemistry simulations are somewhat sensitive to the gridbox resolution. For example: **Wild2006** show that reduced resolution increases OH concentrations and ozone production rates. **Christian2017** find small changes in OH (< 10%) in OH, HO₂ and ozone concentrations local to the north American arctic, when changing from 4 by 5 to 2 by 2.5°resolution. **Yu2016** show how only at higher resolution (0.25 by 0.3125°) does isoprene oxidise under the correct NO_x scheme (through high or low NO_x pathways, see section 1.3.3.1) in variable NO_x environments. This leads to an increase of high NO_x pathway oxidation of isoprene at the lower resolutions, which leads to an overestimation of HCHO but not ozone at coarser resolutions. However, for many global scale analyses, errors from resolution are less important than those from chemistry, meteorology, and emissions (**Christian2017; Christian2018**).

1.5.4.3 Chemistry mechanisms

There is still much work to be done in models to correctly simulate emissions and processes which lead to HCHO and ozone. Often HCHO is used as a way of checking if precursors are correctly modelled since HCHO measurements are more readily available (for instance from satellites). Recently **Christian2017** analysed GEOS-Chem (A global CTM; see section 2.4) for ozone and oxidant (OH and HO_2) sensitivity to the driving processes and inputs. They found that GEOS-Chem ozone was most sensitive to NO_2 photolysis, the $\text{NO}_2 + \text{OH}$ reaction rate, and precursor emissions such as VOCs.

Many models lack in-situ measurements with which to verify their chemical mechanisms, leading to large discrepancies (**Marvin2017**). **Marvin2017** suggest that isoprene mechanisms in several contemporary models (including GEOS-Chem) are inadequate. They show that (for a specific measurement campaign) HCHO concentrations are underestimated in a way that can not be easily fixed through reaction rate changes. They compared five global CTMs isoprene mechanisms by evaluating simulated HCHO mixing ratios compared to in situ measurements from the Southeast Nexus (SENEX) aircraft campaign (in southeastern USA). Five models (GEOS-Chem, CB05, CB6r2, MCMv3.2, and MCMv3.3.1) all are found to underestimate HCHO concentrations (by 15 – 30%).

1.5.4.4 Clouds

One of the major uncertainties in chemical, climate, radiation, and weather models is cloud formation and dynamics. Clouds are remarkably complex at a much finer scale than can be accurately modelled by global chemistry models (with current processing power). Globally over half (50-60%) of the world is covered by clouds, with $\sim 10\%$ of them being rain-clouds (**Kanakidou2005**). Wet scavenging performed in clouds not only depends on large scale cloud processes, but also on the microphysics of aerosols being scavenged, differing between aerosol sizes and hygroscopic properties.

1.5.4.5 Soil Moisture

Modelled emissions are sensitive to soil moisture, especially near the wilting point, below which trees stop emitting isoprene and other VOCs completely as they can no longer draw water (**Bauwens2016**). MEGAN accounts for soil moisture through a parameterisation which drops plant emissions to zero below a prescribed soil moisture level (the wilting point). Recently an update to this has been shown to improve modelled isoprene emissions in drought conditions (**Jiang2018**). **Jiang2018** found that improving the parameterisation of drought based on a measurement campaign in the U.S. would lower isoprene emissions globally by $\sim 17\%$. Many environmental parameters are affected by soil moisture, which all play a role at fine scales to surface emissions (**Rowntree1983; Chen2001**). Droughts effects can be difficult to measure, as they are a multi-scale problem which affects various aspects of the land-air interface including plant emissions and dry deposition (**Wang2017**).

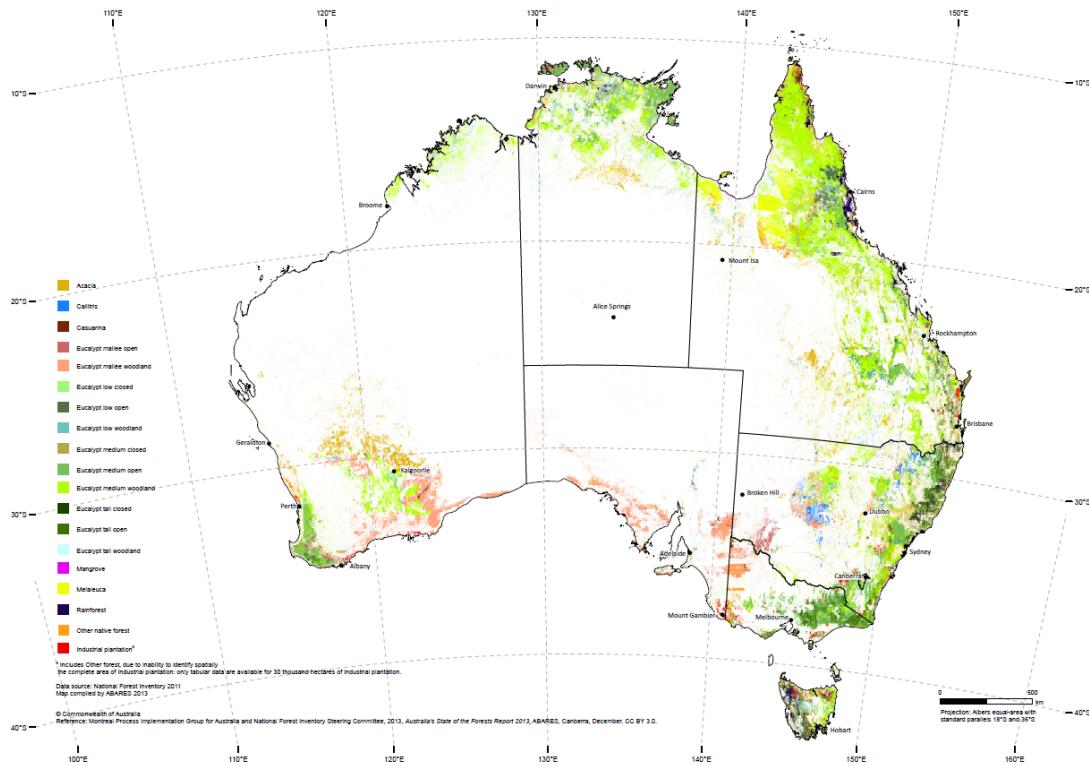


FIGURE 1.11: Forest types in Australia (<http://www.agriculture.gov.au/abares/forestsaustralia/australias-forests>)

1.6 Australia and the southern hemisphere

Australia has a unique climate, along with soil moisture, clay content and other important properties which affect VOC emissions. These properties are only sparsely measured in Australia due to the spread out distribution of population centres, which make many areas very difficult or expensive to reach. In Australia most long term air quality or composition measurements are performed in or near large cities. Australia is dominated by areas with little anthropogenic influence and few ground based measurements of the natural emissions taking place (VanDerA2008). Since many Australian cities are on the edge of regions with rich VOC emissions, it is very important to clarify the quantity, type, and cause of VOC emissions. Understanding of emissions from these areas is necessary to inform national policy on air pollution levels.

The vegetation in Australia is diverse, a summary is provided by ABARES using the national forest inventory at <http://www.agriculture.gov.au/abares/forestsaustralia/australias-forests>. Figure 1.11 shows the different forest types and their locations within Australia, highlighting that much of our forested lands are near population centres along the east coast. 16% of Australia is covered by forest, most (75%) of which is Eucalyptus.

Ozone enhancements above the background levels are most sensitive to emissions (of precursor gases), with meteorology, and atmospheric composition also important. Anthropogenic emissions of ozone precursors are important but relatively stable,

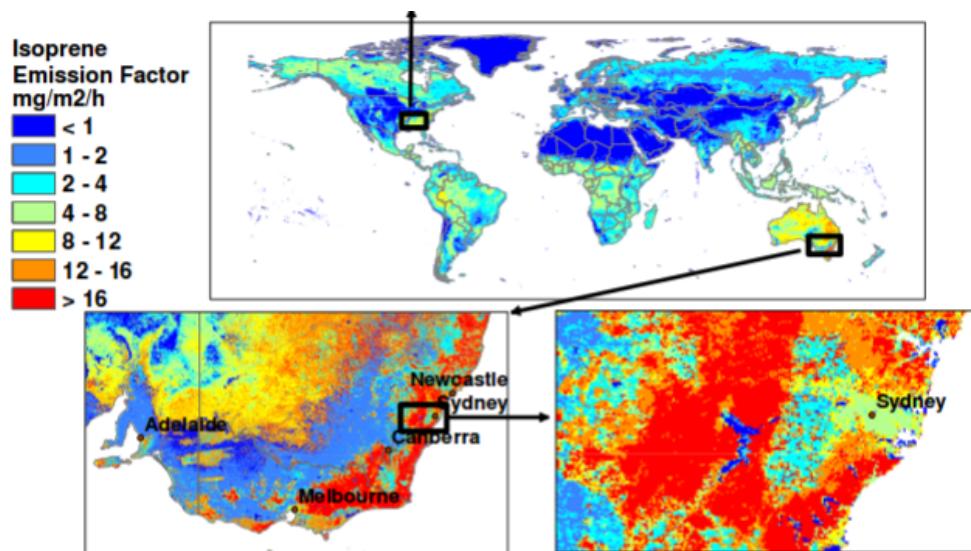


Fig. 2. Global distribution of landscape-average isoprene emission factors ($\text{mg isoprene } \text{m}^{-2} \text{ h}^{-1}$). Spatial variability at the base resolution ($\sim 1 \text{ km}$) is shown by regional images of the southeastern U.S. and southeastern Australia.

FIGURE 1.12: Part of a figure from Guenther2006 showing global isoprene emission factors.

while pyrogenic sources are greatly variable and dependent on weather, fuel, and fire intensity (Lawson2017). Emissions from burning include a range of chemical compounds and particulates and each year the effects of fire or burning seasons blanket the northern and southern hemispheres independently. Biomass burning in southern Africa and South America has previously been shown to have a major influence on atmospheric composition in Australia (Oltmans2001; Gloudemans2006; Edwards2006), particularly from July to December (Pak2003; Liu2016). Local fires are even more influential and the burning season for Australia can be all year, with severity depending on regional vegetation, recent and current weather, and El-niño.

It has been estimated by MEGAN that the Australian outback is among the world's strongest isoprene emitters with forests in SE Australia having emission factors greater than $16 \text{ mg m}^{-2} \text{ h}^{-1}$ (see figure 1.12) (Guenther2006; Guenther2012). Measurement campaigns in SE Australia have since cast doubt on the emission factors used by MEGAN, potentially due to poor characterisation of Eucalyptus trees and soil moisture (Emmerson2016). These emissions factor estimates are not well verified and measurements of isoprene (or other BVOC) emissions are sparse and infrequent in Australia (Sindelarova2014; Bauwens2016). In addition, monoterpene emissions are 2-4 times too low, which may be due to underestimated emission rates for many Eucalypt species (Winters2009; Emmerson2016).

1.6.1 Ozone

Surface ozone levels over Australia are relatively low ($\sim 20 \text{ ppb}$) (Young2018), however it remains unclear how much we would expect this to change in the future as relatively little is known about precursors and influx for the continent. Australian air quality is monitored independently within each state, using several metrics. These metrics

are measured by varying numbers of monitoring stations in each state. Measurement stations are generally located in population centres, and do not regularly measure isoprenoid emissions. This is an important omission as these naturally emitted precursor gases often get transported into cities where they affect air quality through production of O₃ and other pollutants.

Generally STT of ozone over Australia affects the upper troposphere only, however ozone enhancements can reach quite low during heavy storms and cyclonic weather patterns (**Alexander2013**). The contribution of STT to overall tropospheric ozone budgets remains uncertain, especially in the southern hemisphere (SH) (**Skerlak2014**). STT can enhance surface ozone concentrations above legal air quality limits (**Lelieveld2009; Lin2015**). It is easier to determine tropospheric ozone enhancements over the relatively clean southern ocean atmosphere. However measurements of tropospheric ozone over this region are relatively sparse (**Skerlak2014**), and quantification of transported ozone is difficult without large scale extrapolations. Ozone enhancements over the southern ocean signify either transported pollution or stratospheric influx (**Jacobson2000**). Quantifying ozone processes over the southern ocean would be helpful towards understanding chemistry in the “clean background environment”, which is important when validating models and satellite datasets.

1.6.2 VOCs

Bottom up inventories of VOCs remain largely uncertain due to extensive extrapolation over plant functional types, changing land cover, and parameterised environmental stressors (**Guenther2000; Kanakidou2005; Millet2006**). **Muller2008** show how isoprene (a key VOC) is poorly captured by the MEGAN model and analyse the affect of changing the soil moisture parameter. **Sindelarova2014** show reductions in modelled Australian isoprene emissions of 50% when incorporating soil moisture in MEGAN estimates. Uncertainties in isoprene emissions could explain why models of HCHO over Australia are poor at reproducing satellite measurements (**Stavrakou2009**). Improved parameterisation of the affect of drought on plant emissions could also lower modelled isoprene emissions (**Jiang2018**).

Australia suffers from poor characterisation of plant emissions, partly because emission factors are based on northern hemispheric data. Many plant emissions rates have not been published, such as those for any Australian acacias. Some Eucalypt emissions are based on samples from young trees, which may emit more isoprene than older trees (**Emmerson2016**). Additionally soil moisture is not well quantified which has a large effect on emissions. Soil type and moisture, along with drought thresholds, have poorly understood effects on plant emissions in Australia Changes in parameterisation of soil moisture in the MEGAN lead to massive changes in Australian isoprene emission estimates (**Sindelarova2014**). Over Australia MEGAN suffers from a lack of studied plant functional types and their emissions (**Muller2008**). Emission rates from various species of Eucalyptus and other flora are highly complex, depending on current and recent weather, temperature, tree age, health, etc. (**Guenther2012**). With this complexity added to the diversity of tree species in Australia as well as sparse rural data collections it is hard to model and verify emissions.

Emmerson2016 analysed isoprene and monoterpene emissions sensitivities in a regional model of atmospheric chemistry over southeast Australia, using four campaigns which are also examined in this thesis. They show that modelled emissions require spatially and temporally resolved changes. **Emmerson2016** suggest that monoterpenes may be emitted in similar quantities to isoprene, with more measurements required to determine if this is so. They compare emissions estimates from MEGAN against field campaign data and see overestimated isoprene emissions, as well as underestimated monoterpene emissions. Their work suggests that MEGAN estimates of isoprene emissions may be 2-6 times too high, and monoterpene emissions ~ 3 times too low over southeast Australia.

Improvements to emissions models require improved understanding of regions and their behaviour. Satellite measurements can be used to improve understanding of Australian emissions. As HCHO is produced with relatively high yield after isoprene is emitted, we can use satellite measurements to estimate isoprene emissions (**Palmer2001; Millet2006; Bauwens2016**).

1.6.3 Measurements

Isoprene and many of its products can be difficult to measure accurately due to their short lifetimes, high reactivity, and optically thin natures. There are relatively few measurements in the southern hemisphere, including MUMBA (**PatonWalsh2013**), SPS(**Dunne2018**), and Tumbarumba (**Emmerson2016**). These campaigns focus on air quality or biogenic emissions and use several different instruments (including PTR-MS and GC-FID) to detect metrics such as air particulates, HCHO, isoprene, and meteorological information. An air-flight campaign (HIPPO) measuring over the Pacific ocean, with one flight passing along the NSW coastline, also provides isoprene and ozone concentrations in November 2009 (**Wolfsy2011**). Wollongong also has 20 years of DOAS calculated HCHO measurements from a solar FTS. Satellite total columns suffer from orography amongst other limitations when compared to this data (**Demol2010**). For further details on these campaigns see Section ??.

Detecting ozone from the surface up to the top of the stratosphere requires different techniques such as remote sensing and ozonesonde releases. Ozonesondes are weather balloons (with attached ozone detectors) which detect ozone concentrations up to the mid stratosphere (~ 30 km), providing a vertical profile over a single location. Since 1986, Lauder, New Zealand ($45^{\circ}\text{S}, 170^{\circ}\text{E}$) has released ozonesondes allowing a multi-decadal analysis of ozone concentrations over the city (**Brinksma2002**). Kerguelan Island ($49.2^{\circ}\text{S}, 70.1^{\circ}\text{E}$), also has a record of ozonesonde profiles, which are directly in the path of biomass burning smoke plumes transported off shore from Africa (**Baray2012**). SHADOZ is the southern hemispheric additional ozone project, which have released sondes from 15 sites at different times <http://tropo.gsfc.nasa.gov/shadoz/>. A smaller network of ozonesonde release sites (including Davis, Macquarie Island, and Melbourne) is available from the world ozone and ultraviolet radiation data centre <http://woudc.org/data/explore.php> and is used in Chapter 4 to examine stratospheric impacts on tropospheric ozone (see ?? for more info on these ozonesondes).

1.7 Aims

In this thesis I aim to improve understanding of natural contributions to ozone over Australia and the southern ocean. The two largest contributors to tropospheric ozone concentrations are chemical production (driven by precursor emissions) and stratospheric transport. I aim to improve understanding of both of these sources using existing satellite and ground-based datasets along with GEOS-Chem modelled outputs.

Estimation of BVOC emissions in Australia can be improved through satellite measurements of one of the primary oxidation products HCHO. Satellites which overpass daily record reflected solar (and emitted terrestrial) radiation, and give us measurements over all of Australia. Combining satellite data with model outcomes provides a platform for the understanding of natural processes, which are uncertain over Australia. Satellite measurements use modelled a priori vertical profiles of HCHO to estimate total column amounts. **I aim to recalculate satellite vertical columns of HCHO using updated model a priori information.** In this effort I aim to improving the understanding of the importance of relevant parameters (within GEOS-Chem) in calculating vertical columns of HCHO measured by satellite. This includes an examination of how well GEOS-Chem simulates several species such as NO_X, isoprene, and HCHO compared to both in-situ and remote measurement data that exists for Australia. Additionally I detail the construction and effects of satellite data filters. The work towards this aim is in Chapter 2.

The technique of determining isoprene emissions from satellite detected HCHO is called satellite inversion. **I aim to determine isoprene emissions in Australia using a top-down inversion of satellite HCHO, through an estimated yield from isoprene to HCHO** HCHO amounts and the yield of isoprene to HCHO over Australia is required to create top-down estimates. This process also requires careful examination of when the assumptions required within the inversion process are not valid. Due to the low availability of in-situ data over most of the Australian continent, a combination of modelled and satellite data could reduce the uncertainties of isoprene emissions from Australian landscapes. Improved emissions estimates will in turn improve the accuracy of CTMs, providing better predictions of atmospheric composition and its response to ongoing environmental change. The work towards fulfilling this aim is in Chapter 3.

To improve understanding of ozone transported to the troposphere from the stratosphere in Australia and the southern ocean. Stratospheric transport is the second largest driver of tropospheric ozone concentrations, and an improved understanding of transported ozone can be determined from ozonesonde measurements. Ozonesondes provide a glimpse of the vertical ozone profile up to ~ 30 km, and we use a Fourier filter to determine how often stratospheric transport is occurring at three sites: Melbourne, Macquarie Island, and Davis Station. Combining transport event frequency analysis with modelled ozone distributions is used to derive a new method of detection and quantification of transported ozone in Chapter 4.

I aim to describe relative importance of sources of tropospheric ozone in Australia, as well as seasonality. I will describe how modelled ozone is affected by updated isoprene emissions, comparing changes in GEOS-Chem outputs. Trends of isoprene emissions and their relationship to tropospheric ozone trends could provide

new insight into the future of tropospheric ozone in Australia.

Chapter 2

Data and Modelling

2.1 Introduction

In this thesis the word model is most often used to represent a chemical transport model (CTM), which simulates chemistry and chemical transport through the atmosphere. Models of the atmosphere can be used to interpret measurements, estimate chemical concentrations at any scale, and predict atmospheric composition in the future. In remote sensing measurements, modelling is required in order to produce useful outputs. Models of ozone in the atmosphere are used broadly for international assessments of ozone precursor emissions, and estimating effects from related processes (such as radiation) (Young2018). Models provide an estimate of many trace gas concentrations; however, verification is required, and generally performed using results from measurement campaigns. In situ measurements from campaigns or measurement stations can be used to examine what is happening at a particular location. These data are used to determine how accurate models or estimates are - however the utility is limited to where and when the measurements took place. In this thesis data from several campaigns are compared against model outputs and satellite datasets. Satellite datasets provide large amounts of data over most of the planet. However, they can have high uncertainty due to instrument limitations. Many datapoints can be averaged in order to reduce uncertainty. In this chapter several satellite datasets are combined to estimate biogenic HCHO amounts over Australia.

The first goal is to analyse Australia-specific HCHO concentrations measured by satellite, and determine isoprene sensitivity and any model bias. This leads into Chapter 3 where biogenic HCHO columns are used to estimate isoprene emissions. The second goal is to quantify ozone transported from the stratosphere down into the troposphere (Chapter 4). The focus in this chapter is to describe and analyse model outputs and measurements along with how they are recalculated and compared. Section 2.2 details satellite and campaign datasets, and additionally describes model outputs. Measurement techniques used to retrieve the most utilised satellite dataset are outlined in Section 2.3. Section 2.4 describes the GEOS-Chem model, how it is run and what setup and outputs are used in this thesis. In Section 2.5 the process of using model outputs to recalculate satellite vertical columns is defined and analysed. In order to compare satellite data with other datasets, some work must be undertaken to avoid introducing bias (Palmer2001; Eskes2003; Marais2012; Lamsal2014). One key step is to recalculate the satellite information using modelled data, detailed in Section 2.6. The effects of these recalculations on satellite HCHO is also examined. The

creation and effects of filters used to remove non-biogenic influences are described in Section 2.7.

2.2 Datasets

This section describes the datasets used in this thesis, along with an overview of the measurement techniques used for each. This includes modelled output, satellite measurements, and measurement campaigns. These datasets serve four purposes:

1. Model output validation in this chapter
2. Calculation of biogenic HCHO distribution over Australia in this chapter
3. Recalculated OMI formaldehyde columns are used as a basis for estimating isoprene emissions in Chapter 3
4. Extrapolation of ozone transport in Chapter 4

I will also give details on filtering and interpolations which are undertaken when reading data, as each dataset has its own resolution. While I have not made any measurements myself, it is important to understand the techniques used in datasets I have utilised in order to understand possible anomalous datapoints or trends.

Horizontal geographical coordinates are always discussed in terms of degrees north and east, from -180° to 180° spanning west to east around the globe and -90° to 90° spanning the latitudes from the Antarctic to the Arctic. Vertical resolution is sometimes discussed in terms of metres, sometimes in terms of pressure (hPa), and sometimes in terms of sigma (σ) coordinates. Sigma coordinates represent the fraction of the atmosphere vertically above, with $\sigma = 1$ being the surface and $\sigma = 0$ being the top of the atmosphere (TOA). This can be useful when running global atmospheric models as the ground altitude is always at $\sigma = 1$ and we need not worry about topography. Conversion between σ and pressure (p) coordinates is as follows:

$$\sigma = \frac{p_S - p}{p_S - p_T} \quad (2.1)$$

where p_S and p_T is surface pressure and pressure at the TOA respectively.

Uncertainty (or error) is present in each dataset and where possible the causes are explained. There are two types of error: systematic and random. Arguably the worst of these is systematic error (or bias). Bias normally indicates a problem in calculation or instrumentation. If the systematic error is known, it can be corrected for by either offsetting data in the opposite direction, or else fixing the cause. A proper fix can only be performed if the sources of error are known and there is a way of correcting or bypassing it. Random error is often reported as some function of a datasets variance, or uncertainty. It can be reduced through averaging either spatially or temporally. Temporal and/or spatial averaging decreases uncertainty by a factor of $1/\sqrt{N}$ where N is the number of observations being averaged.

2.2.1 Satellite

Satellite data products are generally classed into several categories, level 0 through to level 3. Level 0 products are sensor counts and orbital swath data, level 1B data calibrates and geo-locates the level 0 data. Level 2 products additionally have temporal, spatial, solar, and viewing geometry information, as well as quality flags. To create level 2 data slant column density is determined and then translated into vertical column density. Level 3 data is a temporally aggregated subset of level 2 data, for instance monthly or yearly averages.

Satellites record near nadir (vertical) reflected spectra between around 250-700 nm split into spectral components at around 0.3 nm in order to calculate trace gases including O₃, NO₂, and HCHO ([Leue2001](#)). Satellite measurements are generally performed using spectral fitting followed by conversion to vertical column densities. Several public data servers are available which include products from satellites, including NASAs Earthdata portal (<https://earthdata.nasa.gov/>) and the Belgian Institute for Space Aeronomy (IASB-BIRA) Aeronomie site (<http://h2co.aeronomie.be/>).

Rayleigh and Mie scattering describe two kinds of particle effects on radiation passing through a medium. Rayleigh scattering is heavily wavelength dependent, and is the dominant form of scattering from particles up to roughly one tenth of the wavelength of the scattered light. Mie scattering more generally involves larger particles, and has less wavelength dependence. The effects of scattering are what gives us the information about substances in the atmosphere. The different particles and gases in the air have measurable properties seen by remote sensing devices such as a satellite. Although instruments will be more or less sensitive to various properties depending on altitude, radiation, and other parameters ([Martin2002](#)).

Difficulties can arise when aerosols interfere with recorded spectra (e.g., clouds, smoke, dust); however, some of these can be detected and filtered out. Instruments including MODIS onboard the Aqua and Terra satellites are able to determine aerosol optical depth (AOD), a measure of atmospheric scatter and absorbance. An AOD under 0.05 indicates a clear sky, while values of 1 or greater indicate increasingly hazy conditions. This is important in order to determine where measurements from other instruments may be compromised by high interference. Cloud filtering is performed on several satellite products used in this thesis, due to the uncertainty introduced by cloud interference. This has been seen to introduce a clear-sky bias in monthly averages since measurements do not include cloudy days ([Surl2018](#)).

Satellite measured AOD requires validation by more accurate ground based instruments like those of AERONET which uses more than 200 sun photometers scattered globally. Soon much more satellite data will be available in the form of geostationary satellite measurements ([Kwon2017](#)). Geostationary satellites can provide temporally rich measurements over an area, as they are not sweeping around the earth but fixed relative to one latitude and longitude.

2.2.1.1 Formaldehyde

OMI spectra are used in several products used in this thesis, including OMHCHO, OMNO2d, and OMAERUVd. Satellite based formaldehyde measurements from the OMI instrument onboard Aura are stored in the OMHCHO product. OMHCHO data



FIGURE 2.1: Example of NO₂ tropospheric columns taken from the OMNO2d product.

is used and modified extensively throughout this thesis, and so is discussed in more detail in Section 2.3. Calculation of column density and AMF are discussed respectively in sections 2.3.2 and 2.3.3.

HCHO products can be found in four satellite instruments: GOME on ERS-2, SCIAMACHY on ENVI-SAT, OMI on EOS Aura, and GOME2 on MetOp-A. These satellites have slightly different spectral and spatial resolutions, as well as using varied processes to estimate HCHO from detected radiances. This leads to different estimates between instruments as described in Lorente2017, and both validation and comparison become more important when using these remotely sensed data. The data set used in this thesis is from the Ozone Monitoring Instrument (OMI) onboard the Aura satellite, as it has data for the entire time line and sufficiently covers the southern hemisphere.

2.2.1.2 Nitrogen dioxide

OMNO2d is a gridded daily level three NO₂ product with good satellite pixels averaged into 0.25°x0.25° horizontally resolved bins. An example figure from Jan 29, 2005 is shown in Figure 2.1, while an average for 2005 (global) is shown in Figure 2.2. OMNO2 pixel resolution is 40 km by 130 km. NO₂ measured by OMI is used to check whether NO₂ is well represented by GEOS-Chem (see Section 2.4.6 for the comparison between this product and GEOS-Chem calculations). It is also used to form the anthropogenic influence filter for OMHCHO (See Section 2.7.2).



FIGURE 2.2: Average 2005 tropospheric NO₂ from OMNO2d with pixels screened for < 30% cloud cover.

Like other satellite products, OMNO2d is influenced by a priori modelling which is required to convert slant path radiance to vertical columns. These models are generally low resolution (~ 110 km by 110 km), which leads to column smearing and difficulty detecting point sources of high NO emissions (Goldberg2018). Uncertainty in this product arises mostly from the calculation of the AMF (up to 50% of total error) (Lorente2017).

2.2.1.3 Aerosol optical depth

Aerosols in the atmosphere can be seen through their affects on light. Smoke and dust can be seen as an increase in aerosol optical depth (AOD) (see Section 2.3.2). This is due these particles scattering and absorbing UV radiation (Ahn2008). A data product provided by Earthdata (https://disc.gsfc.nasa.gov/datasets/OMAERUVd_V003/summary) called OMAERUVd (DOI: 10.5067/Aura/OMI/DATA3003) is used in this thesis.

OMAERUVd provides a useful dataset which allows detection of areas which may be smoke affected. The product contains AOD and aerosol absorption optical depths (AAOD) at three wavelengths (354, 388, and 500 nm), along with UV aerosol index (UVAI). The OMAERUVd product is level three gridded daily data, based on quality filtered level two swath pixels which are then gridded by averaging. The product is most sensitive to error in the form of sub-pixel scale cloud interference, so I select AAOD as the basis for my smoke filter as it is least affected by clouds (Ahn2008).

In this work AAOD is mapped from $1 \times 1^\circ$ horizontal resolution to $0.25^\circ \times 0.3125^\circ$ using nearest value mapping. The AAOD at 500 nm wavelength is used to determine smoke influence, although any of the provided wavelengths would be affected by smoke plumes and could also be used. This daily AAOD is compared to a threshold to create a daily smoke filter, any areas with $\text{AAOD} > 0.03$ are considered to be potentially smoke plume affected (see Section 2.7.1).

2.2.1.4 Active fires

MOD14A1 is a gridded daily satellite based dataset of fire counts at $1 \times 1 \text{ km}^2$ horizontal resolution. Fire observations are performed four times daily from Terra (10:30 LT, 22:30 LT) and Aqua (01:30 LT, 13:30 LT). The fire pixels are detected based on parameters including apparent pixel temperature and the nearby background temperature. The dataset is obtained from NASA Earth Observations (NEO) that is part of the EOS Project Science Office at the NASA Goddard Space Flight Center https://neo.sci.gsfc.nasa.gov/view.php?datasetId=MOD14A1_M_FIRE. This product is downloaded and binned into a lower resolution (using the sum of fire pixels) to create an active fire influence mask (see Section 2.7.1).

2.2.1.5 Carbon monoxide

In Chapter 4, potential biomass burning plumes are identified using satellite observations of CO from the AIRS (Atmospheric Infra-red Sounder) instrument aboard the Aqua satellite (AIRS3STD). CO is used as a proxy for biomass burning plumes, and used to qualitatively attribute ozone intrusion events as explained in Section 4.3.5.

This is a separate method of detecting fire influence near specific sites through visual analysis.

2.2.1.6 Uncertainties

While satellite data is effective at covering huge areas (the entire earth) it only exists at a particular time of day, is subject to cloud cover, and generally does not have fine horizontal or vertical resolution. Concentrations retrieved by satellites have large uncertainties, which arise in the process of transforming spectra into total column measurements, as well as instrument degradation (satellite instruments are hard to tinker with once they are launched). Uncertainty in transforming satellite spectra comes from a range of things, including measurement difficulties introduced by clouds, and instrument sensitivity to particular aerosols (**Millet2006**). Many products require analysis of cloud and aerosol properties in order to estimate concentration or total column amounts (**Palmer2001; Palmer2003; Marais2012; Vasilkov2017**). The main source of error in satellite retrievals of HCHO are due to instrument detection sensitivities, and calculation of the air mass factor (AMF) which converts slanted light path concentrations into a vertical profile (**Millet2006**). Calculations of the AMF performed by different groups tend to agree fairly well, as long as all the a priori and ancillary data is similar. Large differences can occur depending on the a priori vertical profile, trace gas concentrations, and cloud properties (**Lorente2017**). Choice of RTM and interpolation operations have a relatively small affect compared to the assumed state of the atmosphere, with high structural uncertainty introduced at this stage of AMF calculation - as shown in **Lorente2017**.

A common way of reducing satellite uncertainty is through oversampling or temporal averaging. This is done frequently for trace gases (which are often near to the detection limit over much of the globe). For example: **Vigouroux2009** reduce the measurement uncertainty (in SCIAMACHY HCHO columns) by at least a factor of 4 through averaging daily over roughly 500km around Saint-Denis, and only using days with at least 20 good measurements. Another example of this can be seen in **Dufour2009**, where monthly averaging is used to decrease the measurements uncertainty at the cost of temporal resolution.

In cloudy, hazy or polluted areas measurements are more difficult to analyse (**Palmer2003; Marais2014**). Recent work by **Vasilkov2017** showed that updating how the surface reflectivity is incorporated into satellite measurements can change the retrievals by 50 % in polluted areas.

2.2.2 Model datasets

2.2.2.1 GEOS-Chem output

GEOS-Chem model output is used extensively in this thesis and is discussed in more detail in Section 2.4. Section 2.4.7.1 specifically describes the model outputs used in this thesis. These are generally resolved to 47 vertical levels from the ground up to 0.01 hPa, at $2^\circ \times 2.5^\circ$ horizontal resolution.

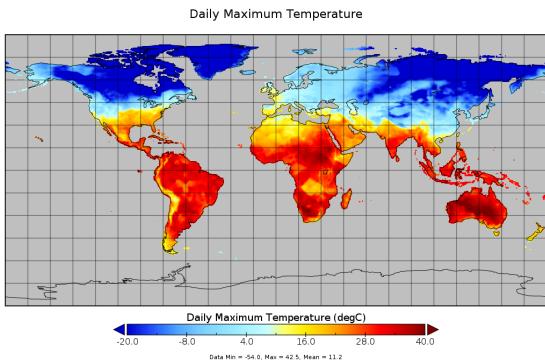


FIGURE 2.3: CPC daily maximum temperature dataset output for 1, Jan, 2005.

2.2.2.2 Meteorological reanalysis

Synoptic scale weather patterns are taken from the European Centre for Medium-range Weather Forecasts (ECMWF) Interim Reanalysis (ERA-I) (Dee2011). These are used in Chapter 4 to determine typical weather systems for stratospheric ozone intrusions. The version used was ERA-Interim, which was the most up to date at the time (2016) but has since been superseded by ERA5.

2.2.2.3 Surface temperatures

The Climate Prediction Center (CPC) provides a product with maximum daily land-surface temperature at $0.5 \times 0.5^\circ$ horizontal resolution. This data is used to check the correlation between HCHO and temperature at a higher resolution than is provided by GEOS-Chem output. A full description of the data can be found at <https://www.esrl.noaa.gov/psd/data/gridded/data.cpc.globaltemp.html>. CPC Global Temperature data is provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their web site at <https://www.esrl.noaa.gov/psd/>. An example of one day of land temperature output is shown in Figure 2.3.

2.2.3 Campaign datasets

In this thesis data from several measurement campaigns are used to examine accuracy of modelled data at specific sites. Figure 2.4 shows the locations of BVOC measurement sites in the top panel, and release sites for ozonesondes in the bottom panel. These took place over disparate times, and are in situ measurements which are hard to directly compare against GEOS-Chem output which is averaged over a large horizontal space.

The campaign datasets provide three separate time series for brief periods of both isoprene and formaldehyde. Figure 2.5 shows these along with the detection limits and also shows isoprene measurements superimposed over a single year. It is apparent that more measurements are required to see more than the daily cycles.



FIGURE 2.4: Locations of VOC measurements (top panel) and ozonesonde release sites (bottom panel). Inlaid in top panel is the flight paths over Australia of the HIPPO campaign.



Figures/OMI_link/campaigns_compared.png

FIGURE 2.5: Top: MUMBA, SPS1, and SPS2 time-series for HCHO (orange) and isoprene (magenta), along with detection limits (dashed).

Bottom: isoprene measurements superimposed onto a single year.

TABLE 2.1: Detection limits for MUMBA

Dates	HCHO (ppb)	Isoprene (ppb)	Ozone (ppb)
21/Dec/2012 - 29/Dec/2012	0.205	0.003	0.5
29/Dec/2012 - 18/Jan/2013	0.105	0.005	0.5
19/Jan/2013 - 15/Feb/2013	0.186	0.003	0.5

2.2.3.1 Measurements of Urban, Marine and Biogenic Air (MUMBA)

The MUMBA campaign ([PatonWalsh2017](#)) measured various compound abundances including isoprene, formaldehyde, and ozone from 21 December 2012 to 15 February 2013. These measurements took place in Wollongong, 10 m above ground level (40 m above sea level). Ozone was measured by Thermo UV absorption with 1-minute time resolution averaged into hourly outputs. Isoprene and HCHO were measured by Ionicon Proton-Transfer-Reaction Mass spectrometer (PTR-MS), with a time resolution of 3-minutes, averaged each hour. Detection limits varied due to instrument conditions, and are listed in Table 2.1. The full dataset has been published on PANGAEA (DOI:[10.1594/PANGAEA.871982](#)) ([Guerette2018](#)).

In this thesis we assume uncertainty in this product is as estimated by [Dunne2018](#) at (50%). Although the uncertainty determined through calibration measurements was only 15% ([Guerette2018](#)), this does not account for competing trace gas interference (such as furan). The readings are re-sampled to hourly averages. Measurements below the detection limit are set to half of the detection limit when reading this dataset.

2.2.3.2 Sydney Particle Studies (SPS1, SPS2)

Two trace gas measurement campaigns took place at the Westmead air quality station. Stage 1 (SPS1) from 5 February to 7 March, 2011 and stage 2 (SPS2) from 16 April to 14 May, 2012. Two instruments measured VOC concentrations: a PTR-MS, and a gas chromatographer (GC) with a flame ionisation detector (FID). The PTR-MS uses chemical ionisation mass spectrometry and can quantify VOCs at high temporal resolution (< 1 s). It was calibrated several times per day against HCHO, isoprene, α -pinene, and several other VOCs, further measurement specifics can be found in [Dunne2018](#).

The output lists hourly averaged ppbv concentrations of trace gases based on the mass to charge ratio (m/z), which for isoprene is 69. It is possible that other chemicals (such as furan, with the same m/z) interfered with this value, especially at low ambient isoprene concentrations and towards the end of autumn (SPS2) when wood fires start to become frequent ([Guerette2018](#)). The GC-FID analysed samples collected in multi-absorbent tubes, with lower temporal resolution but no interference. Further details for this method can be found in [Cheng2016](#). GC-FID data is averaged from 0500-1000 LT, and 1100-1900 LT, while PTR-MS data is averaged hourly. This includes significant differences between measurement devices when detecting isoprene, potentially due to interfering compounds in the PTR-MS ([Dunne2018](#)).

Figure 2.6 shows isoprene and formaldehyde over the course of these two campaigns, as well as the detection limits (dashed lines), as measured by PTR-MS. In order to compare with GEOS-Chem output (see Section 2.4) a daily average and an midday time (13:00-14:00 LT) average are both created from these data. In averaging,

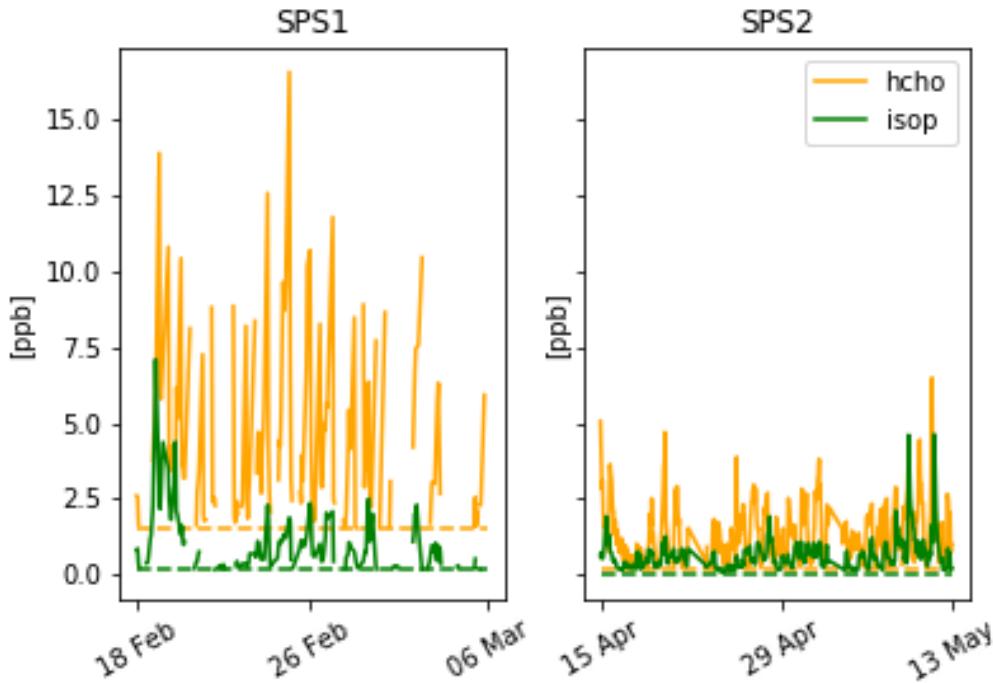


FIGURE 2.6: SPS HCHO (yellow) and isoprene (green) time series, along with detection limits (dashed). SPS 1 (left) took place in late summer 2011, while SPS 2 (right) occurred during autumn 2012.

any measurements below the machine detection limit are set to half of the detection limit, as done in [Lawson2015](#). This should minimise any introduced bias.

2.2.3.3 Ozonesondes

Ozonesonde data come from the World Ozone and Ultraviolet Data Centre (WOUDC). Ozonesondes are weather balloons which measure from the surface to around 35km. Ozonesondes provide a high vertical resolution profile of ozone, temperature, pressure, and humidity. Generally the instrument will perform 150-300 measurements in the troposphere with ozone mixing ratios quantified by an electrochemical concentration cell (<http://www.ndsc.ncep.noaa.gov/organize/protocols/appendix5/>).

Ozonesondes are launched approximately weekly from Melbourne (38° S, 145° E), Macquarie Island (55° S, 159° E) and Davis (69° S, 78° E). Melbourne, a major city with more than 4 million residents ([ABS2016](#)), may be affected by anthropogenic pollution in the lower troposphere. Actual releases are north of the central business district in the Broadmeadows suburb. Macquarie Island is in the remote Southern Ocean and unlikely to be affected by any local pollution events. Davis (on the coast of Antarctica) is also unlikely to experience the effects of anthropogenic pollution. More information on this dataset is given in Section 4.3.

2.2.3.4 Wollongong

Upon the roof of the Chemistry building in Wollongong lies a solar Fourier transform infra-red spectrometer (FTIR) which measures HCHO (amongst other gases) on the path of light between the instrument and the sun. TODO: details about these measurements.

Figure TODO: shows Wollongong FTIR measurement profiles of HCHO with an assumed averaging kernel and then one calculated from GEOS-Chem, along with the two averaging kernel mean and standard deviation. TODO: Discuss this conversion and results seen here... The bottom panel shows the time series of total column HCHO of both GEOS-Chem and the FTIR instrument in molecules cm^{-2} .

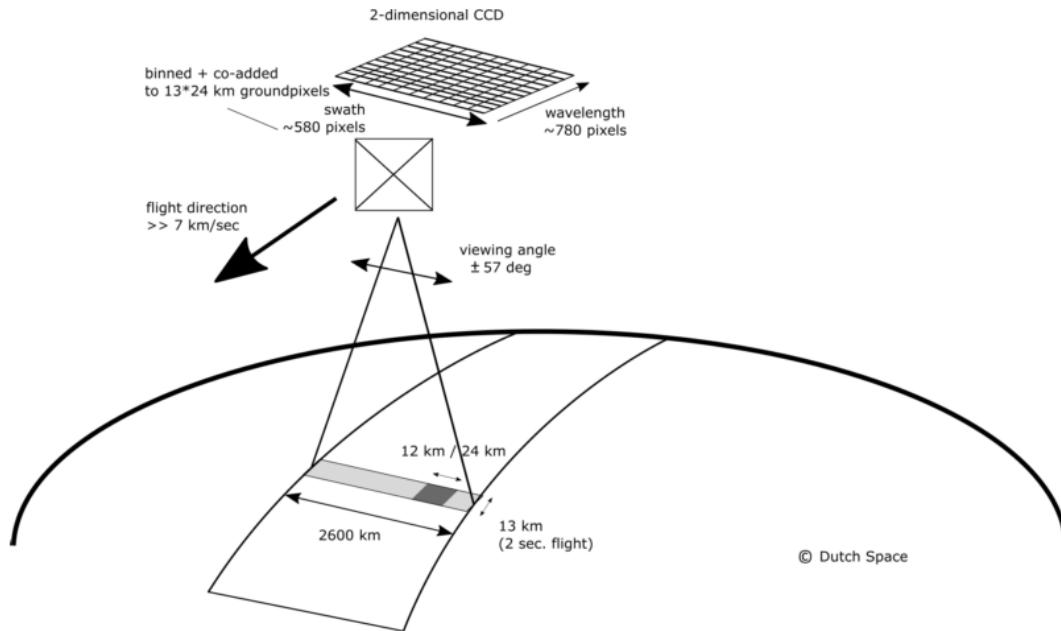
2.2.3.5 Uncertainties

In situ measurements contain errors, and depending on the device used and chemical being measured this error can be significant. The major sources of uncertainty in measurement techniques included interference from non-target compounds and under-reporting (**Dunne2018; Guerette2018**). Overall isoprene uncertainty in measurements analysed by **Dunne2018** was a factor of 1.5 to 2. This can feed into uncertainties in modelling and satellite retrievals, as verification and correlations are affected.

2.3 Satellite formaldehyde

One satellite product used extensively in this thesis is named OMHCHO: from NASA's Earth Observing System's "Aura", which provides several other useful datasets (products). Aura orbits the earth in a polar sun-synchronous pattern, circling the earth on a plane coincident with the sun and the poles. Aura houses the Ozone Monitoring Instrument (OMI), a near-UV/Visible Charged Coupled Device (CCD) spectrometer. The OMI instrument onboard Aura has been active since July 2005, it records spectra from 264-504 nm using an array of 60 detectors with mid-resolution (0.4-0.6 nm). This band of wavelengths allows measurements of trace gases (among other quantities) and the formaldehyde product is detailed here.

From here on the word pixel is used to describe one data point retrieved by OMI, each pixel includes a latitude and longitude within OMI's data product. Figure 2.7 shows the details of OMI's detector array and measurement resolutions. OMI measures atmospheric trace gases including NO_2 , SO_2 , BrO , HCHO , O_3 , and aerosols. OMI measurements occur from right to left on a band covering 115° , resulting in swaths of around 2600 km, with pixel sizes from $13 \times 24 \text{ km}^2$ at nadir to $26 \times 135 \text{ km}^2$ at the swath edges (**Abad2015**). The swaths cover Earth daily, both on the light and dark side of the planet, only daytime measurements provide useful near-UV/Visible information. While satellite measurements can only be used during daytime hours, HCHO lifetimes are sufficiently short that any night-time chemistry will not affect midday observations (**Wolfe2016**).



Channel	Wavelength range	Spectral resolution	Spectral sampling	Ground pixel size
UV1	264–311 nm	0.63 nm = 1.9 px	0.33 nm px ⁻¹	13 × 48 km
UV2	307–383 nm	0.42 nm = 3.0 px	0.14 nm px ⁻¹	13 × 24 km
VIS	349–504 nm	0.63 nm = 3.0 px	0.21 nm px ⁻¹	13 × 24 km

FIGURE 2.7: Figure 1 and Table 1 from Schenkeveld2017, with the following caption “An impression of OMI flying over the Earth. The spectrum of a ground pixel is projected on the wavelength dimension of the charge-coupled device (CCD; the columns). The cross-track ground pixels are projected on the swath dimension of the CCD (the rows). The forward speed of 7 kms^{-1} and an exposure time of 2 s lead to a ground pixel size of 13 km in the flight direction. The viewing angle of 114° leads to a swath width on the ground of 2600 km.” The table shows the optical properties for OMIs three channels.

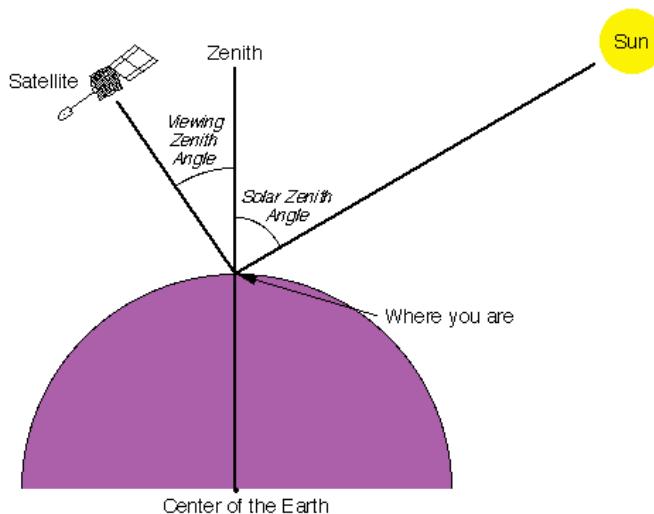


FIGURE 2.8: Solar and viewing zenith angles, image copied from [SZA_Image](#), originally from a NASA website.

The latest OMHCHO algorithm uses a shape factor determined from GEOS-Chem using 47 vertical levels at monthly temporal resolution and $2^\circ \times 2.5^\circ$ latitude by longitude horizontal resolution ([Abad2015](#)). The GEOS-Chem model has been substantially updated since then. In this thesis the more recent version V10.01 is used to recalculate the vertical column HCHO (details are shown in Section 2.6).

OMI uses a Differential Optical Absorption Spectroscopy (DOAS) based technique to read HCHO along the path of light which reaches the satellite instrument. The first step is to determine how much HCHO is in the path of light between the sun and detector. Measurements done using DOAS often apply a forward radiative transfer model (RTM) such as LIDORT (see Section 2.3.4) in order to determine a trace gas' radiative properties at various altitudes. The forward RTM used for satellite data products also involve functions representing extinction from Mie and Rayleigh scattering, and the effect of these on spectra. These RTM are also required to account for (often estimated) atmospheric parameters such as albedo. The next step is to transform the calculated amounts along the non-vertical light path into vertical column amounts. This is done by applying an AMF. In the absence of atmospheric scattering a simple geometric AMF can be defined as a function of the solar zenith angle. The solar zenith angle (θ_s) and the satellite viewing angle (θ_v) are shown in image 2.8. However, in the UV-VIS region of the spectrum, Rayleigh and Mie scattering (see Section 2.3.2) must be accounted for.

Atmospheric HCHO detected by satellite requires that other trace gases with similar features near the HCHO affected wavelengths are accounted for. A DOAS fit determines the total column amount of a trace gas along the path that the instrument views. This uses the Beer-Lambert law where radiance is reduced as light travels through a medium. I use the NASA OMHCHOv003 data product ([Abad2015](#)), with HCHO determined using the spectral window 328.5 nm–356.5 nm. The algorithm used is based on direct fitting of radiances, and accounts for competing absorbers,

under-sampling, and Ring effects. An OMI radiance measurement over the remote Pacific ocean is used instead of an irradiance measurement. This means that the slant columns (Ω_S) are formed from the spectra differential with respect to the radiance reference column over the Pacific. The full method details for slant column retrieval by OMI are outlined in supplemental Section ??, or in the technical document (DOI: 10.5067/Aura/OMI/DATA2015). Slant columns range from $\sim 4 \times 10^{15}$ to $\sim 6 \times 10^{16}$ molec cm⁻², with uncertainties from 30% (larger columns) to over 100% (smaller columns) (Abad2015).

2.3.1 Pixel filtering

This thesis uses the level 2 product swath output from the NASA earth data web portal. OMHCHO level two data includes 14-15 daily swaths of measurements. Each swath contains roughly 9×10^4 pixels, each of which includes latitude, longitude vertical column HCHO, along with all the ancillary data required to make the vertical column and several data quality metrics. The OMHCHO dataset has a quality flag which can be used to remove unlikely or poor satellite measurements. The states represented by this quality flag are shown in Table 2.2 which is reproduced here (originally from Kurosu2014). First all *good* pixels (those with QA flag equal to 0) are read into a long list (roughly 1 million per day). These are then filtered by solar zenith angle (SZA) and latitude, similarly to other works (Marais2012; Barkley2013; Bauwens2016; Zhu2016). Pixels with cloud fraction greater than 40% are removed after being used in determining the reference sector correction, as is performed in Abad2015; DeSmedt2015. Filtering bad or missing measurement pixels is performed prior to any other filtering, this includes the datapoints affected by the row anomaly through a particular flag (see https://www.cfa.harvard.edu/atmosphere/Instruments/OMI/PGEReleases/READMES/OMHCHO_README_v3.0.pdf). This anomaly affects radiance data at particular viewing angles, corresponding to a row on the CCD detectors, and is dynamic over time (Huang2018). The slant columns affected are flagged and removed before any further processing.

Each ~ 90 minutes the Aura satellite sweeps over the sunny side of the planet of which around 50 k – 80 k of the roughly 90 k pixels are classified as good. Each pixel contains several important pieces of data which are needed for recalculation of the HCHO vertical column: the total column of HCHO (Ω ; molec cm⁻²), cloud fraction, associated shape factor, AMF, geometric AMF, scattering weights and their vertical altitudes (hPa), viewing zenith angle, solar zenith angle, latitude, longitude, OMI sensor track, main data quality flag, cross track flag, and total column uncertainty. All of these data are needed in order to reconstruct the total vertical column using a modelled a priori shape factor rather than NASA's included a priori shape factor. Each pixel includes an estimate of the cloud fraction created using the OMI cloud product OMCLDO2. If greater than 40% of a pixel measurement is cloudy (ie. cloud fraction > 0.4) then the pixel is removed from subsequent analysis. This removes around 30% of the pixels which remain after filtering out the bad or missing data. Satellite measurements pole-wards of 60° north or south are removed as well as measurements with SZA greater than 60°. Measurements taken with high SZA are unlikely to have much information from closer to the ground, and often have high uncertainty.

TABLE 2.2: OMI quality flag values table from **Kurosu2014**

Value	Classification	Rational
0	Good	Column value present and passes all quality checks; data may be used with confidence.
1	Suspect	Caution advised because one or more of the following conditions are present: <ul style="list-style-type: none"> • Fit convergence flag is < 300 but > 0: Convergence at noise level • Column $+2\sigma$ uncertainty $< 0 <$ Column $+3\sigma$ uncertainty • Absolute column value $>$ Maximum column amount ($1\text{e}19 \text{ molec cm}^{-2}$)
2	Bad	Avoid using as one of the following conditions are present: <ul style="list-style-type: none"> • Fit convergence flag is < 0 : No convergence, abnormal termination • Column $+3\sigma$ uncertainty < 0
< 0	Missing	No column values have been computed; entries are missing

Due to noise and the differential measurement technique, small negative columns are present in the satellite product which are not removed so as not to introduce a bias. However, some very large negative values which represent an unflagged error in the retrieval process are removed. Unlikely high positive measurements are also removed, leaving only measurements within the density range -0.5×10^{16} to 1×10^{17} molec cm^{-2} , as is performed by **Zhu2016**. This filter is required due to currently unexplained large negative values which occur in the OMI HCHO product increasingly over time. Figure 2.9 shows how unfiltered HCHO columns are affected by a small set of highly negative values which heavily affect the mean column amount over any region. The highly negative values can be seen around the -10^{19} molecules cm^{-2} region.

2.3.2 DOAS

The DOAS technique uses solar radiation absorption spectra to measure trace gases through paths of light. Beer's law states

$$T = I/I_0 = e^{-\tau} \quad (2.2)$$

with T being transmittance, τ being optical depth, and I, I_0 being radiant flux received at instrument and emitted at source respectively. The Beer-Lambert law of extinction allows spectroscopic measurement of absorbing chemical species (absorbers) in the

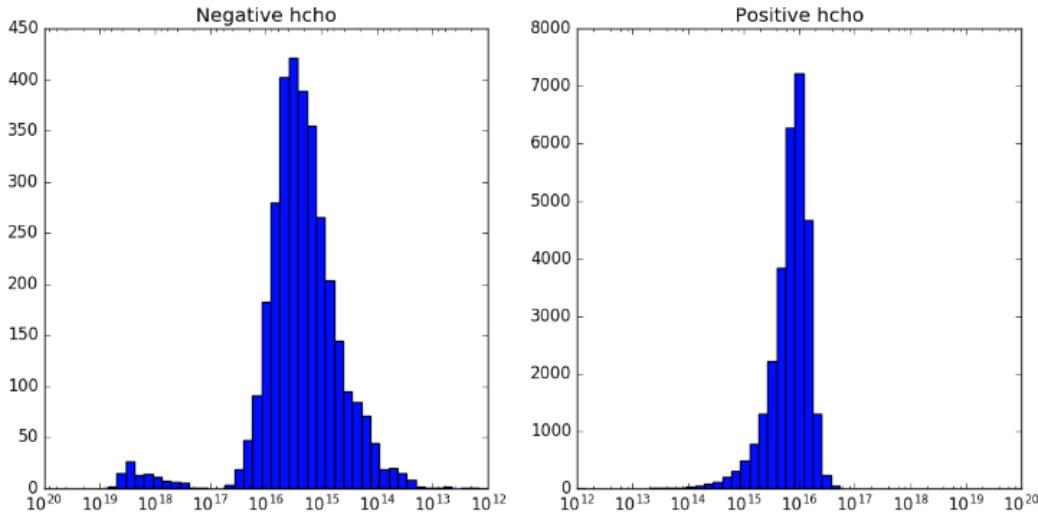


FIGURE 2.9: Column density histograms for a subset of OMI swaths over Australia on the 18th of March 2013. Negative entries are shown in the left panel, positive in the right. Note the different scale between negative and positive panels.

atmosphere:

$$I_B = I_{B_0} e^{-\tau_s} \quad (2.3)$$

where I_B, I_{B_0} is backscattered intensity with and without the absorber respectively, and τ_s is the optical thickness of the absorber along the measured path between source and instrument.

τ can be described using the scattering and absorption cross section area (α , cm²) and density (η , molec cm⁻³) of an absorber as follows:

$$\tau = \int \alpha(s) \eta(s) ds \quad (2.4)$$

τ through a medium is the sum of optical thicknesses of each absorber within the measured path (s) substituting Equation 2.4 into Equation 2.3 leads to

$$I = I_0 \exp \left(\sum_i \int \eta_i \alpha_i ds \right)$$

Where i represents a chemical species index, and the integral over ds represents integration over the path from light source to instrument.

Another way of describing optical depth (also called optical thickness) is the natural logarithm of the ratio of incident radiant power to transmitted radiant power through a material (from Equation 2.3). In the atmosphere we are interested in the optical depth of various chemical species, and we use incoming solar radiation to determine this. The difference between solar radiation at the top of the atmosphere and

the earth's surface defines the atmospheric optical depth along the path of observation.

$$\tau = \ln \frac{\phi_e^i}{\phi_e^t}$$

where ϕ_e^i is radiant flux seen at the earth surface, ϕ_e^t is the solar radiant flux which arrives at the top of the atmosphere. In the atmosphere, optical depth can be due to several factors including scattering, chemical absorbance, and aerosols.

2.3.3 Air mass factor (AMF)

To convert the trace gas profile from a reflected solar radiance column (slanted along the light path) into a purely vertical column requires calculations of an air mass factor (AMF). In satellite data, the AMF is typically a scalar value for each horizontal grid point which will equal the ratio of the total vertical column density to the total slant column density. This value requires calculations to account for instrument sensitivities to various wavelengths over resolved altitudes, and is unique for each trace gas under consideration. An AMF characterises measurement sensitivity to a trace gas at various altitudes **Palmer2001**. **Lorente2017** show that AMF calculations can be the largest source of uncertainty in satellite measurements. Another way of describing AMFs are as measures of how radiance at the top of the atmosphere (TOA) changes with trace gas optical depths at specific altitudes (**Lorente2017**). Calculation of the AMF is important as it is multiplied against the estimated slant columns in order to give vertical column amounts.

DOAS column retrievals are an integration of a trace gas over the instruments viewing path, in order to convert this total to a vertically distributed column a few assumptions and estimates are required. The vertical profile of a trace gas is assumed or estimated via a CTM, while its scattering and radiative properties are calculated at prescribed altitudes using an RTM. These properties are combined into a single array called the AMF. Two examples of this are GOME-2 (on the MetOp-A satellite) products (http://atmos.caf.dlr.de/gome/product_hcho.html), and OMI products which respectively use LIDORT combined with IMAGESv2 and GEOS-Chem for processing (**Chance2002; Abad2015**). AMFs are unique to each trace gas and due to their complexity and the influence of cloud cover they remain one of the largest error sources in remote sensing of BVOCs (**Palmer2001; Millet2006**). **Lamsal2014** recommends that when comparing satellite data to models, the AMF should first be recalculated using the model as an a priori. This is in order to remove any a priori bias between model and satellite columns.

Related to the AMF is the averaging kernal (AK), which is used to handle instrument measurements which are sensitive to gas concentrations at different altitudes through the atmosphere. DOAS methods can be heavily influenced by the initial estimates of a trace gas profile (the a priori) which is often produced by modelling, so when comparing models of these trace gases to satellite measurements extra care needs to be taken to avoid introducing bias from differing a priori assumptions. One way to remove these a priori influences is through the satellite AK (or AMF), which

takes into account the vertical profile of the modelled trace gas and instrument sensitivity to the trace gas (**Eskes2003; Palmer2001**). This process is called deconvolution ($\Omega = AK \times VC_{satellite} + \times(I - AK)VC_{apriori}$) of the AK of the satellite instrument. The AK represents sensitivities to each species at multiple altitudes through the atmosphere and in the case of OMI, can be approximated from the scattering weights ($\omega(z)$) function as follows:

$$AK(z) = \frac{\omega(z)}{AMF} \quad (2.5)$$

This is an approximation for the OMI product, which does not include the AK but does include the ω and AMF, as explained in **Abad2015**.

2.3.4 LIDORT

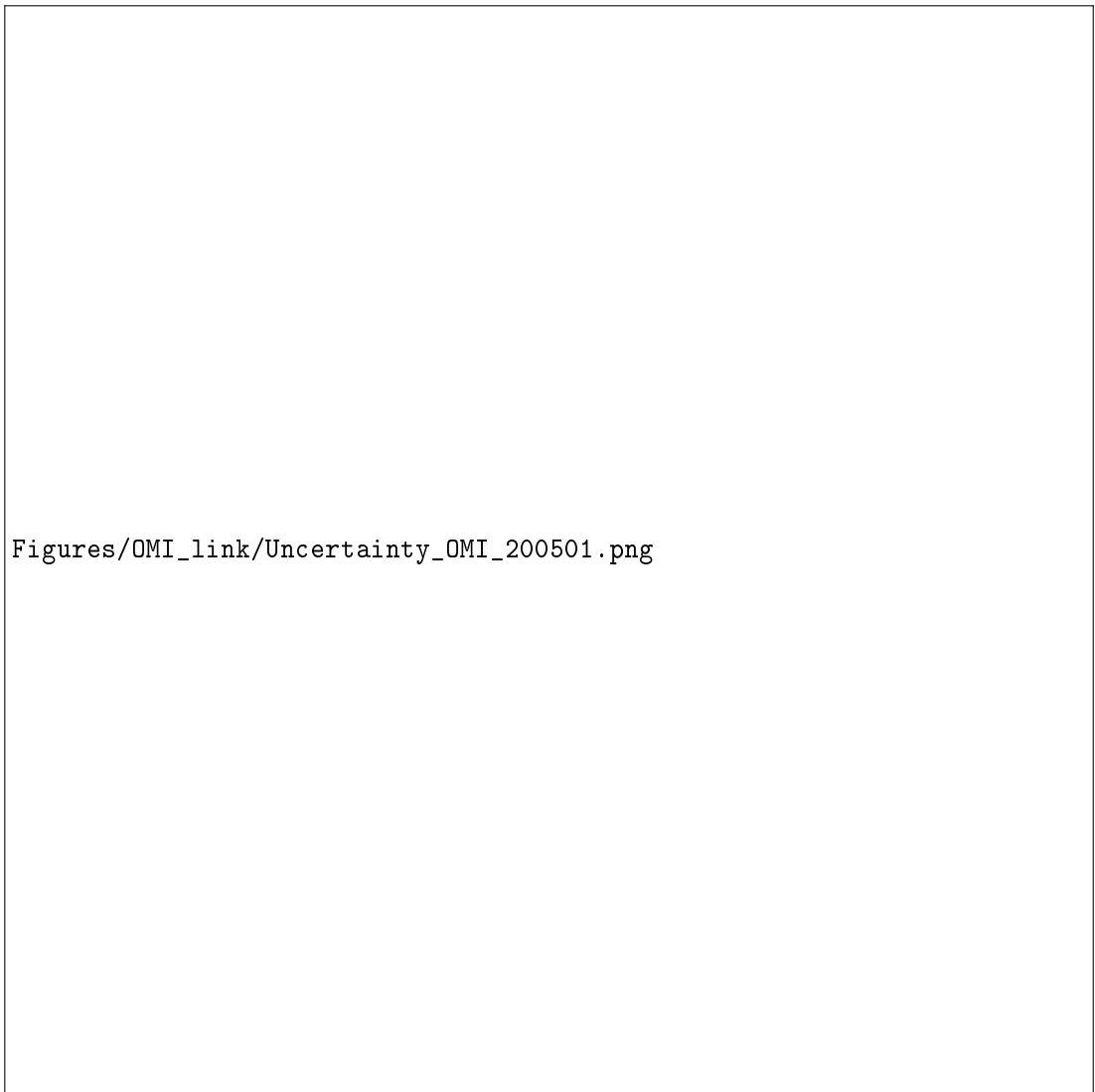
LIDORT is a model of LInearized Discrete Ordinate Radiative Transfer, used to determine backscatter intensities and weighting functions at arbitrary elevation angles (**Spurr2001**). The model solves radiative transfer equations and can be used to determine various atmospheric column measurement attributes such as optical depth, ring effects, and scattering. These radiative properties (or at least estimates thereof) are required when measuring trace gases in the atmosphere through a long path such as seen by satellites (**Palmer2001; Martin2002a; DeSmedt2015; Abad2015**).

2.3.5 Uncertainty

Uncertainty in a single pixel for OMHCHO is roughly the same magnitude as HCHO background levels. Each pixel has $\sim 2 \times 10^{14}$ molec cm $^{-2}$ uncertainty, which is $5\times$ higher than GOME. However, there are $\sim 100 - 200\times$ as many measurements allowing a greater reduction of uncertainty with averaging. This is due to the smaller footprint and better temporal resolution of OMI (**Chance2002; Millet2008**). The finer nadir resolution of OMI (13 by 24 km 2) compared to other satellites also reduces cloud influence (**Millet2006; Millet2008**). The top row in Figure 2.10 shows OMI HCHO columns binned to at $0.25^\circ \times 0.3125^\circ$ longitude by latitude averaged over one day and one month (with and without filtering). Row two shows uncertainty of the satellite data after averaging. It is clear that one day of satellite data is too uncertain when binned at $0.25^\circ \times 0.3125^\circ$ horizontal resolution; however, after a month (with or without filtering) the uncertainties become manageable. If we assume the uncertainty is random error, and not bias introduced through calculation techniques, then we are able to reduce the uncertainty through averaging. High resolution low detection limit estimates can be built up using “oversampling”, which averages satellite measurements over time (**Zhu2014**). Uncertainty in satellite recalculations, along with other factors is analysed in Section ??.

In this thesis, HCHO columns (pixels) with cloud fractions over 40% are filtered as done in **Palmer2001**, which introduces a clear-sky bias to any monthly averages. This is due to HCHO being lower on unrecorded cloudy days. This bias has been measured as a 13% positive monthly mean bias (**Palmer2001; Surl2018**).

Recently **Schenkeveld2017** analysed the performance over time of the instrument and found irradiance degradation of 3-8%, changed radiances of 1-2%, and a stable wavelength calibration within 0.005-0.020 nm. These changes are measured excluding



Figures/OMI_link/Uncertainty_OMI_200501.png

FIGURE 2.10: Top row shows $0.25^\circ \times 0.3125^\circ$ binned OMHCHO columns with one day, one month, and one month with non-biogenic masking applied from left to right respectively. Bottom row shows the uncertainty for each grid square after averaging.

the row anomaly (RA) effect, which is relatively stable since 2011, although it is still growing and remains the most serious concern. Their analysis of OMI concludes that data is still of high quality and will deliver useful information for 5-10 more years, with radiance only changing by 1 – 2% outside of RA impacted areas. An analysis of the row anomaly by **Huang2018** state that measurements remain suitable for scientific use, with recommendation for further evaluation. The RA began in June 2007, with some cross-track rows seemingly blocked. The most likely cause is some instrument insulation partially obscuring the radiance port (**Schenkeveld2017**).

In satellite HCHO products, concentrations over the remote Pacific ocean are sometimes used to analyse faulty instrument readings. This is due to the expected invariance of HCHO over this region. For instance GOME (an instrument which measures trace gases onboard the ERS-2) corrects for an instrument artefact using modelled HCHO over the remote Pacific (**Shim2005**). OMI HCHO products use a similar technique to account for sensor plate drift and changing bromine sensitivity (**Abad2015**). Uncertainty in the OMI satellite instrument is calculated by the Smithsonian Astrophysical Observatory (SAO) group using the uncertainty in backscattered radiation retrievals (**Abad2015; Abad2016**). Another method of calculating the uncertainty is used by the Belgian Institute for Space Aeronomy (BIRA) group, who determine uncertainty from the standard deviation of HCHO over the remote Pacific ocean (**DeSmedt2012; DeSmedt2015**).

For many places the tropospheric column HCHO measured by satellite is biased low, **Zhu2016** examine six available datasets and show a bias of 20 - 51% over south east USA when compared against a campaign of aircraft observations (SEAC⁴RS). **DeSmedt2015** also found OMI and GOME2 observations were 20 - 40% lower than ground based vertical profiles, and **Barkley2013** determine OMI to be 37% low compared with aircraft measurements over Guyana. These bias can be corrected by improving the assumed a priori HCHO profiles which are used to calculate the AMFs of the satellite columns. **Millet2006** examine OMI HCHO columns over North America and determine overall uncertainty to be 40%, with most of this coming from cloud interference. **Millet2008** shows that there also exists some latitude based bias, as well as a systematic offset between the OMI and GOME instruments. This does not appear to be due to the different overpass times of the two instruments.

AMF calculation often dominates the total uncertainty in satellite retrievals, especially in polluted regions (**Lorente2017**). In scenarios where the gas is enhanced in the lower troposphere, AMF calculation is the largest uncertainty in satellite measurements. In polluted environments the structural uncertainty is estimated at 42 %, or 31 % over unpolluted environments (**Lorente2017**). Another impact often not included in uncertainty calculations is the structural uncertainty of retrieval methods. The structural uncertainty of AMF calculation approaches used by different retrieval groups comes from how the AMF is calculated, rather than uncertainty in the calculation components. The importance of a priori and ancillary data (such as surface albedo and cloud top height) sharply affects the structural uncertainty (**Lorente2017**).

2.4 GEOS-Chem

2.4.1 Overview

GEOS-Chem is a well supported global, Eulerian CTM (see Section 1.5.2) with a state of the science chemical mechanism, with transport driven by meteorological input from the Goddard Earth Observing System (GEOS) of the NASA Global Modeling and Assimilation Office (GMAO). Chemistry, transport, and meteorology are simulated at 15 minute time steps within a global set of 3-D boxes. Emissions are either prescribed by inventories or modelled (e.g., fire emissions are pulled from the global fire emissions database GFED4).

GEOS-Chem simulates more than 100 chemical species from the earth's surface up to the edge of space (0.01 hPa) and can be used in combination with remote and in situ sensing data to give a verifiable estimate of atmospheric gases and aerosols. It was developed, and is maintained, by Harvard University staff as well as users and researchers worldwide. In this thesis I use version 10.01 of GEOS-Chem, which outputs up to 66 chemical species (tracers) in the standard run, at $2^\circ \times 2.5^\circ$ horizontal resolution, with 47 levels up to the top of the atmosphere (TOA at 0.01 hPa).

Global CTMs are often run using one or several emission models (or the output from them) to determine boundary conditions. Some of the inventories used by GEOS-Chem are described here. Meteorological fields are taken from NASA's GEOS-5 dataset ($0.5^\circ \times 0.666^\circ$) (Chen2009), which exists up to April 2013. GEOS-5 meteorological fields are used as the boundary conditions driving transport. Fire emissions come from the global fire emissions database (GFED4) product (Giglio2013). Anthropogenic VOC emissions come from the Emission Database for Global Atmospheric Research (EDGAR) inventory, while biogenic VOC emissions are simulated using the MEGAN model (see Section 2.4.5). MEGAN is used to determine biogenic emissions for our default GEOS-Chem simulation. The estimated biogenic VOC emissions are important for accurately simulating chemistry within models, as discussed in Section 1.1.2.

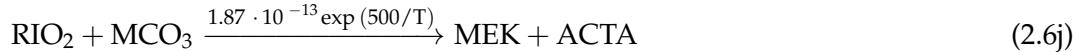
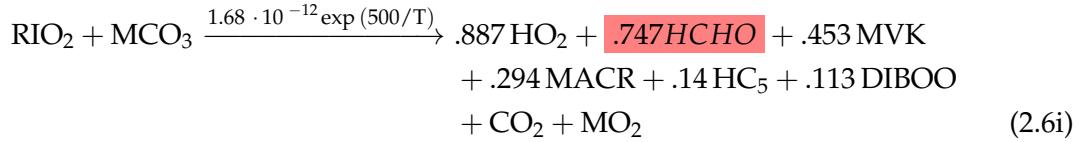
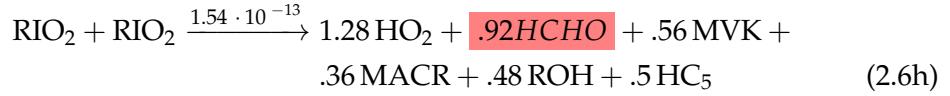
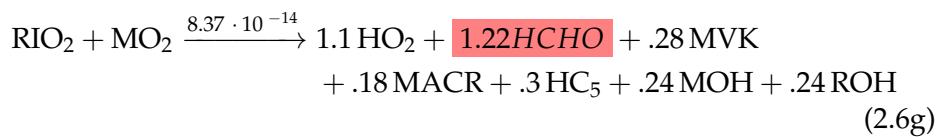
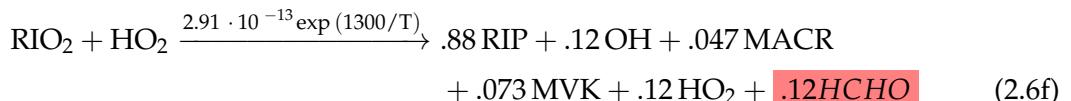
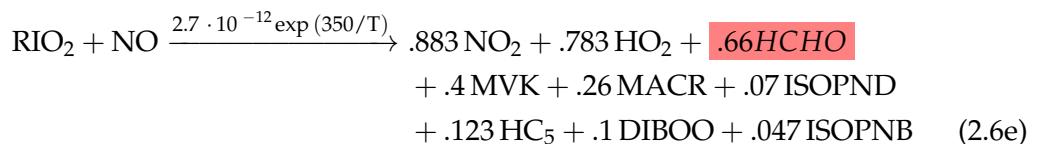
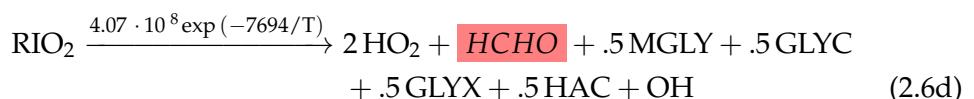
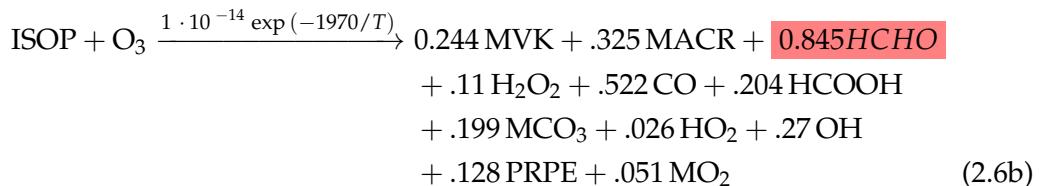
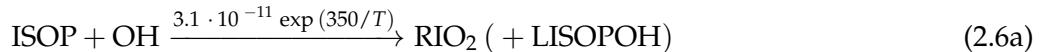
2.4.2 Installing and running GEOS-Chem

GEOS-Chem instructions for download, compilation, and running can be found in the user guide provided by Harvard: <http://acmg.seas.harvard.edu/geos/doc/man/>. In order to build and run GEOS-Chem a high-speed computing system is optimal, as globally gridded chemical calculations can take a long time to perform (for us ~ 70 computation hours per month). I installed GEOS-Chem onto a suitably configured workspace on the Raijin supercomputer at the National Computational Infrastructure (NCI, <http://nci.org.au/>). This workspace included access to compilers and libraries which are needed to build the Fortran based GEOS-Chem source code, and IDL, Python, and various editors and scripting languages to read, run, edit, and analyse both GEOS-Chem and its output. After downloading GEOS-Chem, the code can be compiled with different options for resolution and chemical mechanisms.

2.4.3 Chemical Mechanism

A chemical mechanism is a closed system of chemical reactions and their associated rate constants. Chemical reactions are represented by systems of differential equations to be solved for each gridbox in GEOS-Chem. Simplifications are required due to the large number of reactions which occur in the atmosphere, and the coupled and stiff nature of these reactions which slow down computation of the solutions (**BrasseurJacob2017**). Stiffness in chemical systems of differential equations is due to the massively differing reaction time scales - for instance hydroxyl radicals react within seconds while methane has an atmospheric lifetime of 8-10 years (**Wuebbles2002**).

Some of the important reactions involving isoprene are reproduced here, including reaction rates (k) in the form $k = A \exp -E/RT$, where T is temperature, E is activation energy, and R is the gas constant, A , E , and R are predefined for each reaction. Equation 2.6 lists the main isoprene and sundry reactions, with terms defined in 2.3. LISOPOH is added in order to allow the model to keep track of how much isoprene is oxidised by OH. The mechanisms used in version 10.01 (and its history) are described online at http://wiki.seas.harvard.edu/geos-chem/index.php/NOx-Ox-HC-Aer-Br_chemistry_mechanism.



2.4.4 GEOS-Chem isoprene

The isoprene reactions simulated by GEOS-Chem were originally based on Horowitz1998. The mechanism was subsequently updated by Mao2013, who changed the isoprene nitrate yields and added products based on Paulot2009a; Paulot2009b. They used the yields and reactions of various positional isomers of isoprene nitrates, and their oxidation products. Further mechanistic properties, like isomerisation rates, are based on results from four publications: Peeters2009; Peeters2010; Crounse2011; Crounse2012.

In a chamber with clean air and high NO concentrations, isoprene photooxidation is initially driven by OH addition, followed by NO_x chemistry (150 min - 600 min),

TABLE 2.3: Species or classes from the GEOS-Chem mechanism.

Name	Definition
ACTA	Acetic acid: CH ₃ C(O)OH
DIBOO	Dibble peroxy radical
GLYC	Glycoaldehyde: HOCH ₂ CHO
GLYX	Glyoxal: CHOCHO
HAC	Hydroxyacetone: HOCH ₂ C(O)CH ₃
INO2	RO ₂ from ISOP+NO ₃
ISNP	an isoprene nitrate
ISOPNB	β isoprene nitrates
ISOPND	δ isoprene nitrates
MACR	Methacrolein: CH ₂ =C(CH ₃)CHO
MCO3	Peroxyacetyl radical: CH ₃ C(O)OO
MEK	Methyl ethyl ketone: RC(O)R
MGLY	Methylglyoxal: CH ₃ COCHO
MO2	Methylperoxy radical: CH ₃ O ₂
MOH	Methanol: CH ₃ OH
MVK	Methylvinylketone: CH ₂ =CHC(=O)CH ₃
PRPE	>C ₃ alkenes: C ₃ H ₆ , ...
RIO2	isoprene peroxy radical: ROO
ROH	>C ₂ alcohols

and finally HO_x dominated chemistry. Formation of isoprene nitrates (ISOPN) affect ozone levels through NO_x sequestration, the yields and fate of these nitrates was analysed in **Paulot2009a**. Prior to 2012, oxidation chamber studies were performed in high NO or HO₂ concentrations, giving peroxy lifetimes of less than 0.1 s (**Crounse2012; Wolfe2012**). In most environments NO and HO₂ concentrations are not so high, GEOS-Chem uses production rates for different NO concentrations and peroxy radical lifetimes determined by **Crounse2012**.

2.4.4.1 Oxidation

Crounse2011 examined the isomerisations associated with the oxidation of isoprene to six different isomers of ROO formed in the presence of oxygen. The primary oxidation pathway of isoprene is reaction with the OH radical (*ISOP + OH → RIO₂*) shown in Equation 2.6a. Isoprene undergoes OH addition at the 1 and 4 positions, becoming β (71%) or δ (29%) ROO, although these are not distinguished in the GEOS-Chem mechanism. Secondary oxidative pathways are from ozonolysis (Equation 2.6b) and reaction with NO₃ (Equation 2.6c). These pathways are much slower but can be important in particular scenarios such as inside a pollution plume or at night when NO₃ radicals can build up.

Following isoprene oxidation, seven potential ROO reactions (RIO₂ in Equation 2.6), can occur depending on reactant concentrations and local temperature. **Crounse2011** determined rates and uncertainties involved in these reactions, and studied the rate of formation of C₅-hydroperoxyenals (C₅-HPALD) by isomerisation. Reactions 2.6d -

[2.6j](#) compete to determine the fate of ROO[·], which is often grouped into high or low NO_X concentration pathways.

2.4.4.2 Nitrogen oxide impacts

GEOS-Chem reactions account for high and low NO_X scenarios as determined in [Mao2013](#), based on [Paulot2009a](#). High NO_X has been defined as conditions where NO reactions are the main cause of losses for ROO (Reaction [2.6e](#)) ([Palmer2003](#)). High NO_X concentrations are roughly 1 ppb, while low NO_X concentrations are around 0.1 ppb.

In low NO_X ROO losses occur from isomerisation or reaction with several potential compounds. NO_X + HO₂ (Reaction [2.6f](#)) produces mostly hydroxy hydroperoxides (ISOPOOH), and some HCHO. ISOPOOH can be oxidised (by OH) to produce epoxy-diols, recycling OH ([Paulot2009b](#)). Isomerisation of ROO (Reaction [2.6d](#)) largely produces HCHO, while recycling OH and producing HO₂. This isomerisation accounts for 1,5-H shifts producing MACR, MVK, HCHO, or 1,6-H shifts producing HPALDs. HPALDs can photolyse to regenerate OH and small VOCs ([Crounse2011](#); [Wolfe2012](#); [Peeters2014](#)). Under low NO_X conditions the expected production of HCHO, MVK, and MACR is 4.7%, 7.3%, and 12% respectively. Refer to Section [1.3.3](#) for more information. Less frequently, ROO reacts with itself, MO₂, or MCO₃.

Under high NO_X conditions, the fate of ROO differs depending on how it was formed. The β -hydroxyl reacts with NO_X and produces HCHO (66%), methylvinylketone (40%) (MVK), methacrolein (26%), and β -hydroxyl nitrates (6.7%) (ISOPNB). The δ -hydroxyl reacts with NO to form δ -hydroxyl nitrates (24%) (ISOPND), and ISOPNB (6.7%). These two pathways are combined in Reaction [2.6e](#), which shows the production from ROO and NO.

2.4.4.3 OH

Until recently, models struggled to represent OH in the atmosphere, due to missing chemistry and issues in measurement techniques. Prior to [Mao2012](#), measurements of OH in high VOC regions may have been up to double the real atmospheric OH levels, due to formation of OH inside the instrument. OH regeneration through photolysis of HPALDs in areas with high isoprene emissions are included from [Peeters2010](#). Photolysis of photolabile peroxy-acid-aldehydes generates OH, improving model agreement with continental observations. OH and HPALD interactions are central to maintaining the OH levels in pristine and moderately polluted environments, which makes isoprene both a source and sink of OH ([Peeters2010](#); [Taraborrelli2012](#)). Isoprene chemistry in GEOS-Chem includes OH regeneration from oxidation of epoxydiols (not shown in Equation [2.6](#)) and slow isomerisation of ROO (Equation [2.6d](#)) ([Mao2013](#)). This regeneration dealt with the problem seen in older models where ISOPOOH production titrated OH, which was not backed up by measurements ([Paulot2009b](#); [Mao2013](#)). [Mao2013](#) showed that drastically lowering (by a factor of 50) the rate constant for ROO isomerisation lead to better organic nitrate agreements with measurements. These chemical updates led to more accurate modelling of OH concentrations, especially in low NO_X conditions common in remote forests. The updates to isoprene chemistry by

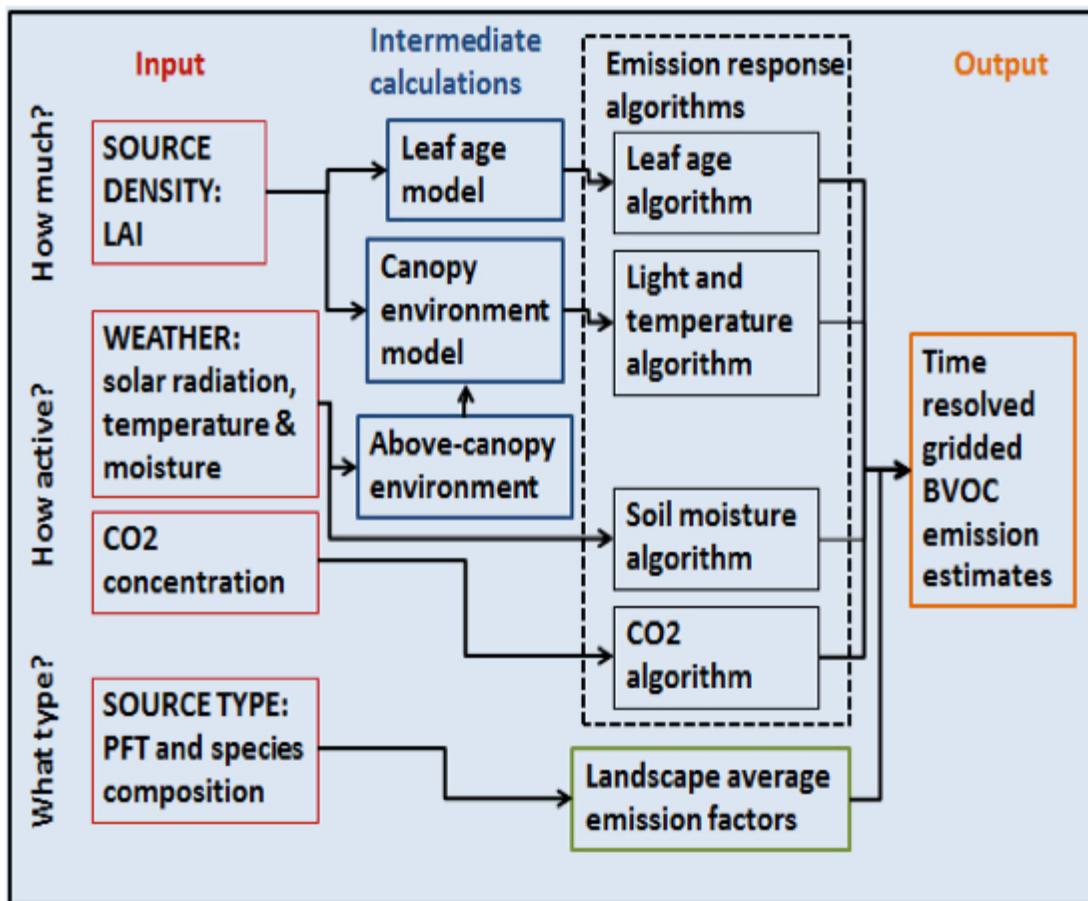


FIGURE 2.11: MEGAN schematic, from <http://lar.wsu.edu/megan/>

Mao2013, and those shown in Crounse2011; Crounse2012 are the last before GEOS-Chem version 11.

2.4.5 Emissions from MEGAN

GEOS-Chem runs the Model of Emissions of Gases and Aerosols from Nature (MEGAN) to determine biogenic emissions globally. GEOS-Chem V10.01 uses MEGAN V2.1 with biogenic emissions from Guenther2012. MEGAN is itself a global model with resolution of around 1 km, which uses globally estimated or measured (including remote sensed satellite data) inventories of parameters to estimate emissions. MEGAN uses leaf area index (LAI), global meteorological data, plant functional types (PFT), and photosynthetic photon flux density to simulate terrestrial isoprene emissions (Kefauver2014). The schematic for MEGAN, from <http://lar.wsu.edu/megan/>, is shown in Figure 2.11.

MEGAN was developed as a replacement for two earlier canopy-environment emission models (the biogenic emission inventory system, and the global emissions initiative), initially including a simple canopy radiative transfer model which parameterised sun-lit and shaded conditions through a canopy. Early models did not account for abiotic stresses, such as drought, prior rainfall and land use changes.

These stresses influence species-specific emissions by more than an order of magnitude (**Niinemets1999**). Isoprene emissions were based on temperature, leaf area, and light, but have since been updated to include leaf age activity (**Guenther2000**), and a leaf energy balance model (**Guenther2006**) in MEGANv2.0. This update included a parameter for soil moisture, to account for drought conditions; however, this parameter is currently (as of version 2.1) not applied to isoprene (**Sindelarova2014**). Instructions to run version 2.1 are available at http://lar.wsu.edu/megan/docs/MEGAN2.1_User_GuideWSU.pdf. Version 2.1 (updated from 2.0 (**Guenther2006**)) includes 147 species, in 19 BVOC classes, which can be lumped together to provide appropriate output for mechanisms in various chemical models. GEOS-Chem uses an embedded version of MEGAN 2.1.

GEOS-Chem computes some emissions using predefined emission factor (ϵ) maps from MEGAN source code, and others using PFT maps and associated emission factors. For isoprene (the focus in this thesis) MEGAN calculates emissions online, using local meteorological conditions. Emissions E of species i are calculated using these EF for classes of plant types j and associated grid-box coverage ξ and an activity factor γ which accounts for response to environmental conditions. The following equation is reproduced from **Guenther2012** showing how E_i are determined in MEGAN:

$$E_i = \gamma_i \sum_j \epsilon_{i,j} \chi_j \quad (2.7)$$

For example: isoprene emissions are tied to sunshine, temperature, plant emission strength (ϵ), emitting species coverage, etc., for each grid box.

2.4.6 Nitrogen oxides

NO_X concentrations affect atmospheric oxidative capacity, which changes many factors important in estimating isoprene emissions including isoprene to HCHO yield, isoprene lifetime, and isoprene oxidation pathways. In GEOS-Chem, NO_X concentrations are regulated by O_3 , VOC, HO_X , Bromine and aerosols reactions. NO_X emissions (from power generation and combustion transport) drive enhancements, while conversion to nitric acid (HNO_3) followed by deposition is the primary removal mechanism (**Delmas1997; Ayers2006**). Anthropogenic emissions (TODO total Tg N a^{-1}) are from the Emissions Database for Global Atmospheric Research (EDGAR), for more details visit http://wiki.seas.harvard.edu/geos-chem/index.php/EDGAR_v4.2_anthropogenic_emissions. Other NO_X emissions arise from sources including soil emissions ($\sim 10 \text{ Tg N a}^{-1}$), and lightning. Soil emissions are layed out in **Hudson2012**, parameterised using biome specific emission factors, and an explicit fertiliser dataset. Lightning based NO_X production is created based on inventories of lightning flash rates scaled within GEOS-Chem.

If GEOS-Chem is misrepresenting NO_X yields and the effects of transport may be incorrectly accounted for. In order to determine the accuracy of GEOS-Chem simulated NO_X over Australia, modelled NO_2 amounts are compared to satellite data

(where available) for 2005. Figure 2.12 shows tropospheric NO₂ columns from GEOS-Chem, OMNO2d, and their differences over four seasons in 2005. Simulated GEOS-Chem tropospheric NO₂ columns averaged from 13:00-14:00 local time (LT) are compared against OMNO2d data which are averaged into seasonal $2^\circ \times 2.5^\circ$ bins. Both data sets show Sydney and Melbourne as NO₂ hotspots; however, Sydney itself is underestimated by GEOS-Chem throughout the year, likely due to low or poorly aligned emission maps (shown in Figure 2.13). Over much of the country GEOS-Chem overestimates NO₂ by 10-60%, except in the northern areas where up to 50% underestimation occurs in Summer. An underestimation in southern NSW can also be seen in the later half of the year. Notably GEOS-Chem anthropogenic NO does not have any seasonality (Figure 2.13), while soil shows clear differences in emission hotspots between summer and winter.

Figure 2.14 shows scatter plots with one data point for each land square over Australia, for each season, coloured by total NO emissions. The correlation between model and satellite NO₂ columns is reasonable throughout the year over Australia, with a general positive bias throughout the year in modelled amounts. A comparison between the bias (GEOS-Chem - OMNO2d) with anthropogenic or soil emissions (columns 2 and 3 respectively) shows no real correlation, suggesting the bias is not driven by either anthropogenic or soil NO emissions in any season. Without a clear link between emissions and biases, alterations would be too subjective, and a change to NO_X chemistry is beyond the scope of this thesis. Additionally, high satellite NO₂ columns are filtered in later calculations which assume biogenic air masses as they suggest anthropogenic influence. The conclusion drawn is that modelled anthropogenic and soil NO emissions do not show sufficient evidence of biasing GEOS-Chem NO₂ columns away from satellite measurements over Australia. For this reason modelled NO emissions are not modified in model runs in this thesis.

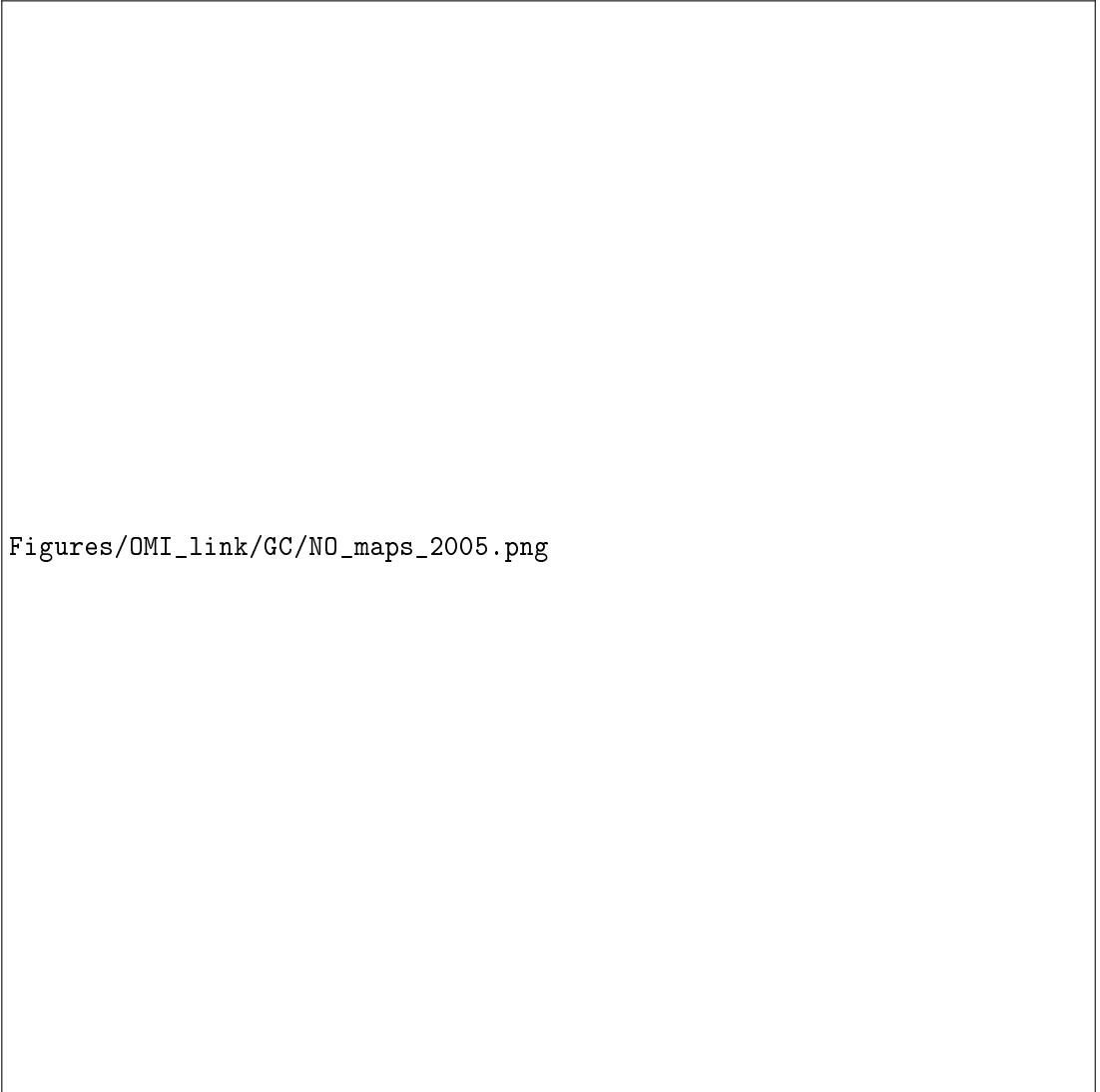
2.4.7 GEOS-Chem simulations

GEOS-Chem is run five times independently in this thesis, with different outputs from each simulation used to determine specific information. The different output types are first described in Section 2.4.7.1. Following this is the list of model runs, including a summary of the run, outputs created, and a summary of how they are used (Section 2.4.7.2). Finally a brief comparison between a subset of the runs is performed (Section 2.4.7.3).

2.4.7.1 GEOS-Chem outputs

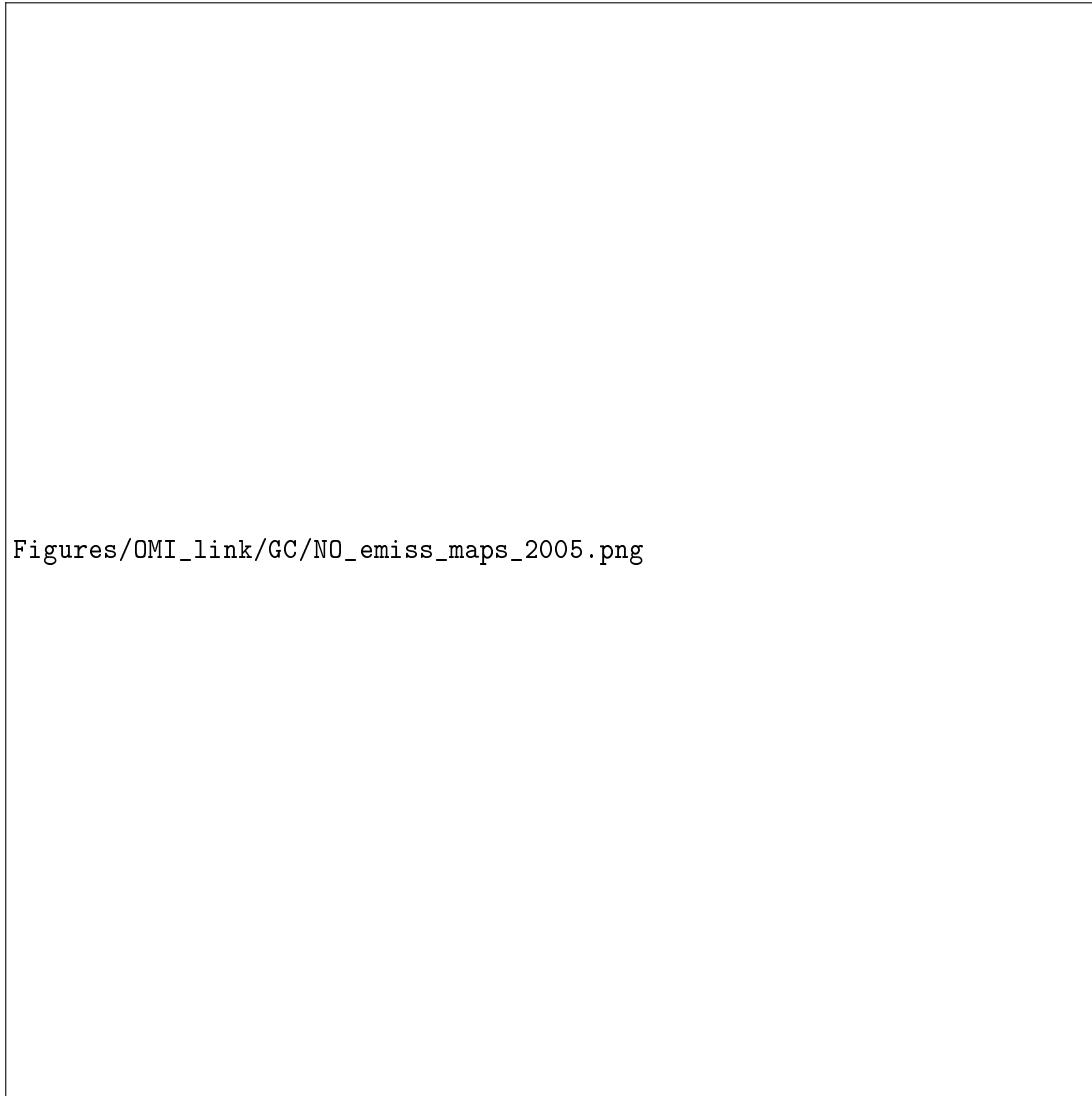
There are various outputs available when running GEOS-Chem, some of which require manipulation in order to compare against observations. GEOS-Chem in this thesis is run with a 15 minute time step for transport and 30 minutes for chemistry, at $2^\circ \times 2.5^\circ$ horizontal resolution over 47 vertical levels. Output is the average of these time steps. Optionally one can save high temporally resolved data for a single (or list of) column(s). This has been used here to compare modelled ozone with ozonesonde profiles at three sonde release sites discussed in Chapter 4.

Midday output is output from averaging over a window of local time for each grid-box. Output averaged between 13:00-14:00 local time is saved to match with



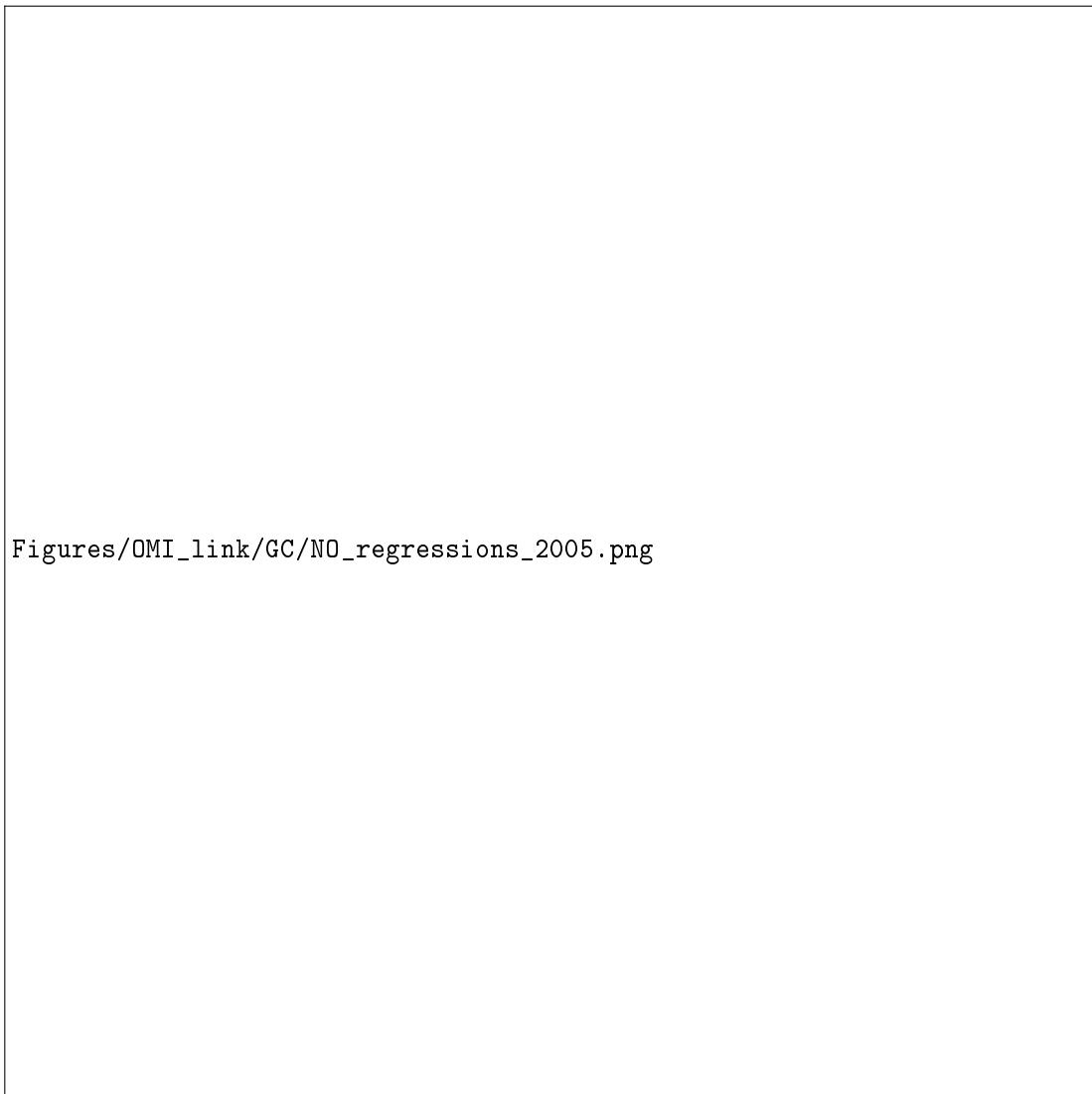
Figures/OMI_link/GC/NO_maps_2005.png

FIGURE 2.12: Left to right columns show tropospheric NO₂ columns (Ω_{NO_2} ; molec cm⁻²) from GEOS-Chem (GC), OMNO2d (OMI), and the differences. Each row shows one season from 2005, the left two columns use the left colour scale, while the third column uses the right colour scale.



Figures/OMI_link/GC/NO_emiss_maps_2005.png

FIGURE 2.13: Left and right columns show anthropogenic and soil emissions of NO respectively from GEOS-Chem in molec $\text{cm}^{-2} \text{ s}^{-1}$. Each row shows one season from 2005. Anthropogenic and soil emissions use a logarithmic and linear colour scale respectively.



Figures/OMI_link/GC/NO_regressions_2005.png

FIGURE 2.14: The first column shows the scatter plots of tropospheric NO₂ columns between GEOS-Chem (y-axis) and OMNO2d (x-axis) at $2^\circ \times 2.5^\circ$. The reduced major axis linear regression is drawn in red, and the equation, regression coefficient (r), and number of grid squares (n) used is inlaid as a legend. The second and third columns show the scatter between emissions and the bias between GEOS-Chem and OMNO2d (GC-OMI), for anthropogenic and soil emissions respectively. These two columns share the far right axis; however, emissions are from anthropogenic and soil sources respectively. All scatter points are coloured by the sum of anthropogenic and soil NO emissions (from GEOS-chem), as per the colour bar shown at the bottom.

Aura satellite measurements, as Aura overpasses at \sim 1330 local time each day. Midday output is saved both for comparison with, and recalculation of, satellite measurements.

HEMCO diagnostics :the Harvard-NASA Emissions Component (HEMCO) deals with emissions inventories used in GEOS-Chem. When working with globally gridded data, handling local time offsets becomes more important. The hourly averaged emissions of isoprene saved using GMT which needs to be offset based on longitude in order to retrieve local time. This is done by setting up a latitude by longitude array which matches the horizontal resolution of the data, filling each gridbox with its local time offset. This offset is determined as one hour per 15° (as 360° is 24 hours), and then used to retrieve global data at any specific local time. The retrieval of a daily local time global array is done by index matching the GMT+local time (modulo 24) with the desired hour on this grid over the 24 GMT hours.

Tracer averages are daily or monthly averaged gridbox concentrations.

Time series are vertical columns with diagnostics saved at a temporal resolution of up to 15 minutes.

2.4.7.2 GEOS-Chem runs

The following list summarises each model run as well as enumerating the outputs (described above), and how the run is used in the thesis.

UCX The model was run using the Universal tropospheric-stratospheric Chemistry eXtension (UCX) mechanism with 72 vertical levels from the surface to the top of the atmosphere (TOA \sim 0.1 hPa). UCX runs a chemistry mechanism with combined tropospheric and stratospheric reactions, with an increased number of stratospheric calculations performed online ([Eastham2014](#)).

1. Outputs: Midday output, daily tracer averages, and time series over three stations.
2. The Midday output is used to check how shape factors for AMF recalculation are affected by vertical resolution and stratospheric chemistry. TODO: Make a plot of AMF with and without 72 vertical levels
3. This run shows what influence the stratospheric chemistry additions have over tropospheric isoprene or HCHO concentrations.
4. The daily tracer averages are used to determine ozone intrusion quantification (Section [4.6.1](#)), and ozone concentration seasonality (Section [4.5](#)).
5. Time series outputs are compared against ozonesonde releases (Section [4.5](#)) both over time and vertically.

Tropchem (standard) default settings for GEOS-Chem 10.01, using 47 vertical levels at $2^\circ \times 2.5^\circ$ horizontal resolution. Additional midday output is created to allow AMF recalculation code to run on OMI satellite measurements (Section [2.6.3.2](#)).

1. Outputs: Midday output, daily tracer averages, and HEMCO diagnostics

2. These are used in recalculation of the satellite AMF (Section 2.5), and the modelled background HCHO over the remote Pacific which is used in the reference sector correction for OMI column retrievals (Section 2.6.5). Additional diagnostic outputs are used hard coded to allow AMF recalculation.
3. Midday output is combined from two different runs in order to determine smearing (Section 3.2.7)
4. Total yearly ozone concentrations are compared before and after scaling isoprene emissions using the top-down estimate.

Tropchem (isoprene emissions halved) identical to standard tropchem except isoprene emissions are halved.

1. Outputs: Midday output, and monthly tracer averages
2. Check modelled ozone sensitivity to isoprene emissions TODO: plot ozone between the two runs and reference here
3. Combined with standard run to determine model sensitivity transport (Section 3.2.7)

Tropchem (biogenic emissions only) identical to standard tropchem except all non-biogenic emissions inventories are disabled.

1. Outputs: Midday output, and hourly biogenic emissions from MEGAN
2. Midday and HEMCO outputs are used to determine isoprene to HCHO yield, after removing days with high smearing (Section 3.2.4)
3. HEMCO diagnostics are compared against top-down estimations of isoprene emissions (Section 3.3.1)

Tropchem (altered MEGAN scaling factor) Identical to standard tropchem except isoprene emissions are scaled to match multi-year monthly averaged top-down estimates.

1. Midday output, time series, daily averaged tracers
2. Compare to campaign datasets after altering isoprene emissions (see Chapter 3)
3. Compare against satellite column HCHO as a sanity check on improving isoprene emissions

2.4.7.3 UCX vs tropchem

Here we compare the model output with and without enabling the Universal tropospheric-stratospheric Chemistry eXtension (UCX). Both runs use $2^\circ \times 2.5^\circ$ latitude by longitude; however, the UCX mechanism is run with 72 vertical levels from the surface to the top of the atmosphere (TOA ~ 0.1 hPa), while the standard (tropchem) run uses 47 levels. The reduced vertical levels are created by lumping together sets of the higher resolved levels from around 70 hPa to the top of the atmosphere. For both runs the input parameters such as MEGAN emissions and GEOS-5 meteorological fields are identical.

Figure 2.15 shows an example of surface HCHO amounts, averaged over Jan and Feb, 2007, with and without (A, B respectively) the UCX mechanism enabled. Surface HCHO (first model level; up to ~ 100 m) does not differ much between runs. The differences do not exceed 3% over Australia, and absolute differences are minor (note the scale in A-B). The major notable difference occurs over northern Africa, where HCHO is around 20% lower in the UCX run. Additionally a slight ($< 8\%$) decrease in HCHO over the oceans can be seen. The comparison is repeated using total columns (instead of surface values) in Figure 2.16, showing that differences affecting HCHO between the model run are spread over the entire vertical column.

Figure 2.17 shows the differences in surface isoprene concentrations over Australia, averaged over 1, Jan to 28, Feb, 2007. Here we start to see a higher relative difference in concentrations, although this is generally over the areas with less absolute concentrations. Very little isoprene is seen away from the continents (4-5 orders of magnitude less), due to its short lifetime and lack of oceanic sources. Generally isoprene is 0-30% higher over mid to western Australia when the UCX mechanism is turned on; however, this increase is lower in the regions with high isoprene emissions (north-east to south-east coastline). This enhancement can be seen throughout the entire tropospheric column as shown by Figure 2.17. There is a greater effect in Africa and South America in the tropics, with high relative differences in many regions with low absolute amounts.

The difference in isoprene between UCX and tropchem is likely caused by differences in the modelled radiation reaching the troposphere due to differences in simulated ozone in the stratosphere. Figure 2.19 shows the total column ozone between UCX and non-UCX run. This shows that UCX ozone is lower everywhere except for a thin band just north and south of the equator. With higher stratospheric ozone levels, less radiation would reach the troposphere. This would slow photolysis limited reactions, such as the splitting of Ozone which leads to OH production in the troposphere, in turn slowing the isoprene loss to OH reactions.

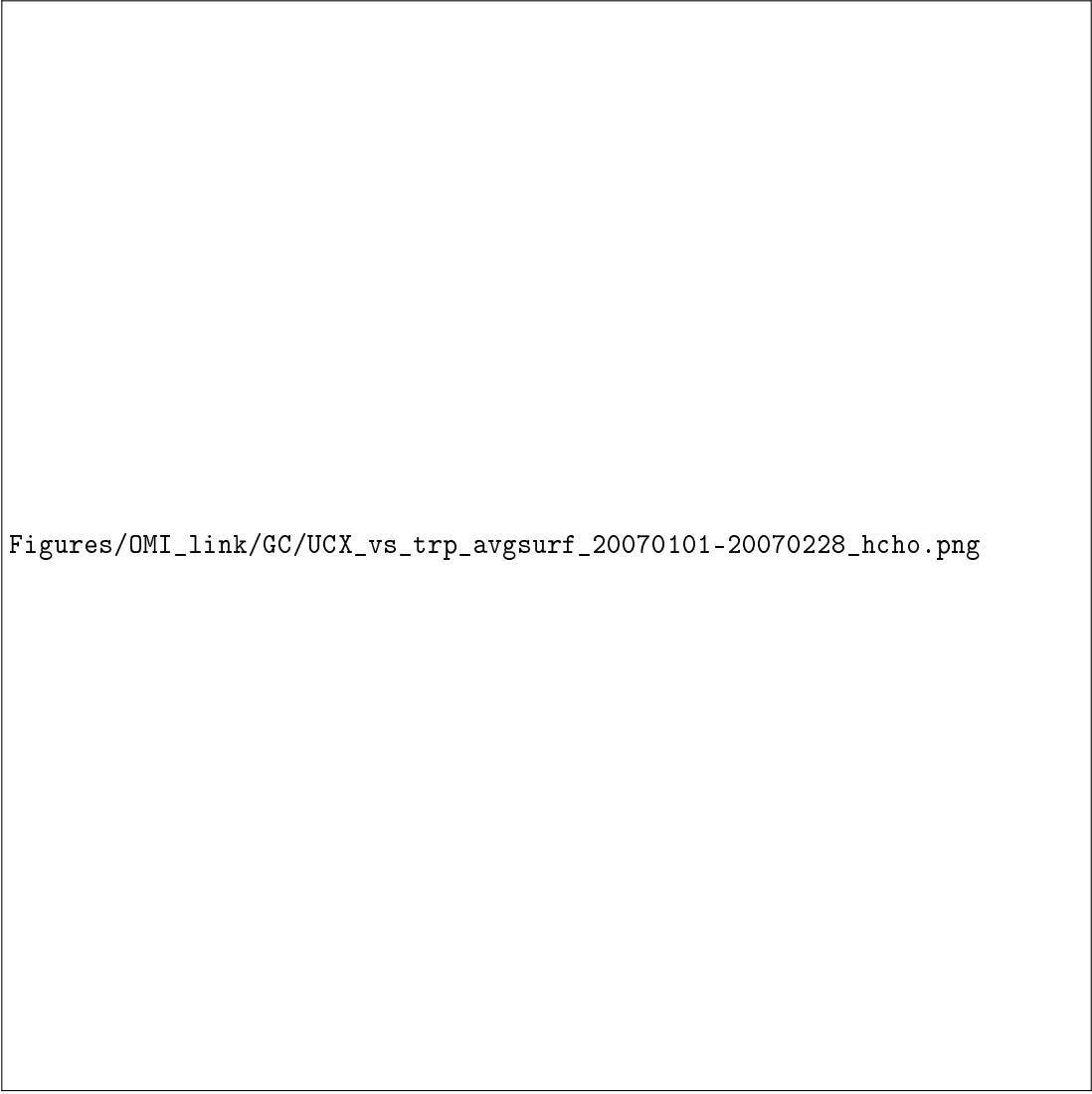
Overall the UCX mechanism and increased vertical resolution lead to slightly higher isoprene and HCHO, and slightly lower ozone over Australia. These differences are on the order of 0-10% and as shown later (see Section 3.4) are minimal compared to other uncertainties in both AMF calculation and emissions estimation. Running UCX requires roughly twice the computation (and real) hours, and for the majority of this thesis the normal tropchem runs are used.

2.5 Calculating new AMF

One of the uses of GEOS-Chem in this thesis is to update the a priori HCHO column used in satellite measurements. The a priori is used when deriving the air mass factor (AMF), which is needed to transform satellite slant columns (SC) into vertical columns (Ω in molecules cm^{-2}): The AMF is the ratio of SC to the vertical column

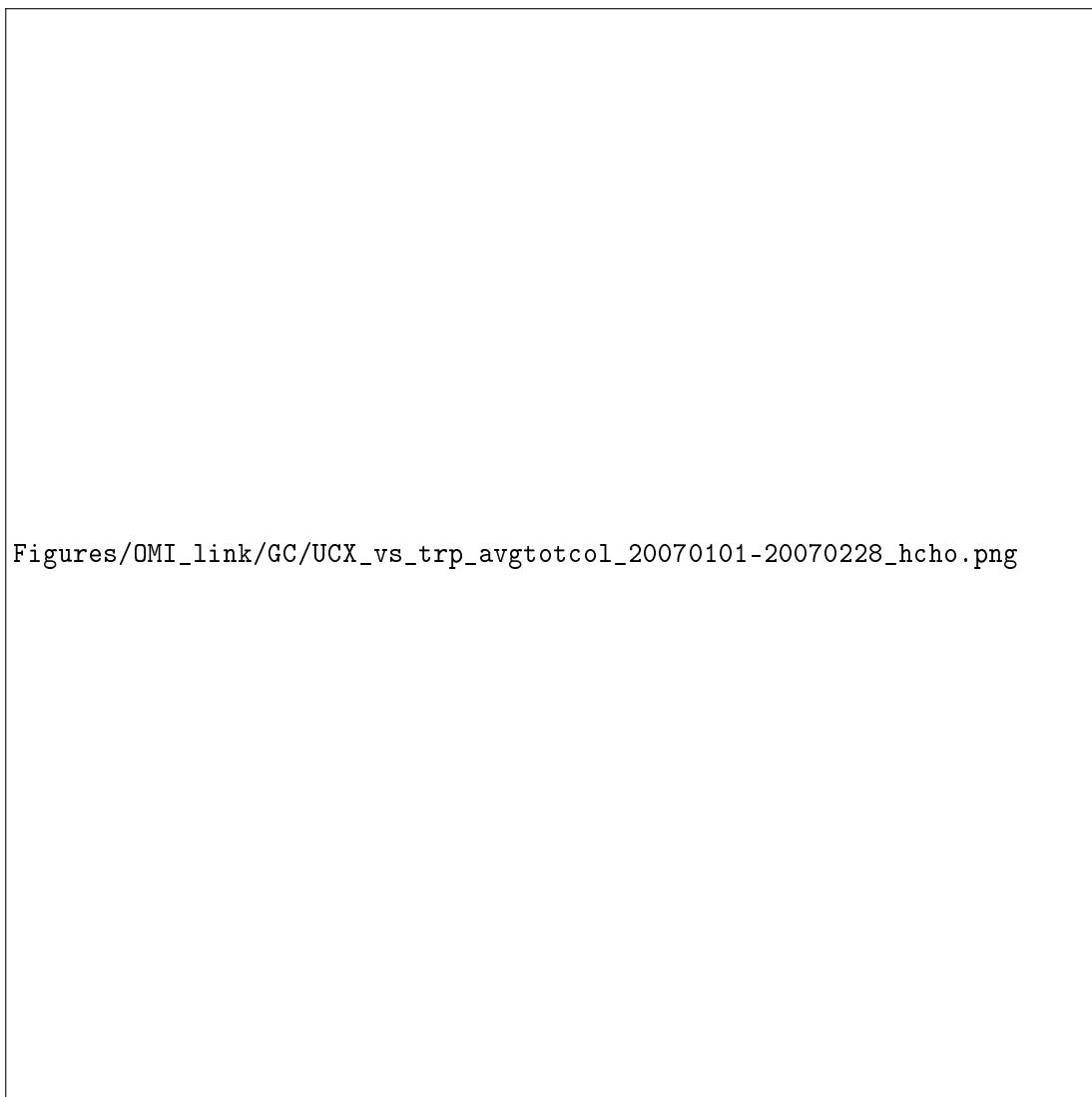
$$\text{AMF} = \frac{\text{SC}}{\Omega} = \frac{\tau_s}{\tau_v} \quad (2.8)$$

with τ being the optical depth or thickness of the absorber through the slant (s) or vertical (v) path of light.



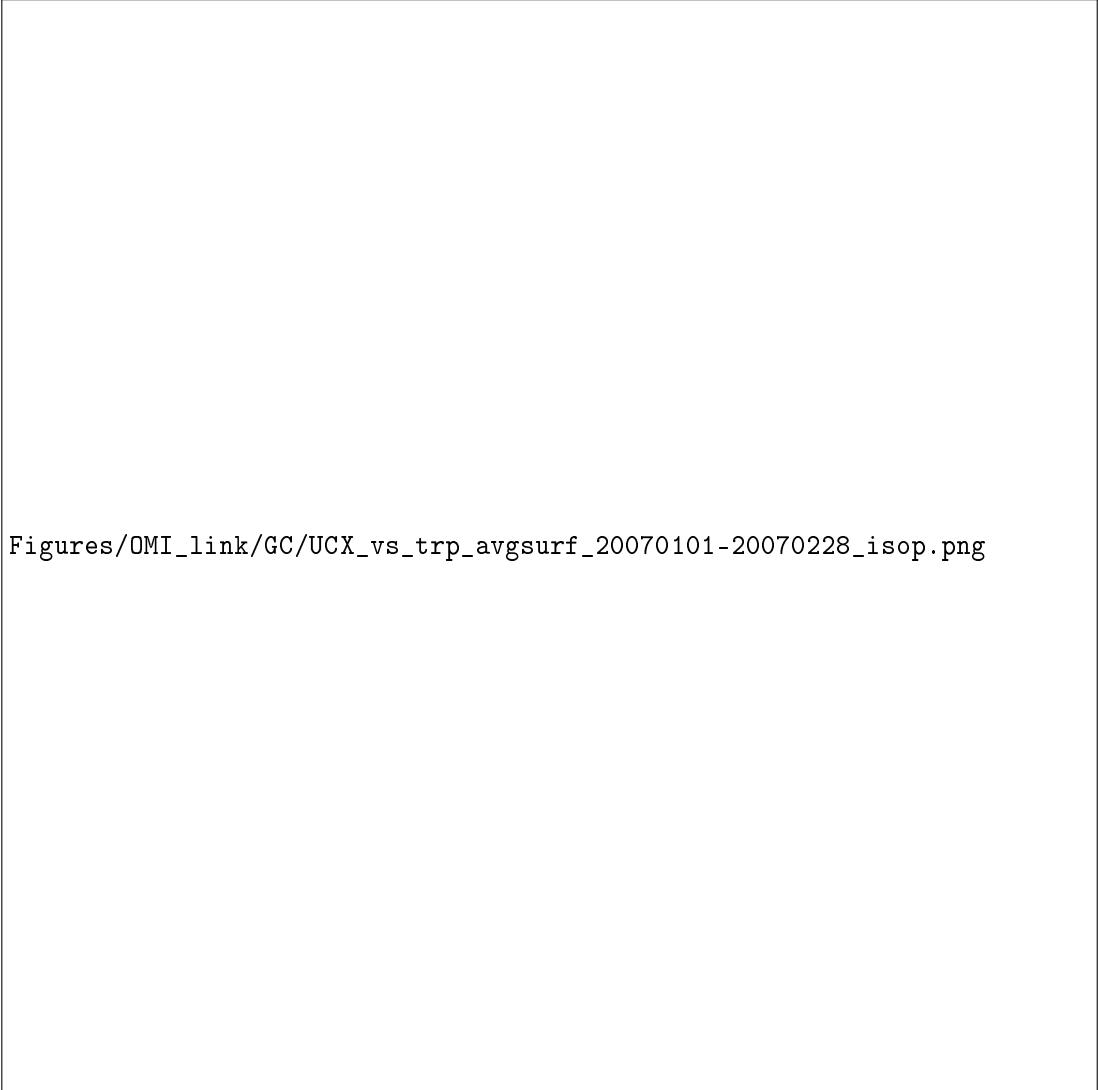
Figures/OMI_link/GC/UCX_vs_trp_avgsurf_20070101-20070228_hcho.png

FIGURE 2.15: Surface HCHO simulated by GEOS-Chem with UCX (top left), and without UCX (top right), along with their absolute and relative differences (bottom left, right respectively). Amounts are the average of all times between 1 Jan and 28 Feb 2007.



Figures/OMI_link/GC/UCX_vs_trp_avgtotcol_20070101-20070228_hcho.png

FIGURE 2.16: As Figure 2.15 using total column amounts instead of surface concentrations.



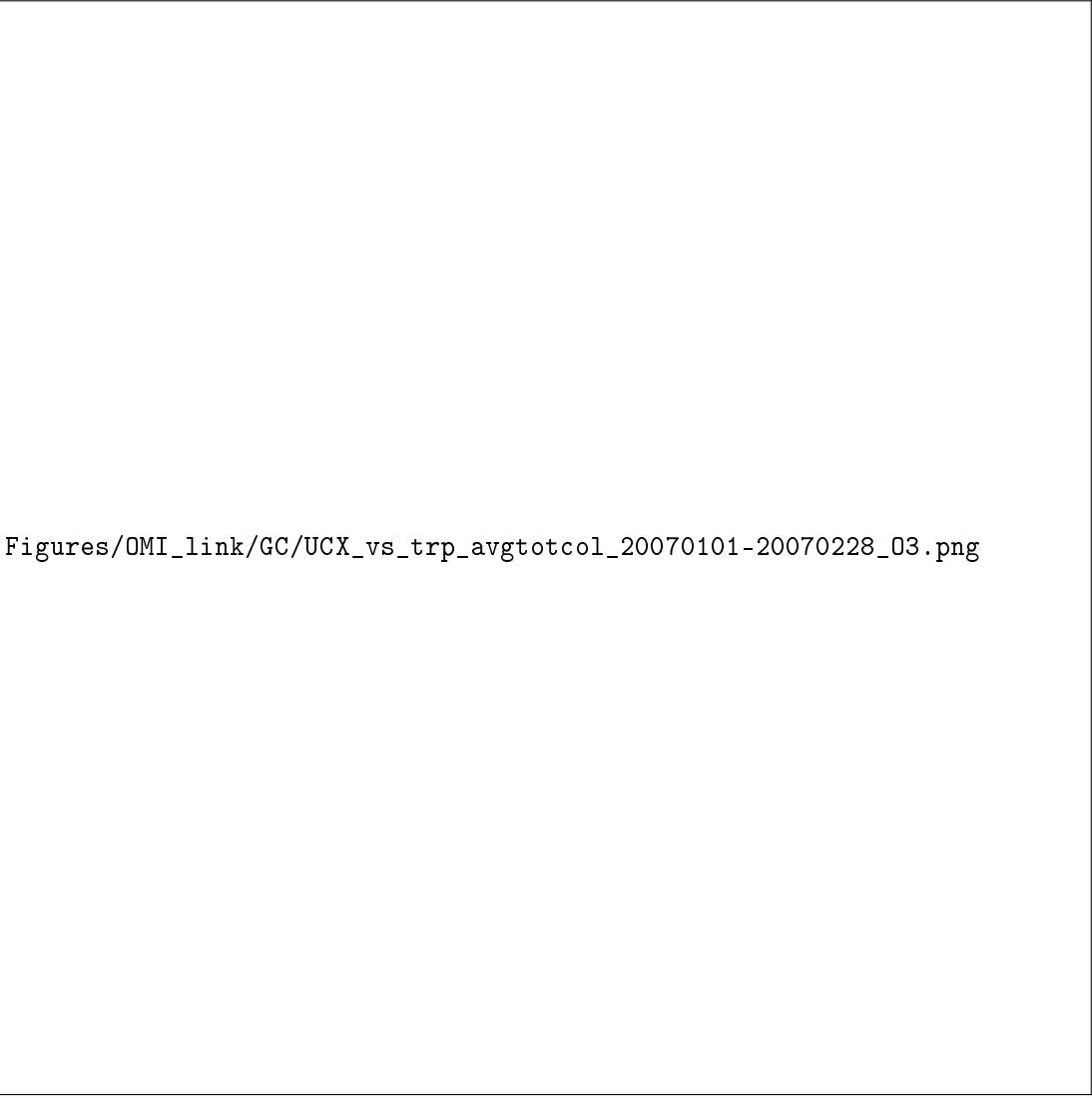
Figures/OMI_link/GC/UCX_vs_trp_avgsurf_20070101-20070228_isop.png

FIGURE 2.17: As Figure 2.15, except showing isoprene surface concentrations.



Figures/OMI_link/GC/UCX_vs_trp_avgtotcol_20070101-20070228_isop.png

FIGURE 2.18: As Figure 2.17, except showing isoprene total column amounts.



Figures/OMI_link/GC/UCX_vs_trp_avgtotcol_20070101-20070228_03.png

FIGURE 2.19: Total column ozone simulated by GEOS-Chem with UCX (top left), and without UCX (top right), along with their absolute and relative differences (bottom left, right respectively). Amounts are the average of all times between 1 Jan and 28 Feb 2007.

The OMI instrument records spectra of light which enters the viewing lens on the Aura satellite. The spectra provide backscattered intensity (I_B) at various wavelengths (see Section 2.3), with the light source (I_{B_0}) being the sun. Using the log of Beers law (Equation 2.2) we get

$$\tau_s = \ln I_{B_0} - \ln I_B$$

which can be subbed into Equation 2.8 to give an expression for the AMF which includes scattering:

$$AMF = \frac{\ln I_{B_0} - \ln I_B}{\tau_v} \quad (2.9)$$

We use $\nabla I = I_B - I_{B_0}$ to represent the change in intensity due to the absorber. For optically thin absorption, $\nabla I/I_B \ll 1$, and we can use:

$$AMF = \frac{\ln \left(1 - \frac{\nabla I}{I_B}\right)}{\tau_v} \approx \frac{-\frac{\nabla I}{I_B}}{\tau_v} \quad (2.10)$$

This is due to the logarithmic property $\ln(1-x) \approx -x$ for $x \ll 1$. ∇I can also be expressed as the integral of the absorption slices over optical depth increments:

$$\nabla I = \int_0^{\tau_v} \frac{\partial I_B}{\partial \tau} d\tau \frac{\nabla I}{I_B} = \int_0^{\tau_v} \frac{\partial \ln I_B}{\partial \tau} d\tau$$

which can be placed into Equation 2.10 leading to

$$AMF \approx \frac{-1}{\tau_v} \int_0^{\tau_v} \frac{\partial \ln I_B}{\partial \tau} d\tau$$

We can then convert $d\tau$ to our path using Equation 2.4 leading to

$$AMF = \frac{-1}{\tau_v} \int_0^{\infty} \frac{\partial \ln I_B}{\partial \tau} \alpha(z) \eta(z) dz \quad (2.11)$$

where $\alpha(z)$ and $\eta(z)$ represent absorption cross section in $\text{m}^2 \text{ molecule}^{-1}$, and number density in molecules m^{-3} respectively. This uses the attenuation cross section relationship to optical depth (see Section 2.3.2).

To represent an average cross section weighted by the absorbing species' vertical distribution, the effective cross section (α_e) is used. This is to account for temperature and pressure dependence of $\alpha(z)$, and is defined as:

$$\begin{aligned} \alpha_e &= \frac{1}{\Omega_v} \int_0^{\infty} \alpha(z) \eta(z) dz \\ &= \frac{\tau_v}{\Omega_v} \end{aligned}$$

Then replacing the τ_v in Equation 2.11 we obtain:

$$AMF = - \int_0^{\infty} \frac{\partial \ln I_B}{\partial \tau} \frac{\alpha(z)}{\alpha_e} \frac{\eta(z)}{\Omega_v} dz \quad (2.12)$$

Often the integrand of this AMF formula (Equation 2.12) is broken apart into two defining terms: the scattering weights $\omega(z)$ and the shape factor $S(z)$, described here:

- ω The scattering weights describing sensitivity of the backscattered spectrum to the abundance of an absorber at altitude z :

$$\omega(z) = -\frac{1}{AMF_G} \frac{\alpha(z)}{\alpha_e} \frac{\partial \ln I_B}{\partial \tau} \quad (2.13)$$

It is worth noting that in the OMI satellite product, the provided $\omega(z)$ term does not include the $\frac{1}{AMF_G}$ term and the calculations which follow therefor do not include this term when utilising the provided ω . This is not noted in any of the papers which recalculate the AMF from the OMI product, due to them recalculating the ω term themselves with a radiative transfer model such as LIDORT.

- S the shape factor describes the profile of an absorber ($\eta(z)$) normalised by its total vertical column amount (Ω_v):

$$S(z) = \frac{\eta(z)}{\Omega_v} \quad (2.14)$$

Plugging Equations 2.13 and 2.14 into Equation 2.12 gives us:

$$AMF = AMF_G \int_0^\infty \omega(z) S(z) dz \quad (2.15)$$

Since we are using the ω provided by OMI, the AMF_G term is removed from this calculation as it is not part Equation 2.13 leading to

$$AMF = \int_0^\infty \omega(z) S(z) dz \quad (2.16)$$

Additionally the AMF can be determined using the sigma (σ) coordinate system, which denotes altitude as going from 1 (ground level) to 0 (TOA) (see Section 2.2). A conversion to the σ vertical coordinate is performed using $p = \sigma(p_S - p_T) + p_T$, where p_T is pressure at the top of the atmosphere and p_S is surface pressure. S_σ is a dimensionless normalised shape factor on the σ coordinate system. In the sigma coordinate system we calculated the shape factor as defined in Palmer2001:

$$S_\sigma(\sigma) = \frac{\Omega_a}{\Omega_v} C_{HCHO}(\sigma) \quad (2.17)$$

where Ω_a is the vertical column of air from the surface to the top of the atmosphere and $C_{HCHO}(\sigma)$ is the mixing ratio of HCHO. We can integrate over the sigma coordinates using the hydrostatic relation $p = -\rho_a g z$, with ρ_a being air density, and g the gravitational acceleration:

$$\begin{aligned} \rho_a g z &= \sigma(p_S - p_T) + p_T \\ d\sigma &= -\frac{\rho_a g}{p_S - p_T} dz \end{aligned}$$

Substitution into 2.16 gives AMF using the sigma coordinates:

$$AMF = \int_0^1 w(\sigma) S_\sigma(\sigma) d\sigma \quad (2.18)$$

2.6 Recalculation of OMI HCHO

OMI HCHO vertical columns are calculated using modelled a priori HCHO profiles (see Section 2.3). When comparing satellite measurements against models it is important to recognise the impact of this a priori on the total column values. This is complicated by how OMI is differently sensitive to HCHO (and other trace gases) vertically throughout the atmosphere. When comparing OMI vertical columns (Ω_{OMI}) to GEOS-Chem (Ω_{GC}), the satellite AMF needs to be recalculated using GEOS-Chem modelled vertical gas profiles as the a prioris. Without performing this step a bias between modelled and measured total column values may be due to the a priori rather than chemistry or measurements (Palmer2001; Lamsal2014).

Here, two new AMFs are calculated, both using GEOS-Chem HCHO profiles as the new a priori. The first (AMF_{GC}) uses the original satellite scattering weights while recalculating the shape factor, whereas the second (AMF_{PP}) also recalculates scattering weights. AMF_{PP} is created using code initially written by Professor P. Palmer. A reference sector correction is determined using the method described in Abad2016. This correction has also been applied to the original vertical column in the OMHCHO product, so the column is available with and without this correction. The reference sector correction is unique for each of the 60 measurement tracks used by OMI. The correction is calculated for each of the two new vertical columns, and applied to each pixel to create the corrected vertical columns. The end product is three sets of corrected vertical columns: the original, one using GEOS-Chem to recalculate the shape factors, and one from Palmer's code using GEOS-Chem to both recalculate shape factors and scattering weights.

2.6.1 Outline

An outline in computational order of what takes place when recalculating the Ω from OMI follows:

1. Section 2.6.2 describes how GEOS-Chem satellite overpass output is used to create new shape factors (S_z and S_σ).
 - (a) Vertical pressure edges and geometric midpoints are determined, along with altitudes (z), and box heights (H).
 - (b) Number density and mixing ratio of HCHO (n_{HCHO} , C_{HCHO} respectively) are taken or created from model outputs HCHO (ppb), air density (molec cm^{-3}), and box heights (m).
 - (c) Total column HCHO from GEOS-Chem (Ω_{GC}) is calculated

$$\Omega_{GC} = \sum_z (n_{HCHO}(z) \times H(z))$$

along with total column air (Ω_A , calculated similarly).

- (d) The shape factor $S_z(z)$ is calculated for each altitude

$$S_z(z) = n_{HCHO}/\Omega_{HCHO}$$

- (e) Sigma coordinates are created from pressures (see Section 2.2, Equation 2.1).
(f) $S_\sigma(z)$ is calculated on each altitude:

$$S_\sigma(z) = C_{HCHO}(z) \times \frac{\Omega_A}{\Omega_{HCHO}}$$

2. For each pixel, new AMF (AMF_{GC} and AMF_{PP}) is created using the GEOS-Chem shape factors and satellite scattering weights in Section 2.6.3:

- (a) Satellite pixels (SC, scattering weights ($\omega(z)$), pressure levels, latitude and longitude) are read from the OMHCHO dataset.
- (b) scattering weights (ω) are interpolated onto the same vertical dimensions (z and σ) as the shape factors.
- (c) Integration (approximated using rectangular method) is performed along the vertical dimension to calculate the new AMF on both coordinate systems:

$$\text{AMF}_z = \Sigma_z (\omega(z) \times S_z(z) \times H(z)) \quad (2.19)$$

$$\text{AMF}_\sigma = \Sigma_\sigma (\omega(\sigma) \times S_\sigma(\sigma) \times d\sigma) \quad (2.20)$$

These two AMFs represent the same thing using different coordinates, and either one can be used as the AMF_{GC} .

- (d) Code from Dr. Surl and Professor Palmer is used to recalculate both the shape factor and scattering weights to produce AMF_{PP} .
- 3. Section 2.6.4 describes how the original AMF, AMF_{PP} , and AMF_{GC} are used to determine the new vertical columns: $\Omega = SC/AMF$.
- 4. A reference sector correction (RSC) is defined each day using these AMFs along with modelled HCHO over the remote Pacific in Section 2.6.5:
 - (a) GEOS-Chem midday output ($\Omega_{GEOS-Chem}$ from 140°W to 160°W) are averaged monthly and longitudinally to provide modelled vertical columns over the reference sector.
 - (b) These vertical columns ($\Omega_{GEOS-Chem}$) are turned into modelled slant columns by multiplication with the AMF over the reference sector:

$$SC_{GEOS-Chem} = \Omega_{GEOS-Chem} \times AMF$$

- (c) For each satellite pixel in the reference sector, a correction ($corr$) is calculated as the measured SC minus the modelled slant column at the nearest latitude:

$$corr[lat, track] = SC[lat, track] - VC_{GEOS-Chem}[lat] \times AMF$$

- (d) These corrections are binned by satellite detector (track: 1-60), and latitude (0.36° ; 500 latitudes from 90°S to 90°N).
 - (e) The median entry of each bin is determined and this forms the $\text{RSC}[\text{lat}, \text{track}]$ (e.g., Figure 2.21).
5. VCC are determined using $\text{VCC} = (\text{SC} - \text{RSC}[\text{lat}, \text{track}]) / \text{AMF}$ for each measured SC and using each AMF, with the RSC linearly interpolated to the latitude of the satellite pixel (see Section 2.6.6).
 6. The VCC (along with most of the pixel and GEOS-Chem data) are binned onto a $0.25^\circ \times 0.3125^\circ$ grid along with other information (see Section 2.6.7).

2.6.2 Creating new shape factors

To visualise and analyse OMI HCHO columns, slant columns are transformed into vertical columns using the AMF. The shape factor (S) is one of the key components in creation of the AMF (see Section 2.5, Equation 2.16). The shape factor is calculated using GEOS-Chem midday output (see Section 2.4.7.1) which provide simulated HCHO concentration profiles ($\eta(z)$) and total columns (Ω) at $2^\circ \times 2.5^\circ$ horizontal resolution. Using Equation 2.14 to determine the shape factor is straightforward $S(z) = \frac{\eta(z)}{\Omega}$. The associated OMI per-pixel scattering weights are not changed in this calculation (unlike in Section 2.6.3.2).

Model output is provided as mixing ratios (C in ppbv), and is converted before being used in the shape factor calculation. The following equation converts model profile output from ppbv into number density (η) in molecules cm^{-3} :

$$\eta_{\text{HCHO}} = C_{\text{HCHO}} \times \eta_a \times 10^{-9} \quad (2.21)$$

where η_{HCHO} is the number density of HCHO, η_a is the number density of air (from model output), and C_{HCHO} is the mixing ratio of HCHO. The modelled total vertical column Ω_{HCHO} is determined by:

$$\Omega_{\text{HCHO}} = \sum_z (\eta_{\text{HCHO}} \times H(z))$$

where $H(z)$ is the box height for altitude z . In effect this equation sums over the molecules per cm^2 in each altitude level.

As a sanity check S_σ is calculated (through Equation 2.17) to confirm that these shape factors are equivalent. Comparing the resulting AMFs created by Equations 2.18 and 2.16 for each pixel provides confidence in the unit conversions (and other factors) applied. These AMF values are nearly identical in practice (correlation coefficient > 0.99).

2.6.3 Creating new AMF using GEOS-Chem

2.6.3.1 Updating shape factors

From Equation 2.16 we have:

$$AMF = \int_0^{\infty} \omega(z)S(z)dz$$

Using the $\omega(z)$ from satellite data, along with our calculated S_z interpolated linearly onto the same vertical grid as $\omega(z)$, the AMF can be determined through integration. The integration is performed using a simple rectangular method, which multiplies the integrand midpoints by the change in height, and then takes the sum for each vertical box. This assumes that the provided scattering weights and shape factors are linear between the 47 resolved values.

2.6.3.2 Recalculating the AMF using PP code

The major limitation of vertical columns which implement AMF_{GC} is that the scattering weights also depend on the model atmosphere, which is not updated when only the shape factor is recalculated. This is addressed using Fortran code written by Paul Palmer and Randal Martin, subsequently updated by Luke Surl. This Fortran code (PP code) is computationally expensive, and is run on a subset of the globe ($50\text{--}10^\circ S$, $160^\circ W\text{--}160^\circ E$) covering Australia and the Pacific ocean. This allows both vertical column recalculations and reference sector correction.. The instrument sensitivity (or scattering weights; ω) and shape factors for each pixel are calculated within the PP code.

The code uses a combination of GEOS-Chem a priori profile information and satellite measurement data to calculate the AMF_{PP} by using LIDORT radiative transfer calculations to determine scattering.

Code for recalculating AMF using satellite swaths and modelled aerosol optical depths and gas profiles can be found at http://fizz.phys.dal.ca/~atmos/martin/?page_id=129. The coded method is detailed in **Palmer2001**, with modifications for clouds and use of the LIDORT radiative transfer model (**Spurr2002**) as described by **Martin2003**. Modifications performed by Luke Surl at University of Edinburgh enabled the PP code to utilise OMI satellite data. Additionally, required tropopause heights averaged within satellite overpass times are created by modifying the GEOS-Chem diagnostic output. The calculation of $\omega(z)$ is simplified by using large “look up tables” of values based on many parameters such as cloud top heights and optical depths. These AMF look up tables can be found in the source code at <https://github.com/LukeSurl/amf581g>. The PP code uses HCHO concentration profiles averaged between 1300 and 1400 LT, including optical depths at 550 nm, and dust concentrations from GEOS-Chem, along with a subset of the OMI pixel information. The required information is taken from OMHCHO daily swath files, and sorted into csv files (one per day) which are then read by the PP code in conjunction with the GEOS-Chem outputs for each day. The PP code then produces a list of recalculated AMF which are read by python code and associated with the corresponding satellite pixel (outlined in ??).

2.6.3.3 Saving the AMF with satellite pixels

OMI satellite pixels are read and filtered for quality (see Section 2.3.1) and have their AMFs calculated as shown above. The process is outlined in figure ?? for a single day. This filtering removes highly uncertain pixels, including row anomaly affected pixels. Additional information is added to each pixel, including the new AMFs. Each pixel and its relevant data are saved in a long list for subsequent processing.

2.6.4 Vertical columns from AMF

All that remains for recalculating the total vertical column using our new a priori shape factor is to apply the new AMF to the slant columns and grid them onto our chosen resolution. Each satellite pixel at this stage has an associated SC along with three AMFs: the original (AMF_{OMI}), one with recalculated shape factors (AMF_{GC}), and one completely recalculated using PP code (AMF_{PP}). These are used to create vertical columns (Ω) through Equation 2.8: $\Omega = SC / AMF$.

Figure TODO shows a comparison between the three satellite Ω over Australia for 2005. The first row shows the yearly average column amounts, as well as coloured rectangles describing regions which are averaged over time in row 3. Row two from left to right shows the relative differences between Ω calculated using AMF_{PP} and AMF_{GC} , AMF_{OMI} and AMF_{GC} , and AMF_{OMI} and AMF_{PP} . TODO Analysis

TODO: Add this plot here.

2.6.5 Reference sector correction

On a medium to long time scale, OMI degradation needs to be accounted for. To remove effects from the deterioration of the satellite instrument, measurements over the remote pacific (our reference sector) are combined with GEOS-Chem simulations to create a reference sector correction (RSC). In this thesis we use OMI HCHO columns to estimate isoprene emissions, and correcting background HCHO does not affect this calculation. The correction should reduce bias caused by satellite degradation, without impacting isoprene emissions estimation. The RSC corrects for several problems; however, it introduces some a priori model influence. One corrected problem is the potential influence of varying dead/hot pixel masks across the OMI 2-D detector array (DeSmedt2015). This method also corrects for the errors introduced through correlations between BrO and HCHO absorption cross sections, which are especially significant at high latitudes (Abad2015).

HCHO products from OMI and SCIAMACHY both use a median daily remote Pacific ocean radiance reference spectrum, over 15°S-15°N, 140°-160°W where it is assumed that the only significant source of HCHO is methane oxidation (DeSmedt2008; Barkley2013; Kurosu2014). Here a new reference sector correction is created using modelled and measured HCHO columns over the remote Pacific to produce corrected vertical columns (VCC). This follows Abad2016, and defines the remote Pacific as the band between 140°W to 160°W. Each satellite slant column measurement is corrected by how much the satellite reference sector measurements at that latitude diverge from modelled amounts over the reference sector.

Modelled slant columns over the reference sector ($SC_{GEOS-Chem}$) are calculated by multiplying modelled vertical columns with the AMF calculated in prior sections using Equation 2.8:

$$SC_{GEOS-Chem} = \Omega_{GEOS-Chem} \times AMF$$

The longitudinal average is taken within the remote Pacific, as corrections are (assumed to be) longitudinally invariant. These modelled slant columns are averaged over the month and interpolated latitudinally to 500 equidistant bins. Figure 2.20 shows the simulated reference sector vertical columns as an example, calculated for January 1st 2005. In this figure the latitudinal resolution is increased from 2° to 0.36° , through linear interpolation, in order to form 500 vertical bins which are used in correcting the satellite data. This resolution is chosen to match that of **Abad2016**.

For OMI swaths, each row of measured data contains 60 “Across track” (track) measurements. The track index i relates the measurement to one of the 60 columns of data. Corrections (molecules cm^{-2}) for each measurement are calculated by taking the difference between the measured slant column and the a priori slant column as follows:

$$\text{Correction}(i, j) = SC(i, j) - \Omega_{GEOS-Chem}(\text{lat}(j)) \times AMF(i, j) \quad (2.22)$$

where j represents a latitude index. The RSC is this list of *Correction* values binned by latitude using medians, and used per pixel based on the linear interpolation to that pixels latitude. The RSC is independently calculated for each of the 60 tracks, at each latitude in the 500 0.36° bins. This provides a different RSC for each of the three AMFs. Due to incomplete latitudinal coverage, the correction for each track is interpolated linearly between measurements, with corrections outside of the highest measured latitudes being equal to the corrections at the highest measured latitudes.

Figure 2.21 shows an example of the 60 track corrections for January 1st 2005. The points are satellite measurements longitudinally averaged over the remote Pacific, coloured by track number. Another way to look at this correction is given in the OMI corrections panel of Figure 2.20, which has tracks along the x axis and latitude on the y axis. This figure shows how corrections are distributed (over this 8-day sample) with more negative values towards the outside tracks, especially in the tropics.

2.6.6 Corrected vertical columns

Corrected vertical columns, or VCC, are created using the difference between slant columns (SC) and reference sector corrections (RSC, detailed in prior section) divided by the AMF.

$$VCC = \frac{(SC - RSC)}{AMF} \quad (2.23)$$

This is equivalent to taking the difference between the slant column and its measured reference sector equivalent (SC_0), and then adding the modelled reference sector column ($\Omega_{GC,0}$):

$$VCC = \frac{(SC - SC_0)}{AMF} + \Omega_{GC,0} \quad (2.24)$$

This method is used in several works, including **DeSmedt2008; DeSmedt2012; DeSmedt2015; Barkley2013; Bauwens2016**. Recently this correction was expanded (for OMI data) to

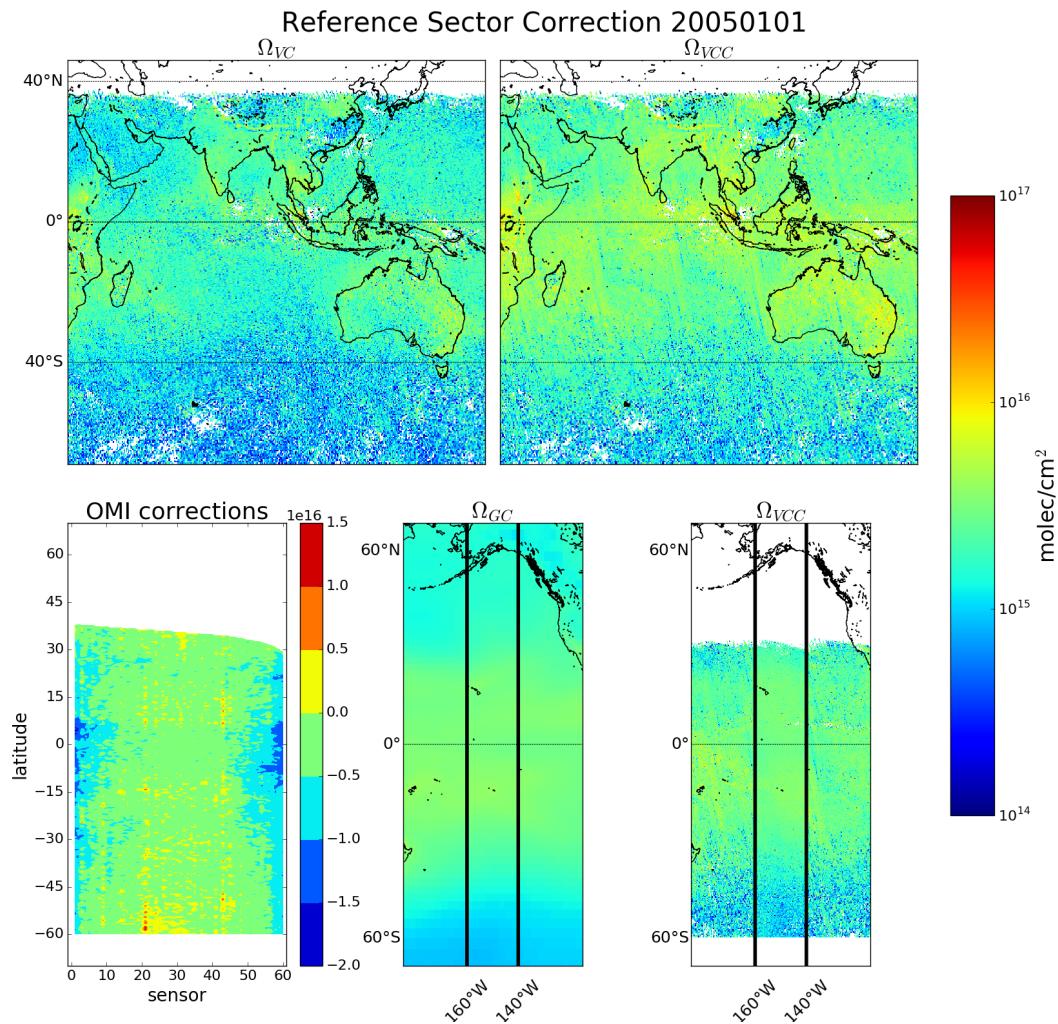


FIGURE 2.20: Example of remote Pacific RSC using 8-day average measurements and one month modelled data. Ω_{VC} shows the uncorrected vertical columns, while Ω_{VCC} shows the corrected vertical columns. OMI corrections shows the correction applied globally based on latitude and OMI track number(sensor). Ω_{GC} shows the GEOS-Chem modelled HCHO VC over the reference sector (region within black vertical lines), with Ω_{VCC} showing the corrected VC over the same area.



Figures/OMI_link/RSC_track_corrections20050101.png

FIGURE 2.21: Example of track correction interpolations for January 1st 2005, points represent the difference between satellite slant column measurements and modelled slant columns over the remote Pacific.

include latitudinal and instrument track influence by **Abad2015**.

One correction (from Equation 2.22) is associated with every good satellite measurement which is used to create a reference sector corrected measurement (Vertical Column Corrected or VCC) through the following equation:

$$VCC(i,j) = \frac{SC_{HCHO}(i,j) - Correction(i, lat(j))}{AMF(i,j)} \quad (2.25)$$

For each good satellite measurement the corrected vertical column is calculated two times, once using each new AMF.

2.6.7 Binning the results daily

Finally the pixels are binned into a gridded dataset named OMHCHORP, as shown in Figure ???. The resolution is chosen to match the native resolution of the GEOS meteorological fields ($0.25^\circ \times 0.3125^\circ$). A bin entry count is used to allow easy re-binning, and can be used to check for sparse data days due to filtering. Data averaged into this dataset:

1. OMI slant column
2. OMI air mass factor
3. OMI vertical column
4. OMI corrected vertical column
5. GEOS-Chem recalculated air mass factor
6. GEOS-Chem recalculated vertical column
7. GEOS-Chem recalculated corrected vertical column
8. GEOS-Chem air mass factor recalculated using PP code (AMF_{PP})
9. GEOS-Chem corrected vertical column based on PP code
10. satellite pixel counts (summed into bins)
11. OMI vertical column fitting uncertainty

This process requires processing time and storage space, and is performed on the National Computing Infrastructure (NCI) supercomputer. In order to reprocess one year of swath files (~ 162 GB = $142 + 16 + 4$ GB OMHCHO, MOD14A1, and OMNO2d respectively per year) of daily data was downloaded and then transformed into ~ 8 GB (per year) of daily gridded data. This takes around 90 minutes per day; however, each day is completely independent and can be run in parallel once model output is available. Initially parallelism was built into the python code; however, simply running python code independently for each date was simpler and more scalable. As much as possible this work uses HDF-5 or NetCDF-4 formats, although GEOS-Chem output is in bitpunch format. The scripts to regrid and reprocess the swath data set are available from github at https://github.com/jibbals/OMI_regridding.

2.6.8 Difference between original and corrected OMI HCHO columns

Corrected vertical columns (VCC) of HCHO are created at $0.25^\circ \times 0.3125^\circ$ horizontal resolution. Figure 2.22 shows how the recalculated columns compare to the original (OMI, left). Recalculation increases the January column amounts while also slightly flattening the distributions (row 3). Recalculated columns show more high concentration ($> 1.5 \text{ molec cm}^{-2}$) columns and lower background concentrations over the ocean. Together this may lead to a slightly steeper continental enhancement which may be expected since elsewhere OMI HCHO is seen to be up to 40% underestimated (Zhu2016; DeSmedt2015; Barkley2013).

Figure 2.23 shows vertical columns of HCHO for the original satellite swaths, those recalculated without changing the provided scattering weights, and fully recalculated vertical columns. Each grid square (at $0.25^\circ \times 0.3125^\circ$ resolution) has been created by binning the recalculated satellite pixels within the month. The average number of pixels per land square is inset as text, changing due to how the fire filter is applied. Each row has a stricter fire filter applied from top to bottom, with no fire filter on the first row up to filtering pixels from squares with fires up to 8 days prior. This figure looks at March 2005 with biomass burning filtered differently in each row. Active fires over the last 0, 1, 2, 4, and 8 days are filtered as the row number increases.

Figure TODO shows an analysis of the differences between running the recalculations with and without updating the ω_z .

TODO: compare how many nans from palmers code to my code The AMF calculated using professor Palmer's produces TODO: MORE OR FEWER reasonable (within screening range of -5×10^{15} to 1×10^{17}) vertical column amounts. Stricter filtering must be balanced against both coverage and the sensitivity of the AMF determination to recalculating ω_z .

Figure TODO: shows global and Australian HCHO averaged total column maps for January 2005, along with the reduced major axis (RMA) regression correlation and percentage difference. This comparison shows how reprocessing with an updated model can have a systematic influence on the total column.

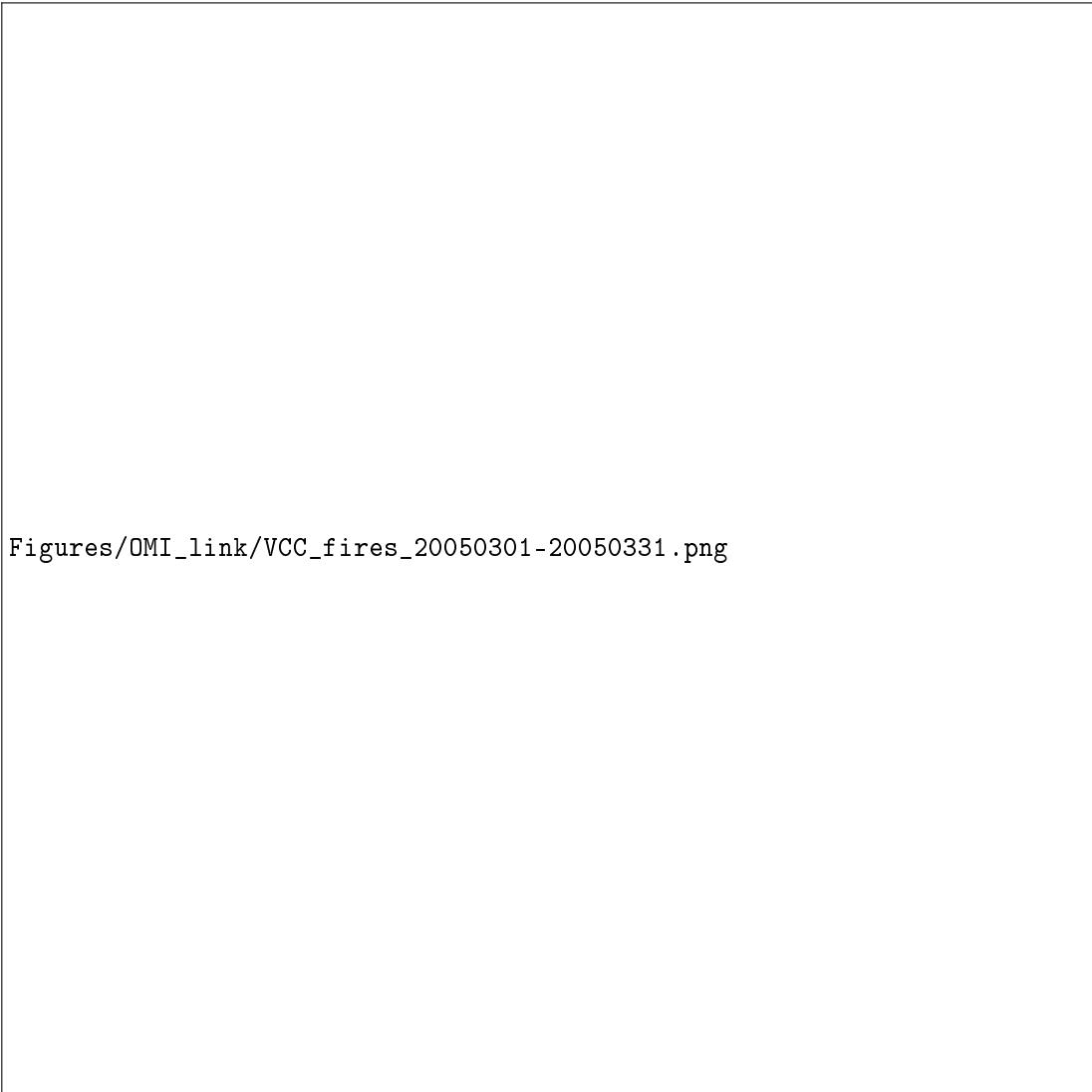
2.7 Filtering Data

A major goal in this work is to infer biogenic isoprene emissions from HCHO columns. Isoprene is the dominant source of HCHO over continental land masses; however, other precursor NVMOCs can contribute to observed HCHO. The main interference in the biogenic isoprene signal in HCHO comes from fire smoke plumes and major anthropogenic source regions. Biomass burning can be a large local or transported (via smoke plumes) source of HCHO, glyoxal (CHOCHO), and other compounds which influence HCHO levels. Anthropogenic emissions from power generation, transport, and agriculture can influence these levels as well. So to infer biogenic isoprene emissions, pyrogenic and anthropogenic influences should be removed (where possible) from modelled and measured data. In GEOS-Chem we simply turn off pyrogenic and anthropogenic emissions; however, in the OMI HCHO satellite product we mask potentially affected pixels.



Figures/OMI_link/VCC_check_200501.png

FIGURE 2.22: Row 1: regridded corrected Ω_{HCHO} from OMHCHO on January 1, 2005: original (left), recalculated using new shape factors (middle), and additionally using updated scattering weights (right). Row 2: shows the monthly average for January 2005. Row 3: shows the distribution over the month for each of the three column amounts. Distribution bins are logarithmic, resulting in wider bins for higher column amounts.



Figures/OMI_link/VCC_fires_20050301-20050331.png

FIGURE 2.23: Column 1: Reference sector corrected HCHO vertical columns Ω from OMHCCHOv003. Column 2: Ω with recalculated a priori shape factors using GEOS-Chem v10.01. Column 3: Ω with recalculated a priori shape factors and scattering weights using GEOS-Chem v10.01 and LIDORT. Row 1-5: increasing number of prior days which have active fires are included when masking fire influence.

TABLE 2.4: Satellite pixels filtered by pyrogenic and anthropogenic masking. In parenthesis are the portion of pixels filtered.

Year	Pixels	Pyrogenic	Anthropogenic	Total
2005	3.9e+06	4.1e+05(10.7%)	5.0e+04(1.3%)	4.4e+05(11.5%)
2006	3.8e+06	5.1e+05(13.5%)	9.2e+04(2.4%)	5.6e+05(14.7%)
2007	3.7e+06	4.4e+05(11.9%)	7.5e+04(2.0%)	4.9e+05(13.0%)
TODO				
TODO				

Pixels: how many land pixels are read over Australia, after cloud fraction filtering.

Pyrogenic: how many of these pixels are removed by the pyrogenic filter.

Anthropogenic: how many are removed by the anthropogenic filter.

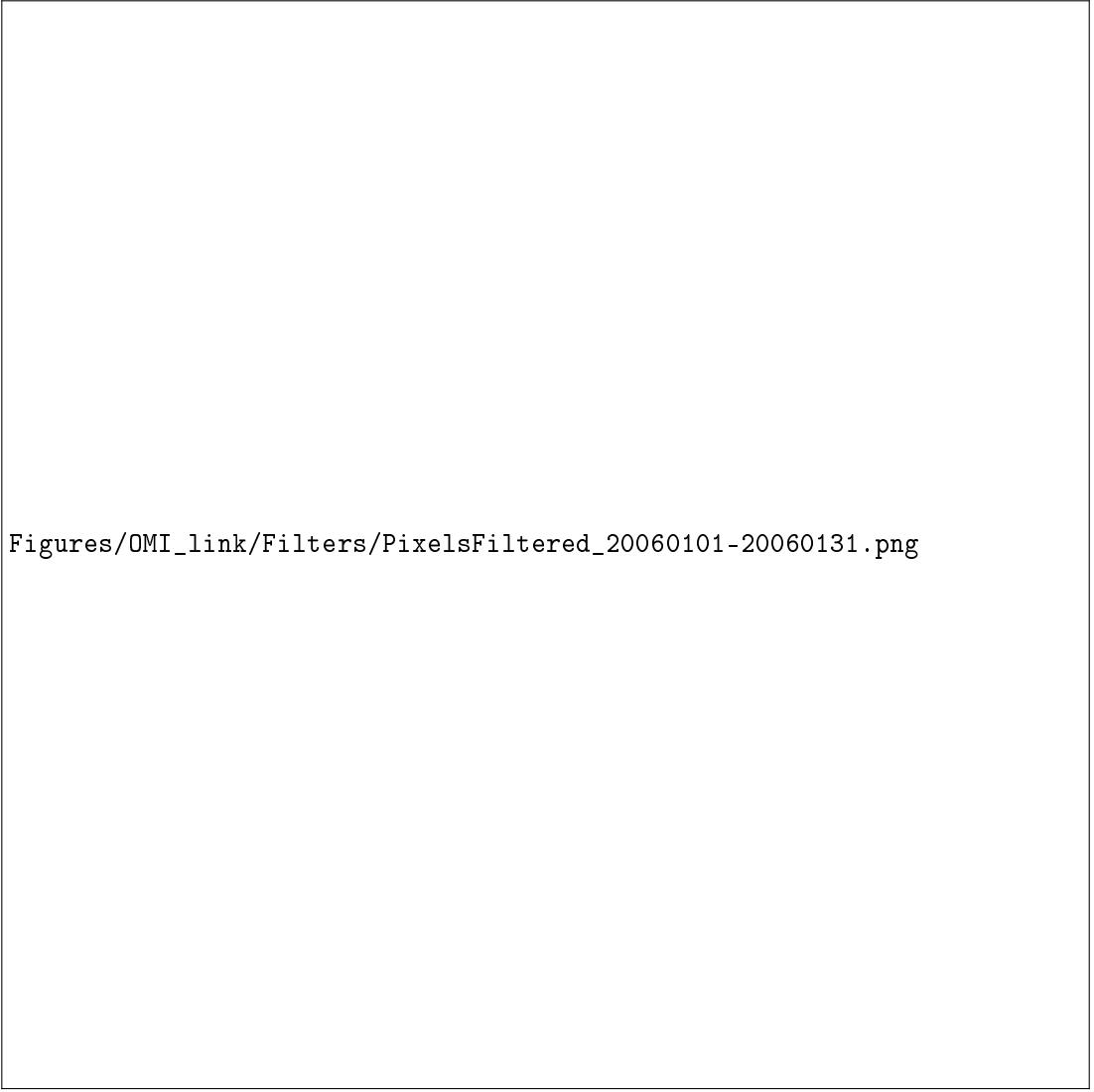
In this work anthropogenic and pyrogenic influences on the OMHCHO satellite HCHO columns are removed by masking active fires, high AAOD, and high NO₂ levels measured by satellite. Active fires and suspected smoke plumes are masked, and together termed the pyrogenic filter. NO₂ measurements are used to mask potential anthropogenic influence. A summary of filtering over Australian land squares at 0.25° × 0.3125° resolution is provided in Table 2.4. Figure 2.24 shows an example month of anthropogenic and pyrogenic filtering, highlighting how many days and pixels are removed across Australia. The anthropogenic filter completely removes grid squares over Brisbane, Melbourne, and Sydney. Other major cities in Australia either do not emit enough NO₂ or are too dispersed and do not breach the threshold to be filtered as anthropogenic. Pyrogenic filtering removes upwards of 10% of the available pixel measurements per year. This filter tends to cover forested areas (as they are more prone to burning) as well as some hot spots that are likely due gas flaring or burning.

2.7.1 Pyrogenic filter

MODIS fire counts are used in conjunction with smoke AAOD enhancements from OMI to remove data points which may be affected by fires or fire smoke plumes. The MODIS fire counts come from a combination of measurements from the Terra and Aqua satellites (Terra overpasses at 10:30, 22:30 LT; Aqua at 13:30, 01:30 LT). satellite AAOD from product OMAERUVd (described in Section 2.2.1.3) is analysed over Australia to determine a suitable filter threshold. AAOD is used instead of an alternative product AOD as it is less sensitive to the presence of clouds (Ahn2008).

OMHCHO total column HCHO Ω is processed into a 0.25° × 0.3125° horizontal daily grid. Pyrogenic filters are interpolated to the same horizontal resolution as Ω to simplify application. The following steps are performed to create the pyrogenic influence mask:

1. MOD14A1 daily gridded Aqua/Terra combined fire counts (1x1 km²) are binned into 0.25° × 0.3125° bins (matching the resolution of binned Ω).



Figures/OMI_link/Filters/PixelsFiltered_20060101-20060131.png

FIGURE 2.24: Top row shows grid squares filtered out by pyrogenic (left) and anthropogenic (right) influence masks during January 2006. Along the bottom is the time series of total pixels over Australian land squares with and without filtering the data.

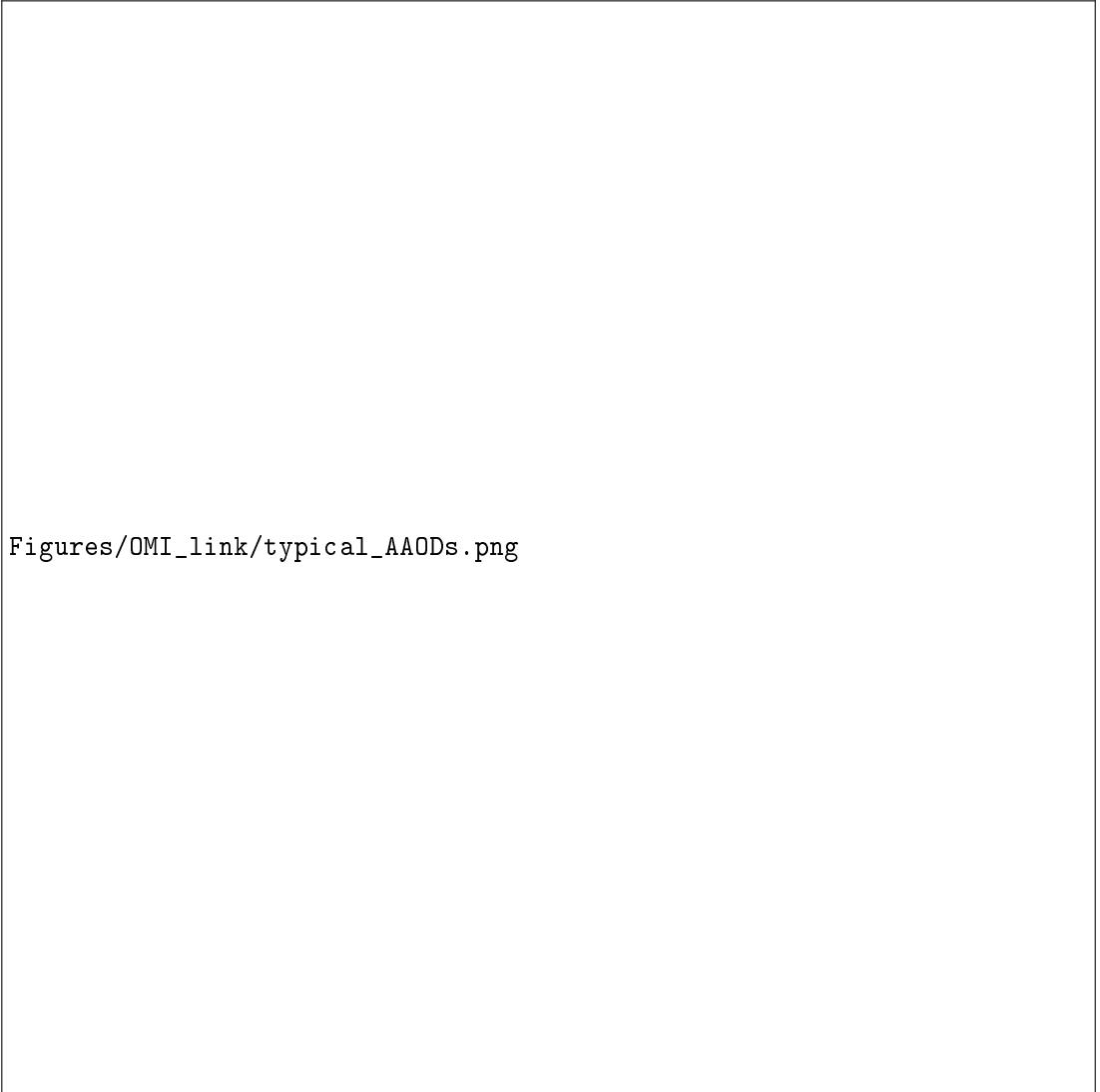
2. A rolling mask is formed which removes Ω if one or more fires are detected in a grid square, or in the adjacent grid square, up to 2 days previously. This includes the current day, making 3 days of fires in total being filtered out on each day.
3. AAOD at 500 nm is mapped from OMAERUVd ($1 \times 1^\circ$ resolution) onto the $0.25^\circ \times 0.3125^\circ$ resolution.
4. An AAOD threshold of 0.03 is determined through visual analysis of AAOD distributions over several days, including days with and without influence from active fires, dust, and transported smoke plumes. Grid squares with AAOD over this threshold are considered potentially affected by transported fire smoke.

This method of masking fires can be compared to **Marais2012** and **Barkley2013**: **Marais2012** removed pixels colocated with non zero fire counts in any of the prior eight days, within grid squares with $1^\circ \times 1^\circ$ resolution, and **Barkley2013** used fires from the preceding and current day, within local or adjacent grid squares, at $0.25^\circ \times 0.3125^\circ$ resolution.

Determining the AAOD due to smoke can be difficult since both smoke and dust absorb UV radiation (**Ahn2008**; **Marais2012**). AAOD filtering is designed to remove pixels affected by smoke; however, it may occasionally remove pixels affected by dust. Dust in Australia is highly episodic and false positives in the smoke filter should not affect more than a few days per month, especially over regions with high tree coverage (**Shao2007**). The threshold is determined through analysing AAOD over Australia in 4 scenarios: normal conditions, active local fires, transported fire smoke, and large scale dust storms. An example of these scenarios and the AAOD distributions is shown in Figure 2.25. This figure shows AAOD maps and distributions, along with satellite imagery on the same day in column 4 (from <https://worldview.earthdata.nasa.gov/>).

Figure 2.26 shows the number of grid square days filtered by the smoke mask and pyrogenic mask (smoke and active fires combined) over 2005. The Lake Eyre basin is the area most affected by smoke filtering, likely due to dust influence in the arid region. The scale of days filtered by the active fires mask is more than 4 times that of the smoke mask, and is more prominent over heavily forested areas such as the northern and eastern coast lines. Masking for active fires impacts analysis to a far greater extent than masking smoke.

Figure 2.27 shows the extent of pixels filtered out by the pyrogenic filter over 2005. The highest concentrations of removed pixels lie along the northern and eastern coastlines, and correspond with forested areas (see Figure 2.29), which suggests that forest fires are being masked properly. Central Australia is largely unmasked, which could be due to a lack of sufficient vegetation to create fires and smoke visible by satellite. The proliferation of petrol or gas wells (see Figure 2.28) may also lead to AAOD enhancement wherever activity stirs up dust, and could be mistaken as active fires wherever gas flaring occurs. In 2005, 388 gas wells existed in Queensland; however, more than 2000 wells (cumulative) were approved by 2013 which may cause more filtering over the analysed time period in this thesis (**Carlisle2012**). To check this the filtering portion for 2012 is also plotted in Figure TODO: plot map of filtered squares for 2012 when it has been run by NCI. One clear hot spot is located over Port Kembla (south of Sydney), most likely due to the flame which burns over the blast furnace stack throughout the year (<https://www.bluescopeillawarra.com>.



Figures/OMI_link/typical_AAODs.png

FIGURE 2.25: AAOD from OMAERUVd (columns 1, 2, 3) over Australia for four different scenarios (rows 1-4). Scenes from the same day are taken from the EOS Worldview website <https://worldview.earthdata.nasa.gov/>. AAOD = 0.03 is demarcated by a horizontal line in the density plots in column 3.



FIGURE 2.26: The left panel shows how many days are filtered using the AAOD threshold, while the right panel shows how many are filtered using the combined smoke and active fire masks. The time series shows pixels available for analysis before being filtered (black), after smoke filtering (cyan) and after the full fire filter is applied (magenta).

[au/community/skylineimages/](#)). Another hot spot can be seen in Western Australia over Kalgoorlie, where a large open cut gold mine is always open and blasting daily TODO: Reference. In Western Queensland over Mount Isa there is again a mining hotspot TODO: Reference. A large area in southern Queensland/northern NSW is also heavily filtered, potentially due to gas flaring in the Surat Basin, which has thousands of petrol and gas wells (<http://www.ga.gov.au/scientific-topics/energy/resources/petroleum-resources/gas>).

Biogenic HCHO concentrations should be correlated with temperature, since isoprene emissions are strongly correlated with temperature (Palmer2006; Zhu2013_poster; Surl2018). Fires emit HCHO precursors and increase HCHO concentrations independently of the relationship between temperature and HCHO, and should be revealed as outliers when comparing HCHO to temperature. Figures 2.30 - 2.32 show the relationship between modelled temperature, and satellite HCHO for January 2005 within subsets of Australia. A reduced major axis regression is used to determine the correlation between surface temperature (X axis) and HCHO (Y axis). Using the natural log of HCHO we can take the linear regression and then exponentiate each side in the equation $\ln Y = mX + b$ to get $Y = \exp mX + b$. This gives us the exponential fit as shown, with the correlation coefficient between $\ln HCHO$ and temperature. The distributions of exponential correlation coefficients and m terms is shown in the embedded plot, with one datapoint available for each grid square where the regression is performed. These figures show that the modelled temperature does not well correlate with corrected recalculated OMI vertical columns (r ranges from -0.24 to 0.47), but is with modelled columns (r ranges from 0.49 to 0.84). Correlations between modelled temperatures and HCHO are further improved when using the spatial average within each region. Further, we see that the relationship is improved in individual grid squares over south Eastern Australia by removing non-biogenic emissions from the model (r increased from 0.58 to 0.75). This improvement is not seen in Northern Australia, nor south Western Australia. Overall this suggests that modelled correlations between temperature and HCHO are spatially dependent, and generally well reflected in satellite measurements over Australia.

One problem with detecting outliers in the temperature and enhanced HCHO relationship is that days when fires occur are likely to be hot. Another problem with correlating heat and HCHO is that increased temperature accelerates HCHO destruction (Zheng2015). We test the fire mask by examining the relationship between modelled temperature and satellite HCHO with and without applying the filters for smoke and active fires. Figures 2.33 and 2.34 show the regressions between OMI HCHO total columns and temperature from GEOS-Chem output and CPC daily maximum temperatures respectively. Comparing against GEOS-Chem modelled surface temperatures at $2^\circ \times 2.5^\circ$ shows??? TODO. The left column in Figure 2.33 shows scatter and RMA correlation within a single grid box over 2 months from Jan 1 to Feb 28, 2005, without having applied either the fire nor anthropogenic masks to OMI HCHO columns. The right column shows the same correlation after applying the fire filter, affected data points are marked in teal (matching the red marked points in the left column). The analysis is repeated for Sydney, Canberra, and three grid squares to the north-west, west, and southwest of Sydney ($w1, w2, w3$ respectively). Figure 2.34 shows the same analysis at higher resolution using CPC daily maximum temperatures (see 2.2).



Figures/OMI_link/Filters/Pyrogenic_filter_annotated_2.png

FIGURE 2.27: Top: percentage of pixels filtered out by fire and smoke masks in 2005. Bottom: percentage filtered out each day from land squares in Australia.

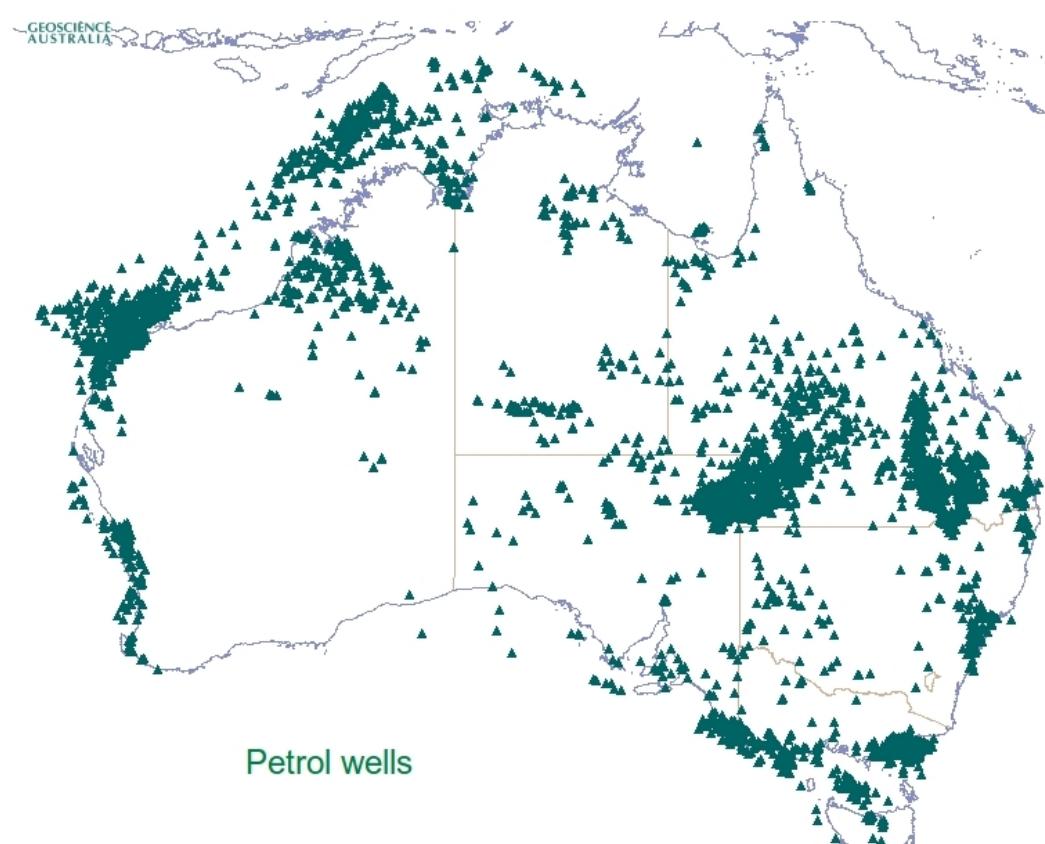
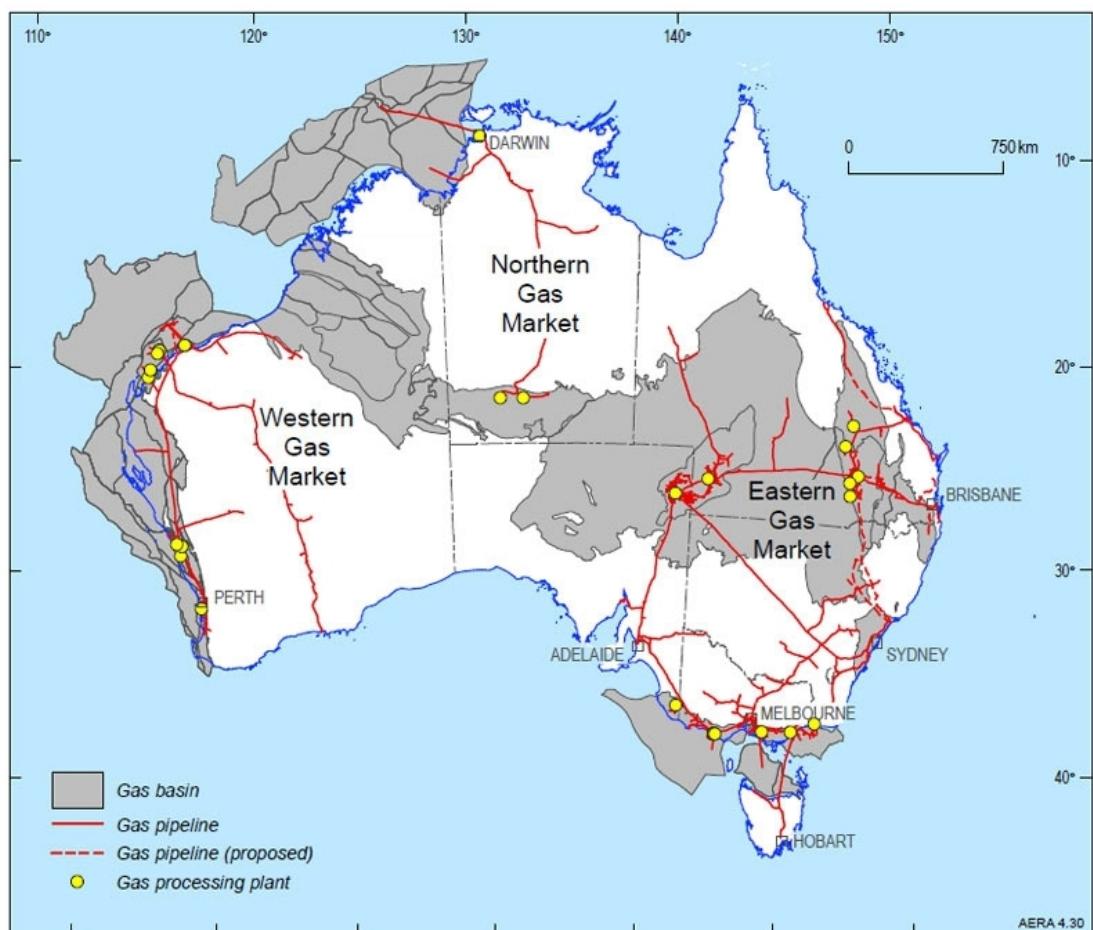


FIGURE 2.28: Top: gas fields and pipelines (2018) for Australia (<http://www.ga.gov.au/scientific-topics/energy/resources/gas-fields-and-pipelines/>). Bottom: petrol well locations over

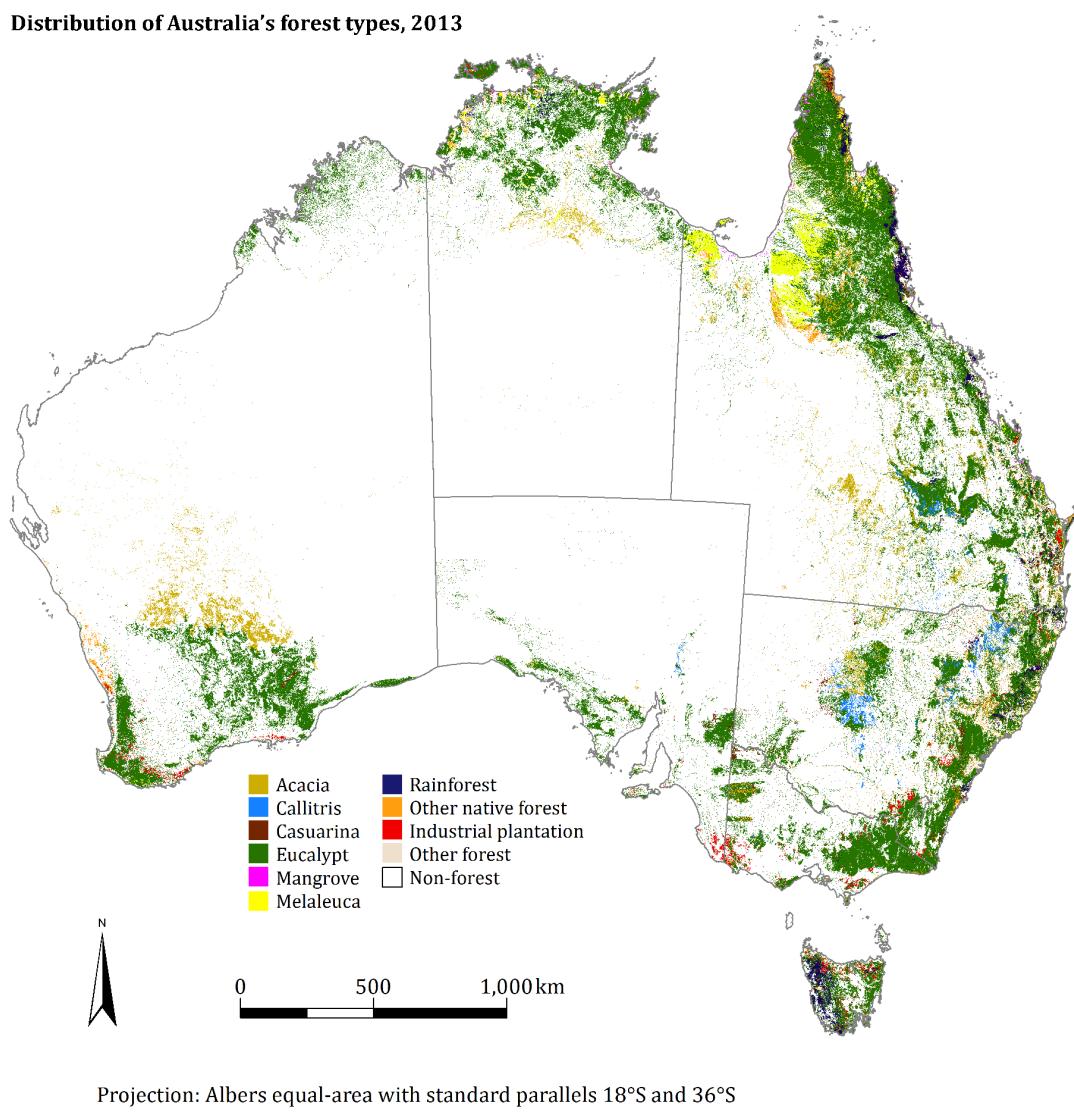
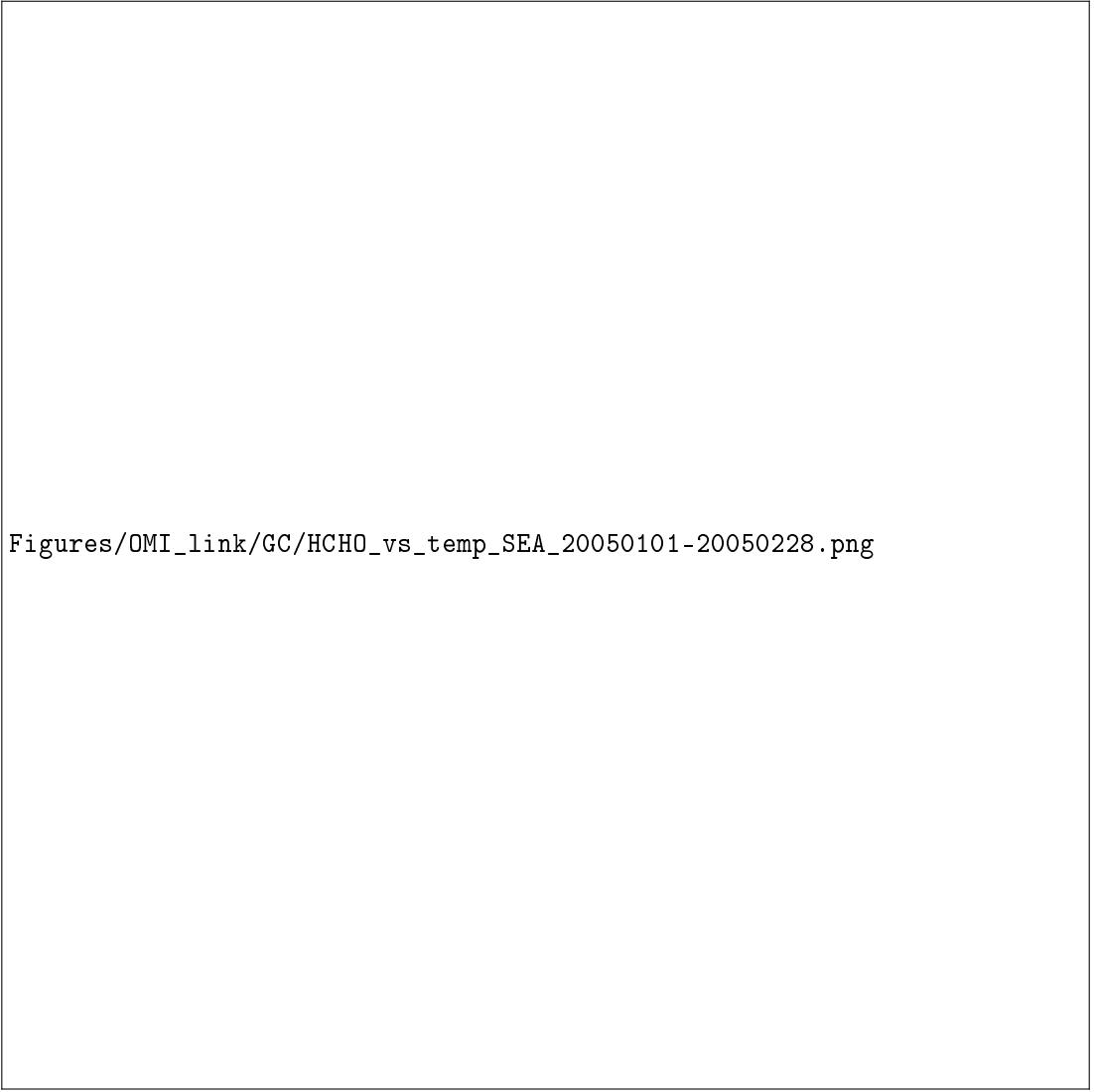


FIGURE 2.29: Forest coverage, coloured by predominant tree species.



Figures/OMI_link/GC/HCHO_vs_temp_SEA_20050101-20050228.png

FIGURE 2.30: Top row (left): surface temperature averaged over January and February 2005. Top row (right): correlation between spatially averaged GEOS-Chem temperatures and recalculated satellite vertical columns. Second row: GEOS-Chem surface temperatures correlated against GEOS-Chem HCHO, with different colours for each grid box, and black showing the spatially averaged correlation over time. Third row: as second row, except GEOS-Chem HCHO comes from the biogenic emissions only simulation. A reduced major axis regression is used within each gridbox using daily overpass time surface temperature and HCHO. The distribution of slopes (solid) and regression correlation coefficients (dashed) for the exponential regressions is shown in the inset panels in rows 2 and 3.



Figures/OMI_link/GC/HCHO_vs_temp_NA_20050101-20050228.png

FIGURE 2.31: As Figure 2.30 but for northern Australia.

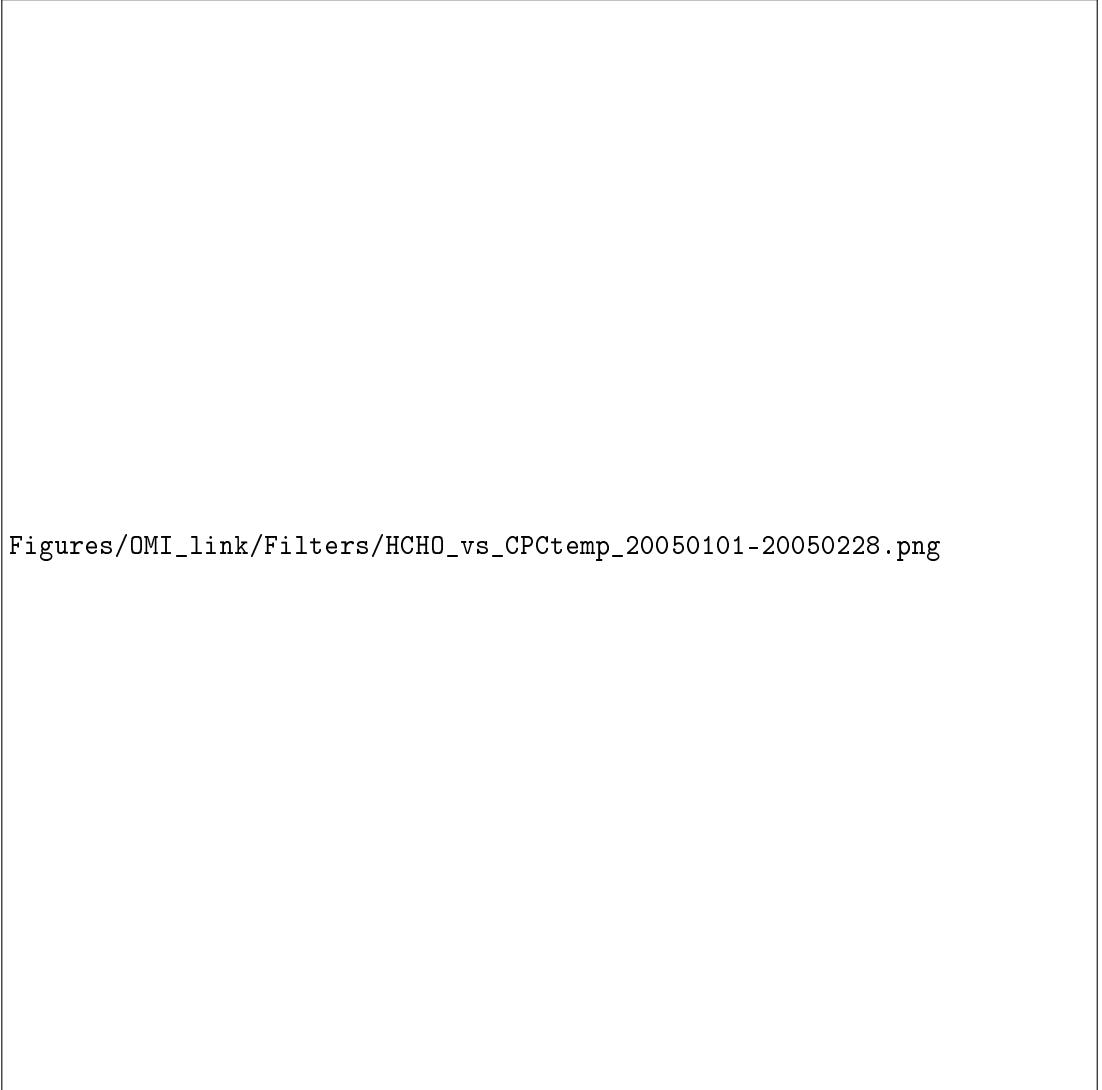
Figures/OMI_link/GC/HCHO_vs_temp_SWA_20050101-20050228.png

FIGURE 2.32: As Figure 2.30 but for south-western Australia.



Figures/OMI_link/Filters/HCHO_vs_temp_20050101-20050228.png

FIGURE 2.33: HCHO vs GEOS-Chem daily midday temperatures
TODO add longer caption



Figures/OMI_link/Filters/HCHO_vs_CPCtemp_20050101-20050228.png

FIGURE 2.34: HCHO vs CPC daily maximum temperatures TODO add
longer caption

TABLE 2.5: NO₂ averages by region before and after filtering for anthropogenic emissions using 2005 data from the OMNO2d product.

Region	NO ₂	NO ₂ after filtering	% Data lost
Aus	1	2	3
BG	1	2	3
Syd	1	2	3
Melb	1	2	3
Adel	1	2	3

2.7.2 Anthropogenic filter

Enhanced NO₂ concentrations indicate anthropogenic influence over Australia. We use NO₂ as a proxy for potential anthropogenic NMVOC emissions, as these could bias the inversion performed in Chapter 3. A filter is designed using the tropospheric NO₂ columns in the OMNO2d product.

NO₂ columns near several major cities in south eastern Australia over 2005 are used to determine a suitable threshold for anthropogenic influence. The mean, standard deviation, and time series over Australia of tropospheric NO₂ measured by satellite is shown in Figure 2.35. Tropospheric NO₂ columns averaged within all of Australia and each region is shown in Figure 2.35. Anthropogenic influences are clearly visible near major cities in Australia, and some influence can be seen along nearly every coastline.

The anthropogenic filter is created for each year from in two steps:

1. Daily grid squares with NO₂ greater than 10^{15} molec cm⁻² are flagged as anthropogenic.
2. After taking the yearly average for each grid square, any tropospheric NO₂ columns greater than 1.5×10^{15} molec cm⁻² are flagged for the whole year.

These thresholds are chosen subjectively through trial and error so as to ensure the removal of definite anthropogenic influence while not removing too many data points over near uninhabited portions of Australia. These thresholds completely remove grid squares over major cities which are likely emitting NMVOCs year round, and also frequently remove grid squares down wind. The affects of applying this filter to the OMNO2d product itself can be seen in Figure 2.36. Areas over major cities in south eastern Australia do not lie outside the IQR of Australian NO₂ tropospheric columns once the filter is applied. Many coastal regions show some effect from the filtering, and Sydney, Melbourne, and east of Melbourne are completely removed.

The same regions as in Figure 2.35 are shown again in Figure 2.7.2, with NO₂ pixel densities for each region shown, along with the threshold of 1×10^{15} molec cm⁻². This led to a reduction of TODO grid squares from the total available measurement space over Australia. The removal of grid squares which went above the yearly averaged limit of 1.5×10^{15} molec cm⁻² further reduced the available data by TODO grid squares.

Vertical dashed lines show the threshold for anthropogenic influence, any columns above this value are filtered out.

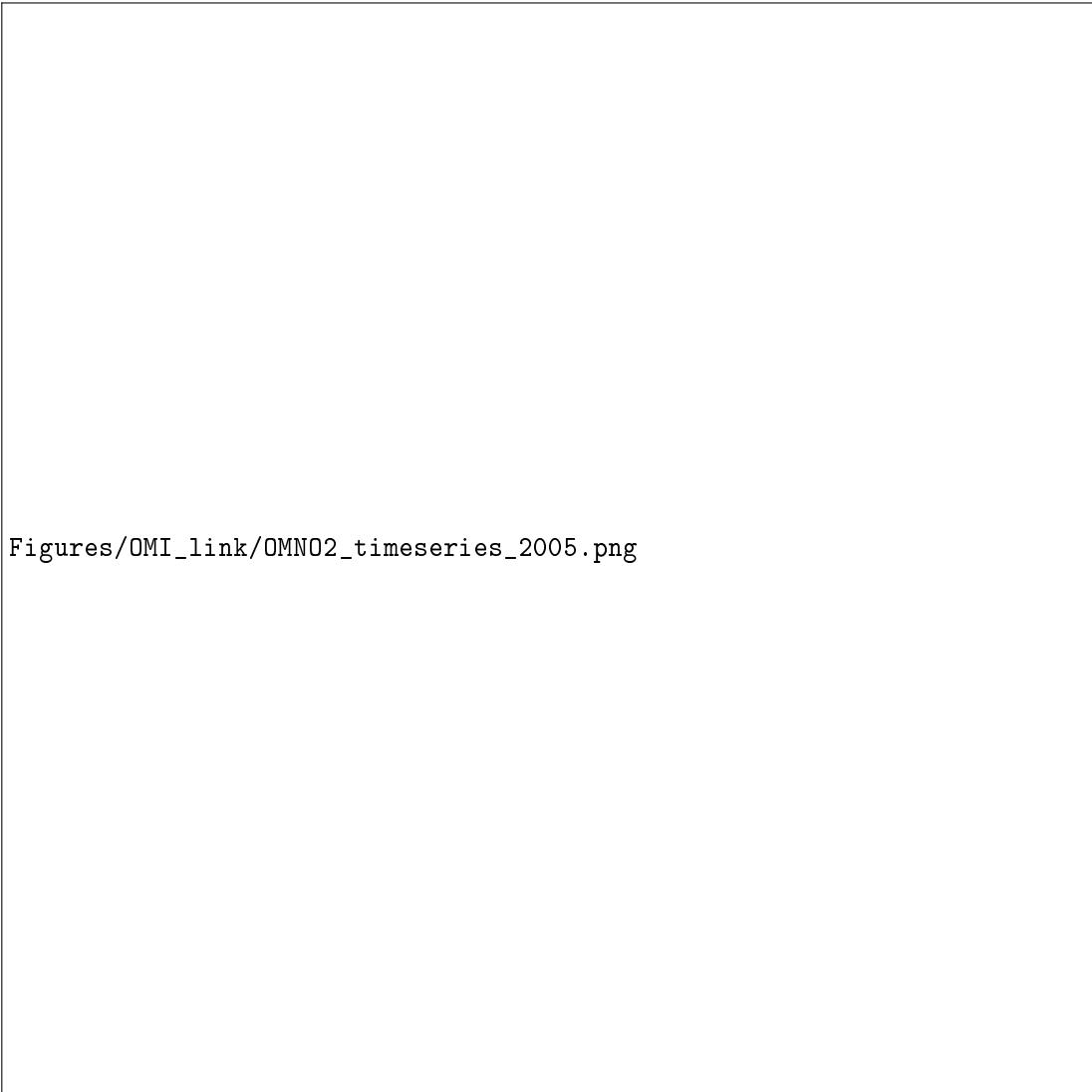
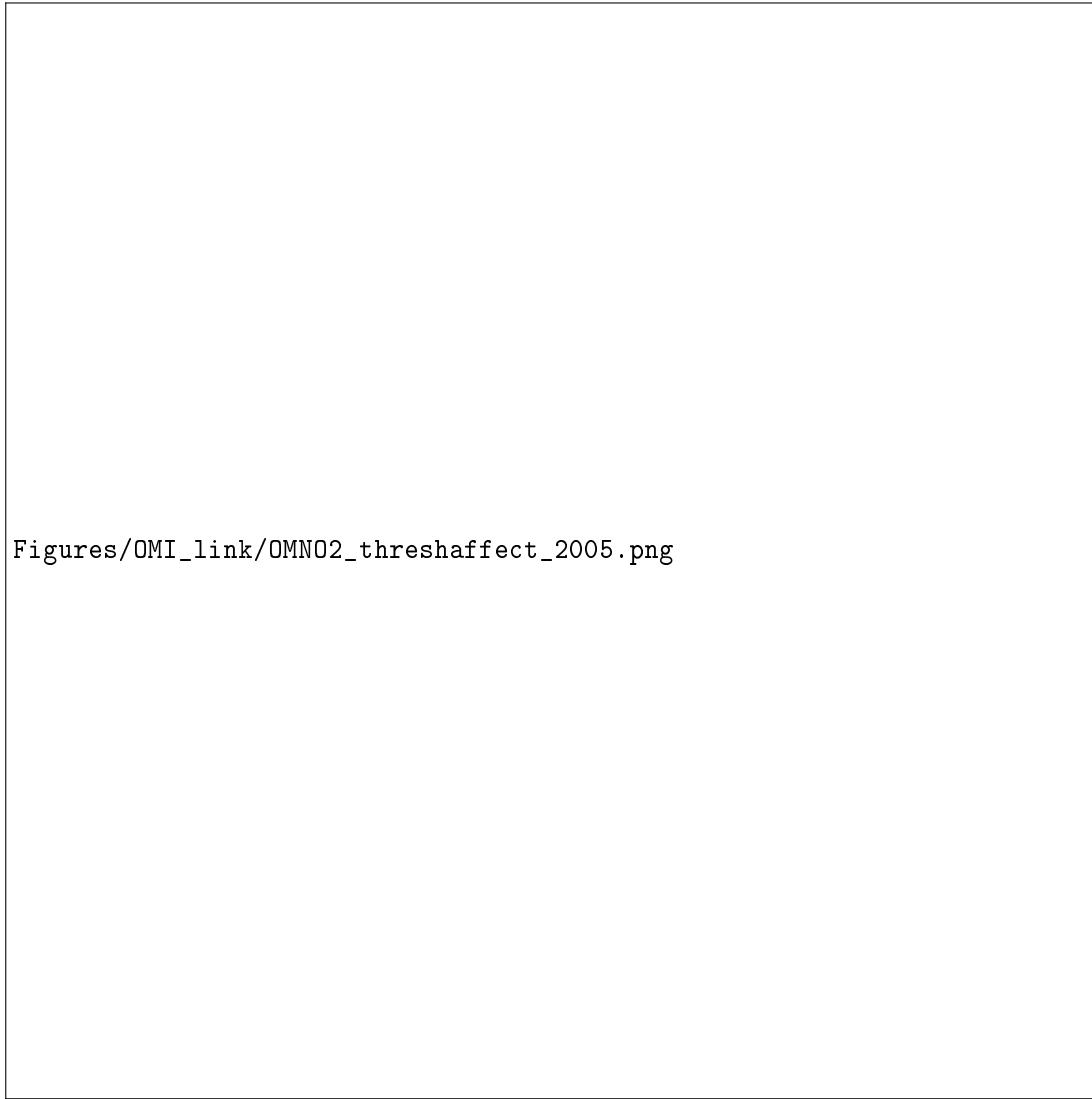
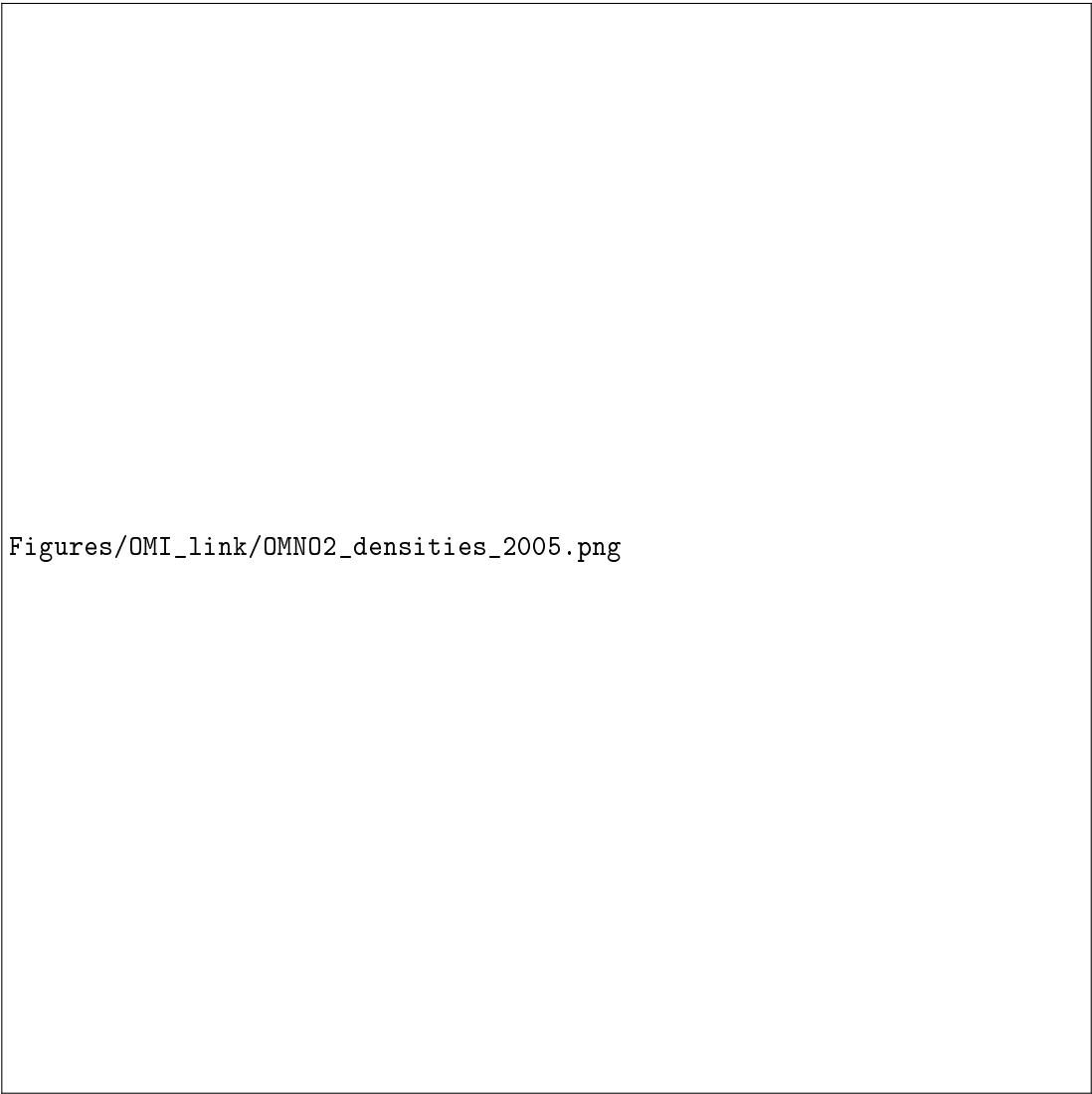


FIGURE 2.35: Mean (top left) and standard deviation (top right) of OMNO2d daily $0.25^\circ \times 0.25^\circ$ tropospheric cloud filtered NO₂ columns. Time series for Australia, and each region (by colour) shown in the bottom panel, with mean for that region shown on the right. The grey shaded area depicts the 25th to 75th percentiles of Australia averaged NO₂ columns for each day in the time series, with a thicker black line showing the Australia-wide mean value.



Figures/OMI_link/OMNO2_threshaffect_2005.png

FIGURE 2.36: Top row: 2005 OMNO2d NO₂ column mean before (left) and after (right) applying the threshold filters as described in the text. Bottom row: time series for Australia, and each region (by colour) shown in the bottom panel, with mean for that region shown on the right. The time series before (left) and after (right) applying the anthropogenic filter are shown, the dashed and dotted horizontal lines show daily and yearly mean thresholds respectively.



Figures/OMI_link/OMN02_densities_2005.png

FIGURE 2.37: 2005 OMNO2d NO₂ column means (top left) and distributions (top right) for Australia, and each region shown in the area map (by colour)

2.8 Process schematic

Figure 2.38 shows an overview of how vertical columns (modelled and measured) along with filters and sundry data are created in this thesis. Output from GEOS-Chem is combined with OMHCHO swath data to produce a gridded data file which contains new and old satellite vertical columns along with modelled vertical columns. The original AMF as well as those recalculated using GEOS-Chem, and associated corrected vertical columns are all binned and stored. Masks created using MODIS and OMI measurements are also stored in the OMHCHORP dataset.

2.9 Data Access

TODO: ADD MORE HERE

OMHCHO Satellite swaths of HCHO slant columns downloaded from <https://search.earthdata.nasa.gov>, with DOI 10.5067/Aura/OMI/DATA2015

OMNO2d Daily satellite NO₂ product downloaded from <https://search.earthdata.nasa.gov/search>, DOI:10.5067/Aura/OMI/DATA3007. For more information in refer to section

OMAERUVd Gridded satellite based AAOD measurements downloaded from <https://search.earthdata.nasa.gov>, DOI 10.5067/Aura/OMI/DATA3003.

SPEI Monthly standardised precipitation evapotranspiration index (metric to determine drought stress) downloaded from <http://hdl.handle.net/10261/153475> with DOI:10.20350/digitalCSIC/8508. See more information in section

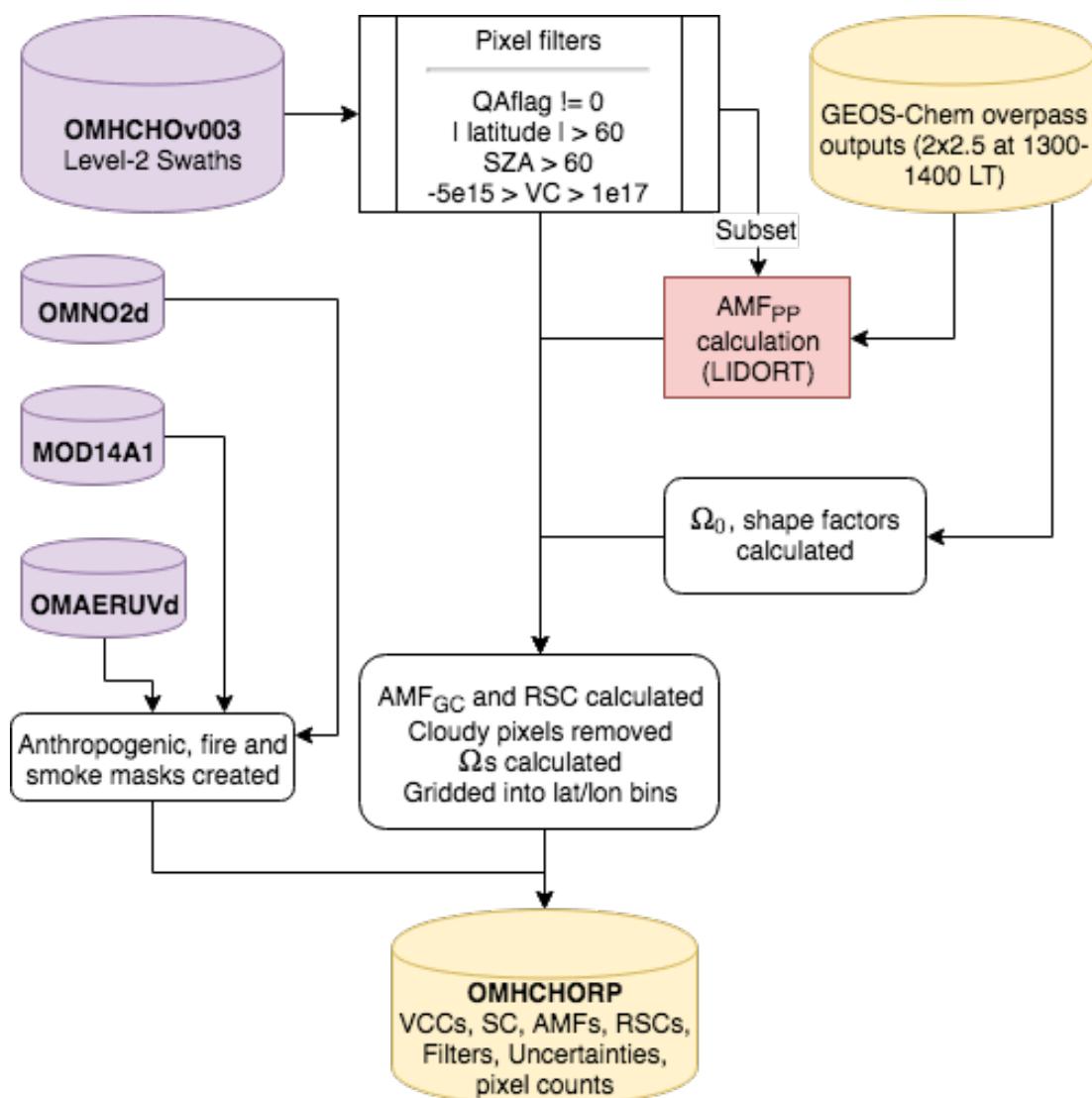


FIGURE 2.38: Flow diagram showing how OMHCHO level two swath data is read, processed, and gridded in this thesis

Chapter 3

Biogenic Isoprene emissions in Australia

3.1 Introduction

Biogenic volatile organic compounds (BVOC) affect the oxidative capacity of the atmosphere and are largely driven by what type of vegetation is in the area (**Kefauver2014**). In the troposphere, BVOC emissions affect hydroxyl radical (OH) cycling, ozone (O_3) and secondary organic aerosol (SOA) production, and methane lifetime. Australian forests are strong emitters of isoprene, the primary BVOC emitted globally (**Guenther2006; Messina2016**). Poor measurement coverage of isoprene, isoprene products, and isoprene emissions within Australia means that emissions are poorly understood. The lack of measurements makes it difficult to estimate the subsequent atmospheric processes. Isoprene (C_5H_8) is relatively difficult to measure due to its high reactivity and short lifetime.

Emission models used to derive estimates of isoprene fluxes are based on understanding the emissions from different plant species (phenotypes) in varying conditions. **Guenther2012** estimated global biogenic isoprene emissions at roughly 535 Tg yr^{-1} , while **Sindelarova2014** estimated around 411 Tg yr^{-1} . Reactions following emissions are complex, and are sensitive to other trace gases in the ambient atmosphere. Uncertainties in several important products such as ozone and SOA are increased due to both isoprene measurement difficulties and its complicated subsequent chemical mechanisms. Isoprene emissions may be overestimated in Australia since they are based on measurements taken from a few young trees (**Winters2009**) that may emit more than older trees (**Emerson2016**). The sample of trees include 4 types of Eucalyptus, which are not representative of the hundreds of species that make up Australian forests, and how these species depend on biological and meteorological stresses is unclear (**Winters2009; FortemsCheiney2012**). Emissions estimates are often used as boundary conditions for atmospheric chemistry models and improving these estimates for Australia is one goal of this thesis.

In this chapter I describe and implement a *top down* technique using satellite measurements of HCHO to calculate surface isoprene emissions. HCHO is a dominant product of most BVOCs, including isoprene, and is measured by satellites via remote sensing. In situ isoprene concentration measurements are costly and sparse within Australia, while satellite HCHO data are plentiful and freely available, making

this technique very attractive. Top down techniques have informed isoprene emission inventories in North America (**Abbot2003; Palmer2003; Palmer2006; Millet2006; Millet2008**), South America (**Barkley2013**), Europe (**Dufour2009; Curci2010**), Africa (**Marais2012**), Asia (**Fu2007; Stavrakou2014**), India (**Surl2018**), and even globally (**Shim2005; FortemsCheiney2012; Bauwens2016**). In this thesis I apply the technique solely focusing on Australia for the first time.

3.1.1 Aims

Recent work suggests that modelled emissions may be overestimated in Australia (**Emmerson2016**). This work tries to improve the understanding of isoprene emissions over the whole of Australia, clarifying the spatial distribution of bias and how these biases impact modelled chemistry. I estimate isoprene emissions in Australia using a top-down technique based on OMI HCHO measurements and GEOS-Chem modelled yields. This a posteriori top-down estimate is then evaluated against sparse available ground-based measurements. The GEOS-Chem model is modified to run with the a posteriori isoprene emissions to determine potential impact on modelled chemistry. Goodness of fit between in situ, satellite, and modelled HCHO is determined before and after scaling emissions estimates.

In this chapter I outline why current isoprene emissions estimates are inadequate and how they can be improved. I discuss literature that shows how the estimates may be too high, and describe how emissions may be calculated using satellite datasets. Section 3.2 lays out how new isoprene emissions are estimated, with results examined in Section 3.3. Updated satellite HCHO columns (Chapter 2) are compared to available measurements in Section 3.3.3. Uncertainties for each step along the way are quantified in Section 3.4. In Section 3.5 I examine how these changes in emissions would affect ozone concentrations in Australia, along with some other chemical processes.

3.1.2 Existing emissions estimates

The Model of Emissions of Gases and Aerosols from Nature (MEGAN) is one of the most widely used sources for biogenic isoprene. Along with other models that rely on measured plant emission rates, it is poorly calibrated for Australian conditions. Emissions of isoprene (C_5H_8) appear to be overestimated in within Australia (**Sindelarova2014; Stavrakou2014; Emmerson2016**). Although the lack of measurements of isoprene emission rates in Australia makes this overestimation difficult to characterise. **Stavrakou2015** saw isoprene emissions overestimated by a factor of 2-3 in January. **Emmerson2016** suggest isoprene emissions are estimated 2-6 times too high compared against available measurements of isoprene concentrations. They also show that no blanket increase or decrease in emission factors is appropriate for the entire southeast of Australia. They compared modelled data to campaign measurements from multiple sites over different seasons and found that scaling emissions did not universally improve model outputs.

Recently **Bauwens2016** estimated isoprene emissions with a top-down technique using the IMAGESv2 global CTM. They calculated emissions that create the closest match between model and satellite vertical columns, and compared these a posteriori data with their a priori (model data) and independent data sets. For Australia

they found emissions ranging from 26–94 Tg C yr⁻¹, with MEGAN a priori emissions of 38 Tg C yr⁻¹ and a posteriori emissions of 36 Tg C yr⁻¹, although the 94 Tg C yr⁻¹ estimate was also based on MEGAN. In this thesis I focus on the analysis of a top-down emissions estimate compared against MEGAN, along with how changed emissions affect modelled ozone levels.

3.1.3 Top-down isoprene emissions estimates

In the remote troposphere HCHO production is dominated by methane oxidation, while in the continental boundary layer production is largely due to non-methane VOCs (NMVOCs) (Abbot2003; Kefauver2014). This leads to a causal relationship between enhanced HCHO concentrations and NMVOC emissions at low (< 1 km) altitudes. NMVOCs are generally short lived (< 1 hr), and the most prominent of these is isoprene. Isoprene is emitted and enters the atmosphere in the gas phase, where it begins a complex series of reactions. HCHO is produced with high yield in many reactions beginning with isoprene oxidation, discussed in more detail in Section 1.3.3, and has a lifetime of a few hours (Kefauver2014).

Top-down estimates determine emissions of a particular species through careful analysis of the measurable products of that species. This generally takes advantage of longer-lived products that may reach an equilibrium in the atmosphere. For isoprene this is done through examination of atmospheric HCHO enhancement, which is a major product of isoprene oxidation that takes place after emission. Since 1997, when GOME measurements were first used to measure HCHO over Asia, satellites have been used to provided a total column measurement of HCHO, providing isoprene emissions estimation by top-down methods (Thomas1998; Palmer2001; Bauwens2016). Using satellite information to improve estimates of biogenic emissions has been highlighted as a valuable use of satellite derived datasets (Streets2013). Here NASA's OMHCHO product based on measurements from the OMI instrument onboard the Aura satellite (see Section 2.3) is the basis for a top down biogenic isoprene emission estimate over Australia.

There are two top-down isoprene emission estimation techniques, Bayesian and linear, which are discussed briefly here. Both the linear and Bayesian techniques assume that modelled chemistry is accurate and only try to correct precursor emissions. This is an additional source of uncertainty given existing uncertainties in chemical mechanisms.

3.1.3.1 Bayesian inversion

Bayesian inversion optimises model parameters in order to minimise the difference between model output and an (ideally) independent dataset such as satellite measurements. Emissions of isoprene (and other precursors to HCHO) will form part of the set of model parameters that are adjusted to make the model HCHO output most closely match satellite measurements. These inversions can be set up to account for effects from transport and allow source attribution (Curci2010; FortemsCheiney2012).

In general; a model (the forward model) is used to determine the relationship between HCHO (y) and the state variable x , which represents isoprene emissions (and

other variable parameters of interest):

$$y = \mathbf{K}\mathbf{x} + b + \epsilon \quad (3.1)$$

where ϵ are the (assumed) independent errors in measurements. K is the Jacobian matrix determined from the forward model representing the sensitivity of y to the state variable x . Essentially the K matrix is the modelled estimation of how y responds to each of the driving parameters represented by elements of x . This K matrix is used in conjunction with error covariance in x to determine the most likely solution to x , given what is known about y .

This method was used by **Shim2005** to optimise isoprene emissions in areas with high HCHO concentrations. They showed model underestimation of isoprene emissions by 14-46%, which when corrected reduced bias between GOME HCHO measurements and GEOS-Chem model output by 3-25%. More recently **Kaiser2018** showed a 40% bias in MEGAN isoprene emissions over the southeast US using a Bayesian inversion based on OMI HCHO.

An advantage (over the linear method described below) of the Bayesian method is that it can account for pyrogenic and anthropogenic emissions, as these form part of the state variable x . However, biases may still arise due to errors in modelled emission estimation (**Curci2010**). More limiting is the fact that the Bayesian method is computationally expensive due to the requirement that model runs take place using many permutations of changed input parameters. In this work I do not use the Bayesian method due to the computational costs surpassing the resources available.

3.1.3.2 Linear inversion

The linear technique is the less complicated of the two, and is performed in this thesis. Vertical columns of HCHO from satellite and modelled yield from isoprene allow the inference of local (grid space) isoprene emissions (**Palmer2003; Millet2006; Marais2012; Bauwens2016**). In Australia the effective molar HCHO yield from isoprene has not been extensively studied, while in other continents this value varies from 1-3 depending on local NO_x concentrations (**Palmer2006; Millet2006; Bauwens2016; Surl2018**). The primary assumption of the linear inversion technique is that HCHO and its precursors (primarily isoprene) are in a linear steady state relationship. This allows one to link isoprene emissions to HCHO measurements using production and loss rates. Essentially a linear relationship between total column HCHO (Ω) enhancement above a background level (Ω_0) and isoprene emissions (E_{isop}) is determined:

$$\Omega = S \times E_{isop} + \Omega_0$$

This uses modelled vertical columns and emissions to estimate the slope (S). Then this modelled S is applied to satellite measurements of Ω (Ω_{sat} and $\Omega_{sat,0}$) to determine \hat{E}_{isop} :

$$\hat{E}_{isop} = \frac{\Omega_{sat} - \Omega_{sat,0}}{S}$$

This is described further in Section 3.2, with an outline in Section 3.2.1.

The calculation requires reaction rates and yields from isoprene to HCHO, which can be determined most readily using chemical modelling. The method for calculating isoprene emissions from HCHO is laid out in **Palmer2003**, taking into account the expected lifetime and reaction rates of the precursor VOCs and HCHO. In their work, isoprene emissions fluxes over the US were derived using the Global Ozone Monitoring Experiment (GOME) satellite instrument. The method has since been applied to many regions using OMI, GOME, GOME-2, and SCIAMACHY satellite data (**Abbot2003; Barkley2013; Stavrakou2014; Surl2018**). In this thesis I apply the technique solely focusing on Australia for the first time.

The linear inversion assumes fast HCHO yield from isoprene and no precursor transport, which is unrealistic in certain scenarios; e.g. high wind speeds can transport precursors, or low NO_x concentrations can slow HCHO production (**Palmer2006; Surl2018**). Filtering out data that do not match assumptions is required but can limit the utility of this technique, and leads to some dependence on environmental factors. Uncertainties in the technique are discussed in more detail in Section 3.4.1. Nonetheless, a major benefit is that the simple nature of the inversion requires very little computational power after acquiring satellite and model datasets, even over large amounts of gridded data. This allows an inversion using more than 8 years of satellite and model data, capturing inter-annual variability over all of Australia. With the computational resources available this would not have been possible using the Bayesian inversion.

3.2 Methods

I broadly follow the method of **Palmer2001** to create a biogenic isoprene emissions estimate over Australia. A relationship is modelled between biogenic-only midday tropospheric columns of HCHO and GEOS-Chem midday biogenic isoprene emission rates, and then this relationship is applied to satellite measured HCHO total columns to derive a new isoprene emissions estimate. Daily modelled values averaged between 13:00-14:00 LT are used to match the overpass time of the Aura satellite. Then the slope is calculated using the reduced major axis (RMA) regression between the a priori isoprene emissions (those from GEOS-Chem, E_{GC}) and tropospheric HCHO columns in each model grid box each month. There is very little HCHO above the tropopause, so differences between total and tropospheric column are negligible. In this work I refer to both total and tropospheric column HCHO using Ω .

3.2.1 Outline

This section provides an overview of the steps involved in creating a top-down emissions estimate. This process is summarised in Figure 3.1.

1. Corrected vertical columns (Ω_{OMI} ; saved in the OMHCHORP dataset) are calculated (see Section 2.6) using level two OMI HCHO satellite data (see Section 2.3), along with GEOS-Chem model runs (see Section 2.4.7). Satellite columns are binned into both $0.25^\circ \times 0.3125^\circ$ and $2^\circ \times 2.5^\circ$ horizontal resolutions. In this step model background values (columns over the remote pacific) are used to correct the vertical columns, which is explained in Section 2.6.5.

2. Level three satellite data are used to make anthropogenic, fire, and smoke influence masks (see Section 2.7). These are applied to remove Ω_{OMI} that may be influenced by pyrogenic or anthropogenic sources.
3. A mask is created showing where the HCHO production is not dominated by local isoprene emissions. This is determined by calculating smearing over Australia using two model runs with differing isoprene emissions. The smearing value is determined as $\hat{S} = \Delta\Omega_{GC}/\Delta E_{OMI}$: the ratio of the differences between model runs of HCHO columns and isoprene emissions. The acceptable range for \hat{S} over Australia is determined as 800 - 4600 s. A full description of the creation of this smearing filter is given in Section 3.2.7.
4. GEOS-Chem modelled biogenic emissions of isoprene (E_{GC}) along with biogenic columns of HCHO (Ω_{GC}), both averaged over $2^\circ \times 2.5^\circ$ horizontally and 13:00-14:00 LT temporally, are used to calculate a reduced major axis linear regression slope ($\Omega_{GC} = S \times E_{GC} + \Omega_{GC,0}$). Calculation of this modelled slope is explained in Section 3.2.4.
5. Satellite HCHO Ω_{OMI} and S then form the basis for top-down estimate of biogenic isoprene emissions (E_{OMI} atoms C $\text{cm}^{-2} \text{s}^{-1}$). This product is our a posteriori, and calculation details are given in Section 3.2.6.
6. A posteriori top-down emissions E_{OMI} are compared against a priori emissions, and analysed in conjunction with independent observations from in-situ measurement (MUMBA, and SPS). Results are examined in Section 3.3.
7. GEOS-Chem is run using the a posteriori emissions (see Section 3.2.8), and HCHO, O₃, isoprene, and NO_x outputs are compared to campaign and satellite measurements where possible (Section TODO).

3.2.2 Masks and reprocessed satellite HCHO

Satellite data pixels are read from OMHCHO, the level 2 OMI HCHO dataset, AMFs are recalculated, and then pixels are gridded daily into $0.25^\circ \times 0.3125^\circ$ horizontal bins. This forms the intermediate product OMHCHORP, which is fully described in Section 2.6.1. This dataset includes gridded satellite HCHO columns (Ω_{OMI}), along with pixel counts (how many satellite datapoints were used for each grid box) to allow averaging, re-binning, and uncertainty analysis. In this thesis I use the OMI product as it has better temporal coverage and increased pixel counts over Australia when compared to GOME or GOME-2 (on board the ERS-2 and METOP-A satellites respectively).

In order to determine biogenic HCHO enhancements from Ω_{OMI} , we require filters for non-biogenic sources. These masks are described in Section 2.7, and a brief recap is provided here. While one source of HCHO production is methane oxidation, the linear regression used to estimate isoprene emissions effectively ignores this source as part of the background, which means a methane filter is not required. Anthropogenic, pyrogenic, and smoke influence masks are created from three satellite products: NO₂ from OMNO2d, fire counts from MOD14A1, and AAOD from OMAERUVd respectively.

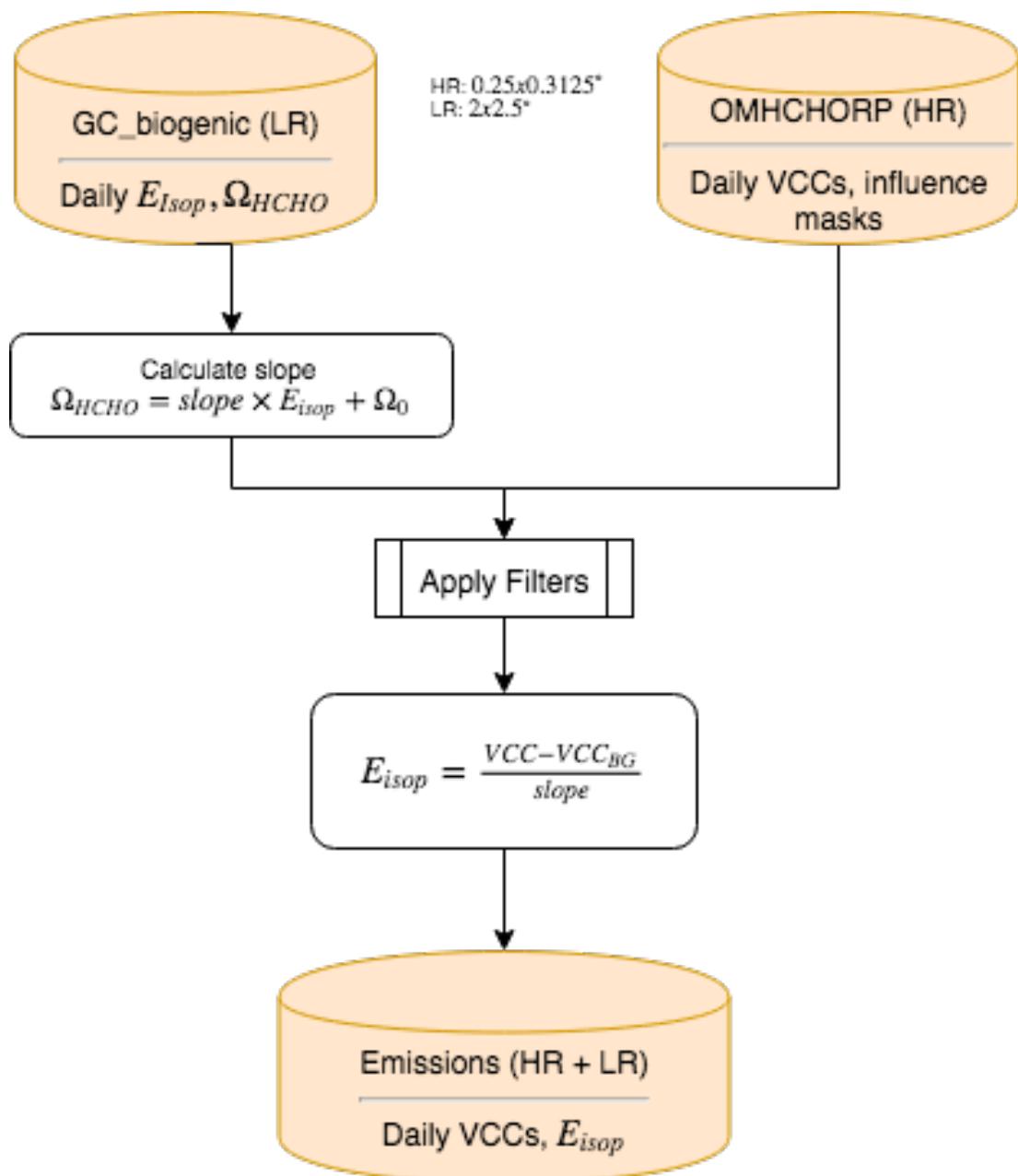


FIGURE 3.1: Top-down isoprene emissions estimate formation using OMHCHORP and biogenic GEOS-Chem outputs.

1. The fire mask is created daily using non-zero (MODIS) fire counts over the prior 2 days that occur in local or adjacent grid squares at $0.25^\circ \times 0.3125^\circ$ horizontal resolution.
2. Influence from transported smoke plumes is removed where OMI aerosol absorption optical depth (AAOD, from OMAERUVd) is greater than 0.03.
3. A filter for anthropogenic influence is created daily using OMNO2d NO₂ tropospheric column amounts; masking any grid squares with greater than 2×10^{15} molec cm⁻² on any particular day, along with grid squares where the yearly average is above 1.5×10^{15} molec cm⁻².

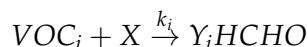
The recalculated corrected vertical columns are saved to OMHCHORP dataset both before and after applying the filters, to allow filter analysis.

3.2.3 GEOS-Chem emissions

In this work MEGAN (version 2.1) is run as a module within GEOS-Chem (version 10.01). The chemical model is coupled to a meteorological model driven by GEOS-5 meteorological fields at $0.25^\circ \times 0.3125^\circ$ horizontal resolution. GEOS-Chem output is averaged onto 47 or 72 vertical levels at $2^\circ \times 2.5^\circ$, based on chemistry and transport calculated every 30 and 15 minutes respectively. Isoprene emissions from the default *tropchem* simulation are referred to as the a priori emissions, when shown as part of a formula the a priori are denoted as E_{GC} .

3.2.4 Relationship between isoprene emissions and formaldehyde

Tropospheric HCHO production is primarily due to the oxidation of VOC precursor species (VOC_i). Background concentrations are driven by methane; a longer lived (~ 8 yr) VOC. Over continental land masses, the variability in HCHO is driven by shorter lived precursor emissions (Chance2000; Palmer2003). The intermediate steps are considered negligible as HCHO is produced quickly from short-lived intermediates:



where X is an oxidant, Y_i is HCHO yield (per C atom in VOC_i), and k_i is the reaction rate constant. In specific conditions described below, HCHO total columns (Ω ; molec cm⁻²) can be linearly related to isoprene emissions.

The isoprene to HCHO relationship is derived using several assumptions that are important to understand. The first assumption is that HCHO is at steady state, which implies production (P_{HCHO}) and loss (L_{HCHO}) are equivalent:

$$\frac{d\Omega}{dt} = 0 = P_{HCHO} - L_{HCHO} \quad (3.2)$$

This is reasonable during midday when isoprene emissions are steady and Ω has had time to stabilise. The second assumption is that loss (k_{HCHO}) is only first order, such

as from photolysis and oxidation:

$$L_{HCHO} = k_{HCHO}\Omega \quad (3.3)$$

This assumption means that loss due to transport must be negligible as it is not first order. This assumption is reasonable for large enough grid boxes as transport becomes negligible relative to the linear (first order) terms. Production and loss are on the order of minutes, and grid box sizes in this work are rectangular with ~ 200 km edge lengths. Monthly averaged wind speeds rarely exceed 20 km hr $^{-1}$ over Australia, meaning HCHO and precursor transport remain minor. transport can still be an issue, and is handled in Section 3.2.7.

These assumptions about Ω production above the background level is due only to precursor emissions (E_i ; atoms C cm $^{-2}$ s $^{-1}$) multiplied by their yields to HCHO (Y_i):

$$P_{HCHO} = \sum_i Y_i E_i \quad (3.4)$$

By combining Equations 3.2, 3.3, and 3.4, we can relate Ω to precursor emissions:

$$\begin{aligned} k_{HCHO}\Omega &= \sum_i Y_i E_i \\ \Omega &= \frac{1}{k_{HCHO}} \sum_i Y_i E_i \end{aligned} \quad (3.5)$$

Finally, we assume isoprene emissions are driving changes in Ω (Palmer2003; Millet2008; Marais2014; Stavrakou2015) and lump other terms together:

$$\sum_i Y_i E_i = Y_{isop} E_{isop} + \sum_{i \neq isop} Y_i VOC_i \quad (3.6)$$

This assumption is reasonable only over continental land masses, and only if pyrogenic and anthropogenic HCHO precursors are accounted for. The linear relationship between isoprene emissions and Ω is determined by equating P_{HCHO} and L_{HCHO} from Equations 3.4 and 3.3, plugging in Equation 3.6, and assuming that the lumped terms make up the background:

$$\begin{aligned} k_{HCHO}\Omega &= Y_{isop} E_{isop} + \sum_{i \neq isop} Y_i VOC_i \\ \Omega &= \frac{Y_{isop}}{k_{HCHO}} E_{isop} + \Omega_0 \\ &= S \times E_{isop} + \Omega_0 \end{aligned} \quad (3.7)$$

Here S is the slope: $S \equiv \frac{Y_{isop}}{k_{HCHO}}$. This assumption can be false when pyrogenic or anthropogenic emissions influence the HCHO column, however these scenarios are filtered using independent satellite measurements (see Section 2.7).

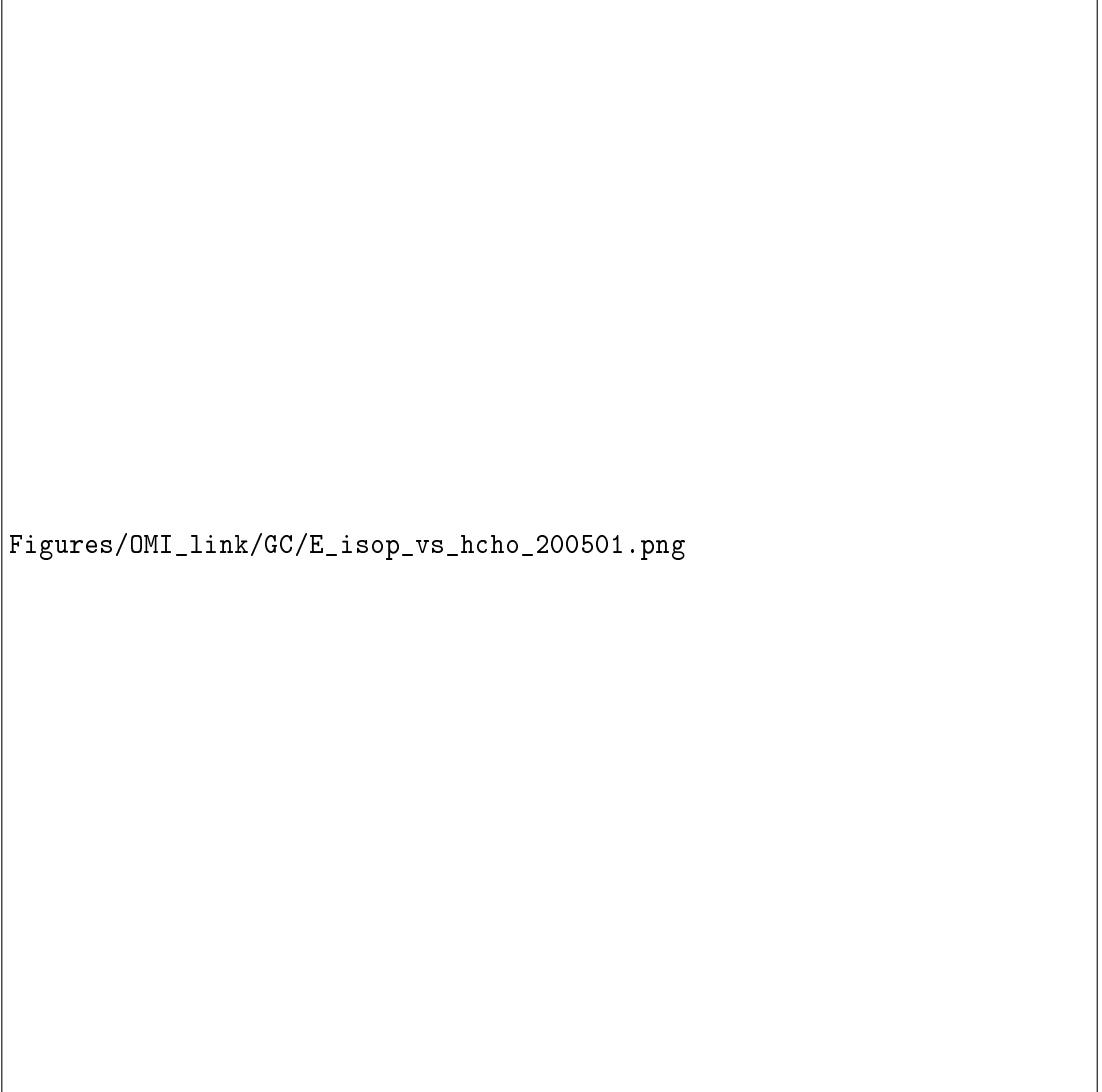
3.2.5 Calculation of modelled slope

To determine S , the link between biogenic isoprene and midday column HCHO, we use GEOS-Chem. The term E_{GC} is used when discussing isoprene emissions estimated within GEOS-Chem and Ω_{GC} is used to represent simulated HCHO. The method to calculate S using GEOS-Chem follows roughly the following three stages:

1. Hourly gridded model output E_{GC} (atoms C $\text{cm}^{-2} \text{s}^{-1}$) at 13:00 LT daily are extracted, along with Ω_{GC} (molec cm^{-2}) output.
2. Filtering removes gridded output on days where grid squares are likely to be affected by smearing (see Section 3.2.7).
3. A reduced major axis regression slope is determined between Ω_{GC} and E_{GC} using a month of modelled output (one value per day) for each grid square (e.g., see Figure 3.2)

Each $2^\circ \times 2.5^\circ$ grid box from daily GEOS-Chem (biogenic only) output of $\Omega_{HCHO} \equiv \Omega_{GC}$ and E_{isop} within Australia, and calculate the RMA slope monthly from January 2005 to December 2012. Modelled background concentrations can be ignored here as they do not affect slope calculation. This effectively provides monthly gridded slope (S) between biogenic isoprene emissions and HCHO columns, in units of seconds. Figure 3.2 (top left) shows how S varies spatially over Australia for an example mid-summer month. Some areas can be seen to be very sensitive to emissions, such as the west coast and Eyre basin, which is likely due to the low precursor and HCHO levels in those areas. The regression coefficients also vary spatially (bottom left), and some areas show little correlation, likely due to weather, transport, and a lack of local emission sources. The slopes shown in the bottom right panel show a small sample of scatter and regression plots. These can range widely due to differences in emission and yield parameters, which plays a role in the smearing filters applied in Section 3.2.7. Due to the $2^\circ \times 2.5^\circ$ horizontal resolution of GEOS-Chem, calculations over coastal grid boxes that are mostly oceanic are often discarded as the change in HCHO is not dominated by emissions of isoprene, as is assumed for equation 3.7.

One issue with slope calculation is potential transport (also known as smearing), either of isoprene transported in from outside the local grid box (before any HCHO is formed), or of HCHO formed by local emissions but transported out of the local grid box. The effects of this are dealt with using a smearing filter (see detailed discussion in Section 3.2.7). Days where we expect smearing may be affecting local levels of HCHO are removed before calculating S , and a quick analysis is performed on how the filter affects monthly slope, correlation, and uncertainty. Figure 3.3 shows the calculated slope for 2005-2012, along with its 95% confidence interval over Sydney. The monthly and multi-year monthly averages are shown before and after the smear filter is applied. The filter slightly reduces the amplitude of the seasonal cycle, raising the January minimum and lowering the June and July maximum. Filtering slightly improves the correlation coefficient throughout the year. Surprisingly, more data is filtered in summer. I believe this is due to higher biogenic isoprene emissions over summer, making transport more noticeable on windy days. Essentially the smearing signal is stronger in summer. Additionally the anthropogenic signal on HCHO of



Figures/OMI_link/GC/E_isop_vs_hcho_200501.png

FIGURE 3.2: Top left: RMA slope between modelled tropospheric column HCHO and isoprene emissions (E_{GC}) using midday (13:00–14:00 LT) values over for January 2005, per grid square at $2^\circ \times 2.5^\circ$ horizontal resolution. Top right: Australia-wide average of midday emissions and tropospheric columns. Bottom left: Squared RMA correlation coefficient for regression in top left. Coloured dots correspond to colour of regressions shown in bottom right panel. Bottom right: Sample of correlations from four grid squares.

non-isoprene precursor emissions from Sydney may lower the efficacy of the smearing filter over winter, which is based on altering modelled isoprene emissions. This plot has been repeated for several grid squares over Australia (not shown). When calculating top-down emissions the smearing filtered slope (S) is used for each grid square month. The multiple year monthly averaged slope is used instead when the regression coefficient (r) is less than 0.4, or then number of data points used in the regression (n) is less than 10. When r for the multiple year slope is also lower than 0.4 (does not happen in the example grid square), no estimation is performed.

There are two simple ways to determine the modelled background HCHO, one of which involves running the model with no isoprene emissions. Since we have assumed variation in HCHO columns only depends on isoprene emissions, our background term is theoretically identical to this simulated HCHO. The other method uses HCHO over the remote Pacific at matching latitudes and times, which emulates how the background is determined for the satellite measured HCHO. Figure 3.4 shows GEOS-Chem total column HCHO with and without isoprene emissions along with amounts over the remote Pacific at the same latitudes. The difference in Ω_{GC} over Australia with no isoprene emissions, and Ω_{GC} over the remote pacific (bottom right panel) shows the difference (~ 1 to 3 molec cm^{-2}) between these methods in an example averaged month (January, 2005). This difference is relatively small, and may be due to non-isoprene HCHO precursors. For consistency with the satellite data, we determine backgrounds using the remote Pacific. Background HCHO for any latitude in this thesis is calculated by averaging longitudinally (140°W to 160°W) the matching latitudes over the remote Pacific.

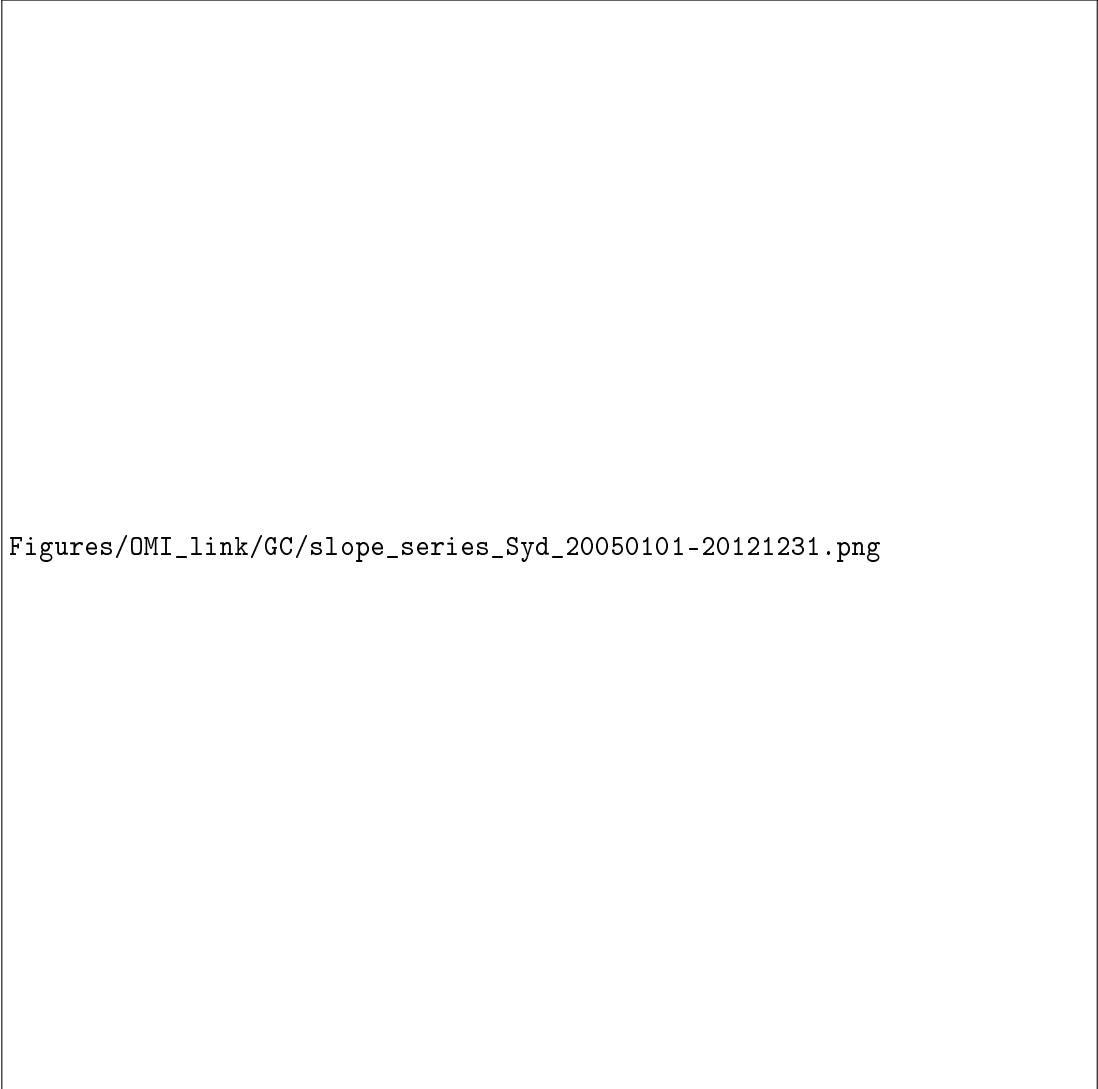
3.2.6 Calculation of Emissions

Top-down emissions estimates are calculated using OMHCHO (see Section 2.3) slant columns and an updated AMF calculated using code by Paul Palmer and Randal Martin, with modifications by Luke Surl (see Section 2.6.3.2). These emissions are referred to as the a posteriori from here onward, or E_{OMI} in formulae.

A posteriori emissions are calculated using the linear relationship described in Section 3.2.4 using the modelled slope S calculated in the prior section and satellite HCHO columns recalculated in 2.6:

$$\begin{aligned}\Omega_{OMI} &= S \times E_{OMI} + \Omega_0 \\ E_{OMI} &= \frac{\Omega_{OMI} - \Omega_0}{S}\end{aligned}\tag{3.8}$$

This is the same as equation 3.7, except now we use the satellite HCHO (Ω_{OMI} , and Ω_0). Ω_0 is calculated using Ω_{OMI} in the remote Pacific averaged monthly and longitudinally, for each latitude. This leaves E_{OMI} as the only unknown once the satellite measurements are processed to match the temporal and horizontal resolution of S . Figure 3.5 shows an example of how the a priori compares to the a posteriori, averaged over January, 2005. This figure gives a single month of output as an example. The a priori exceeds 500% of the a posteriori in many regions, however this is mostly in regions of low emissions and represents only minor absolute differences. Analysis of the full record is discussed in the results (Section 3.3).



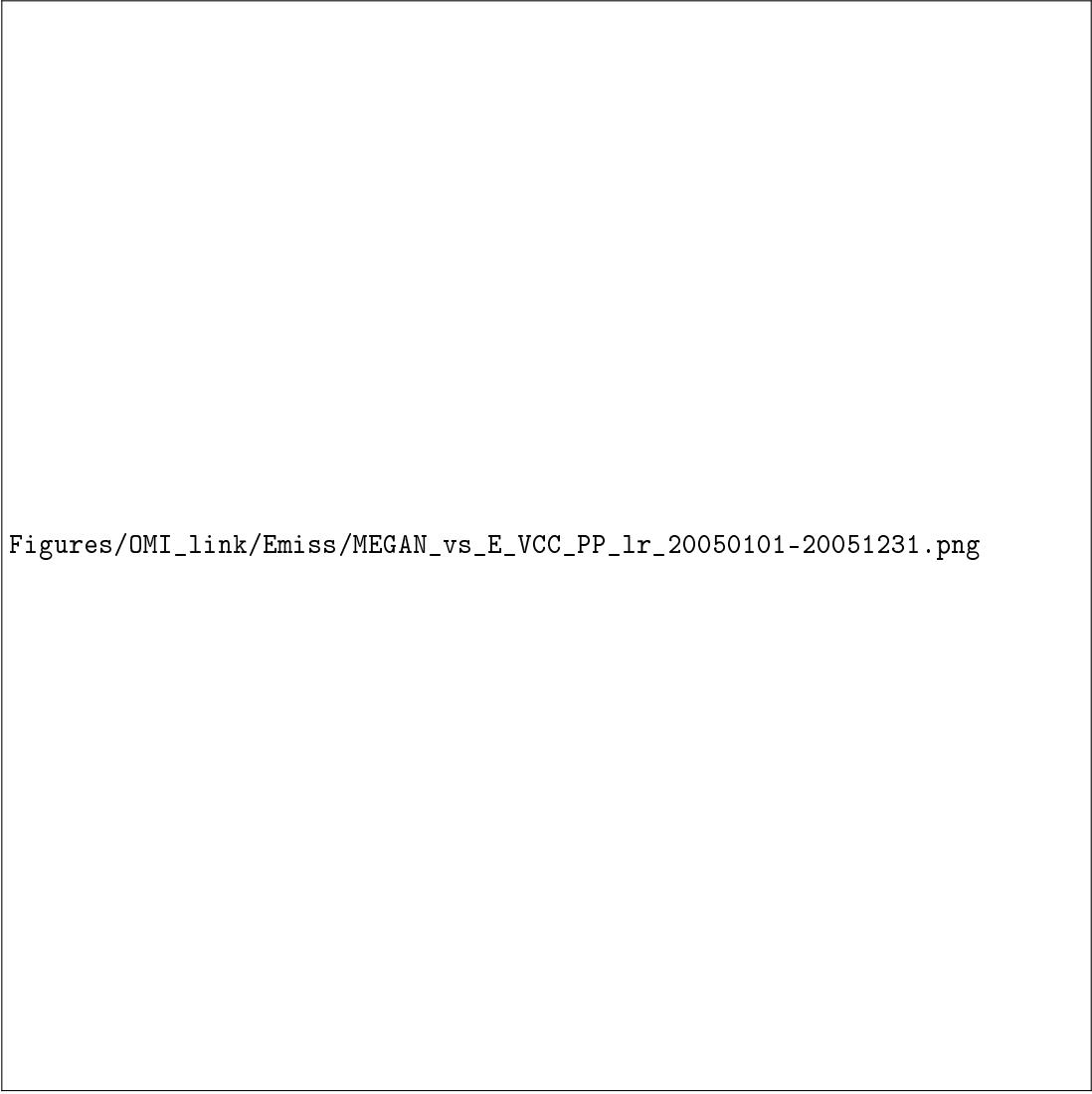
Figures/OMI_link/GC/slope_series_Syd_20050101-20121231.png

FIGURE 3.3: Row 1: monthly slope along with 95% confidence interval both before (left) and after (right) applying the smear filter for the model grid square containing Sydney over 2005-2012. Limits used in creation of the smear filter are shown with dashed lines. Row 2: regression coefficient and data-point counts for slope shown in row 1. Additionally limits for r and n used in slope utilisation (see text) are shown with dashed lines. Row 3: slope and confidence interval using the multi-year dataset for each month. Row 4: regression coefficient and data-point counts for row 3.



Figures/OMI_link/GC/GC_background_hcho_200501.png

FIGURE 3.4: Top left: Total column HCHO over Australia using the standard tropchem GEOS-Chem simulation. Top right: As top left except over the remote Pacific region at southern mid-latitudes. Bottom left: As top left except using the no isoprene emissions GEOS-Chem simulation. Bottom right: Difference between the no isoprene emission HCHO columns over Australia, and the remote Pacific HCHO columns from the standard tropchem run at matching latitudes.



Figures/OMI_link/Emiss/MEGAN_vs_E_VCC_PP_lr_20050101-20051231.png

FIGURE 3.5: Top row: isoprene emissions from GEOS-Chem (a priori, left) simulation and top-down (a posteriori, right) calculations averaged over the month of January, 2005. Bottom row: the absolute (left) and relative (right) differences between the two.

One potential issue in this top down estimation technique is the low number of valid satellite measurements that may occur due to the higher zenith angles in winter and at higher latitudes. When calculating the a posteriori from our modelled slope, negative emissions result wherever measured columns are lower than background amounts (as $E_{OMI} = \frac{\Omega_{OMI} - \Omega_0}{S}$). These are set to zero, which increases monthly averages by TODO: $\sim X - Y\%$ ($\sim XX - YY$ atom C cm $^{-2}$ s $^{-1}$) over Australia, with the highest increases occurring during (TODO winter?).

3.2.7 Accounting for smearing

In high NO_x ($>\sim 1$ ppb) environments, isoprene has a lifetime on the order of 30 minutes, and HCHO can be used to map isoprene emissions with spatial resolution from 10-100 kms (Palmer2003). In low NO_x conditions, isoprene has a longer lifetime (hours) and may form HCHO further from the source area (Fan2004; Liu2016a; Liu2017_hpald). Over Australia, NO_x levels are generally low and smearing is therefore expected to be important. Smearing limits the horizontal resolution of the linear top-down inversion process, as a finer resolution increases sensitivity to transport. Horizontal transport *smears* the HCHO signal so that its source location would need to be calculated using wind speeds and loss rates (Palmer2001; Palmer2003). Smearing is a measure of how much HCHO in a given grid box was produced from isoprene emitted in a different (upwind) grid box. Smearing affects emissions estimates as HCHO enhancements downwind of where precursor emissions occurred lead to misinterpretation of local emissions. Smearing affected grid squares are filtered out prior to application of Equation 3.7.

3.2.7.1 Calculation of smearing

Smearing has been analysed in several publications (Martin2003; Palmer2003; Millet2006; Stavrakou2009; Marais2012; Barkley2013; Zhu2014; Wolfe2016; Surl2018), and is often calculated using the method used in this thesis, as first described in Palmer2003. This involves using two model runs, one of which has isoprene emissions scaled globally by a constant (generally from 0.5 to 2). From Section 3.2.4, Equation 3.7 states that the modelled slope (S) is the yield of HCHO per C of emitted isoprene divided by HCHO loss rate ($S = \frac{Y_{isop}}{k_{HCHO}}$). Using two runs of GEOS-Chem with differing isoprene emissions but otherwise identical we have:

$$\begin{aligned} Run_1 : \Omega_{HCHO} &= SE_{isop} + \Omega_0 \\ Run_2 : \Omega'_{HCHO} &= S'E'_{isop} + \Omega'_0 \end{aligned} \quad (3.9)$$

There are several assumptions that need to be understood, as these are what is tested by the smearing calculation. The initial assumption is that the system is at steady state, with no transport of isoprene affecting HCHO columns, this is the basis for equations 3.9. It is also assumed that background values (Ω_0) are from oxidation of methane and other long lived VOCs, so that $\Omega_0 = \Omega'_0$. Between these two runs we are only changing the E term. Chemistry is unchanged so that the yield and loss rate should

TABLE 3.1: Smearing filters or typical slopes (S) from literature.

Publication	min. (s)	max. (s)	type ^a	Region
Palmer2003	1270	2090	Range	North America ^b
Marais2012		4000	Limits	Africa
Barkley2013^c	1300	1800	Limits	South America
Surl2018	2200	4900	Range	India
In this Thesis	800	4600	Limits	Australia

a: Slope *ranges* are observed or modelled S , while smearing *limits* are the applied acceptable limits for S .

b: Slope range for summer only.

c: Assumed HCHO lifetime of 2.5 hours implies yields from 0.14 to 0.2 per C, consistent with box modelling.

not change between the two runs:

$$S = S' = \frac{Y_{isop}}{k_{HCHO}} \quad (3.10)$$

Equations 3.9 may then be combined as follows:

$$\begin{aligned} Run_1 - Run_2 : \Omega_{HCHO} - \Omega'_{HCHO} &= SE_{isop} - S'E'_{isop} + \Omega_0 - \Omega'_0 \\ \Omega_{HCHO} - \Omega'_{HCHO} &= S(E_{isop} - E'_{isop}) \\ \Delta\Omega_{HCHO} &= S\Delta E_{isop} \\ \hat{S} \equiv & \frac{\Delta\Omega_{HCHO}}{\Delta E_{isop}} \approx \frac{Y_{isop}}{k_{HCHO}} \end{aligned} \quad (3.11)$$

This allows the combination of outputs from the two runs to determine where \hat{S} diverges from expected values for S .

Potential smearing is masked by checking thresholds a daily modelled value for $\hat{S} \approx Y_{isop}/k_{HCHO}$ against thresholds. By assuming that midday HCHO lifetime ($\tau = 1/k_{HCHO}$) typically falls within 1.5 to 4 hrs (as seen in the USA (Palmer2006; Wolfe2016)) and isoprene to HCHO yield (HCHO per isoprene carbon emitted) lies within the range 0.2 to 0.4 (scenarios estimated in Palmer2003), one can set a simple bound on \hat{S} of $[0.2 \times 1.5, 0.4 \times 4]$ hrs or 1080 to 5760 seconds. As NO_x levels across Australia are relatively low, and lower NO_x levels reduce the prompt yield (Palmer2003; Wolfe2016), I reduce the threshold range by 20% and round to the nearest hundred leading for bounding range of 800 to 4600 for \hat{S} . This range strikes a balance between unlikely modelled yields and how much data is lost to filtering. Table 3.1 compares the smearing filter for Australia used in this thesis to typical slopes used in previous work for other regions.

Isoprene reaction chains are diverse, with many branches forming HCHO. HCHO production yields are often classed into two categories: first generation HCHO yield

and total (or molar) yield. First generation yield refers to the amount of HCHO produced per unit isoprene consumed by initial oxidation. Total yield refers to time dependent yield of HCHO over multiple oxidation stages (**Wolfe2016**). In this work yield (Y_{isop}) is not calculated as the slope ($S = Y_{isop}/k_{HCHO}$) can be used instead.

By assuming yield Y_{isop} lies between 0.2 and 0.4, we find a range for midday lifetimes of HCHO using equation 3.7:

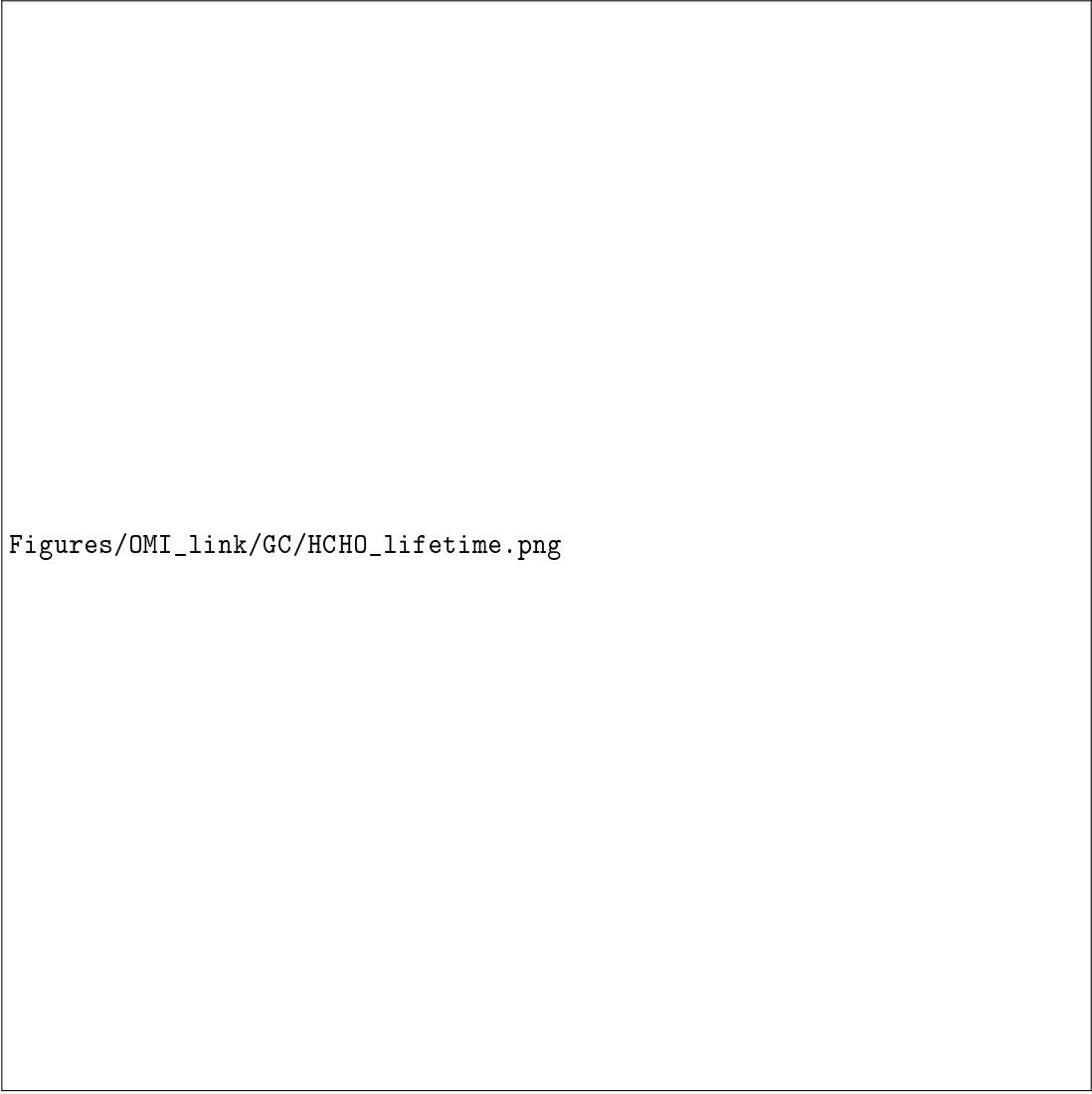
$$\begin{aligned} S &= \frac{Y_{isop}}{k_{HCHO}} \\ \tau &\equiv \frac{1}{k_{HCHO}} \\ \tau &= \frac{S}{Y_{isop}} \end{aligned}$$

τ is heavily influenced by assumed yield, and improved methods of estimating yield over Australia are required to improve this estimate. Figure 3.6: shows the GEOS-Chem HCHO lifetime estimated throughout the year. There is a clear seasonal cycle with longer lifetimes in winter months. A clear June (and sometimes March, July and August) increase in HCHO lifetimes is shown, which is caused by the reduced winter HCHO concentrations, temperature, and insolation. These factors may limit the utility of any top-down emissions estimation technique using HCHO in the winter months. The figure is produced using filtered slope information from 2005, and outliers along with low data availability in some months is an issue. Noise in the southwest and middle regions may be indicative of heavy filtering, potentially driven by westerly winds which can bring transported pollution and also lead to smearing.

To determine which model grid boxes are affected by smearing, we follow **Marais2012**. GEOS-Chem is run with and without isoprene emissions halved, then Equation 3.11 ($\hat{S} = \frac{\Delta\Omega_{HCHO}}{\Delta E_{isop}}$) provides \hat{S} . Here Δ represents the difference (daily 1300-1400 LT) between default and scaled runs. If \hat{S} sits outside the 800-4600 range then we remove that grid square day from both S and subsequent a posteriori calculations. A relatively large change in Ω_{HCHO} compared to local emissions ($\hat{S} > 4600$) suggests HCHO production is from non-local isoprene emissions. Alternatively, a relatively low value of \hat{S} ($\hat{S} < 800$) suggests emissions from the local grid square are being exported before they form HCHO.

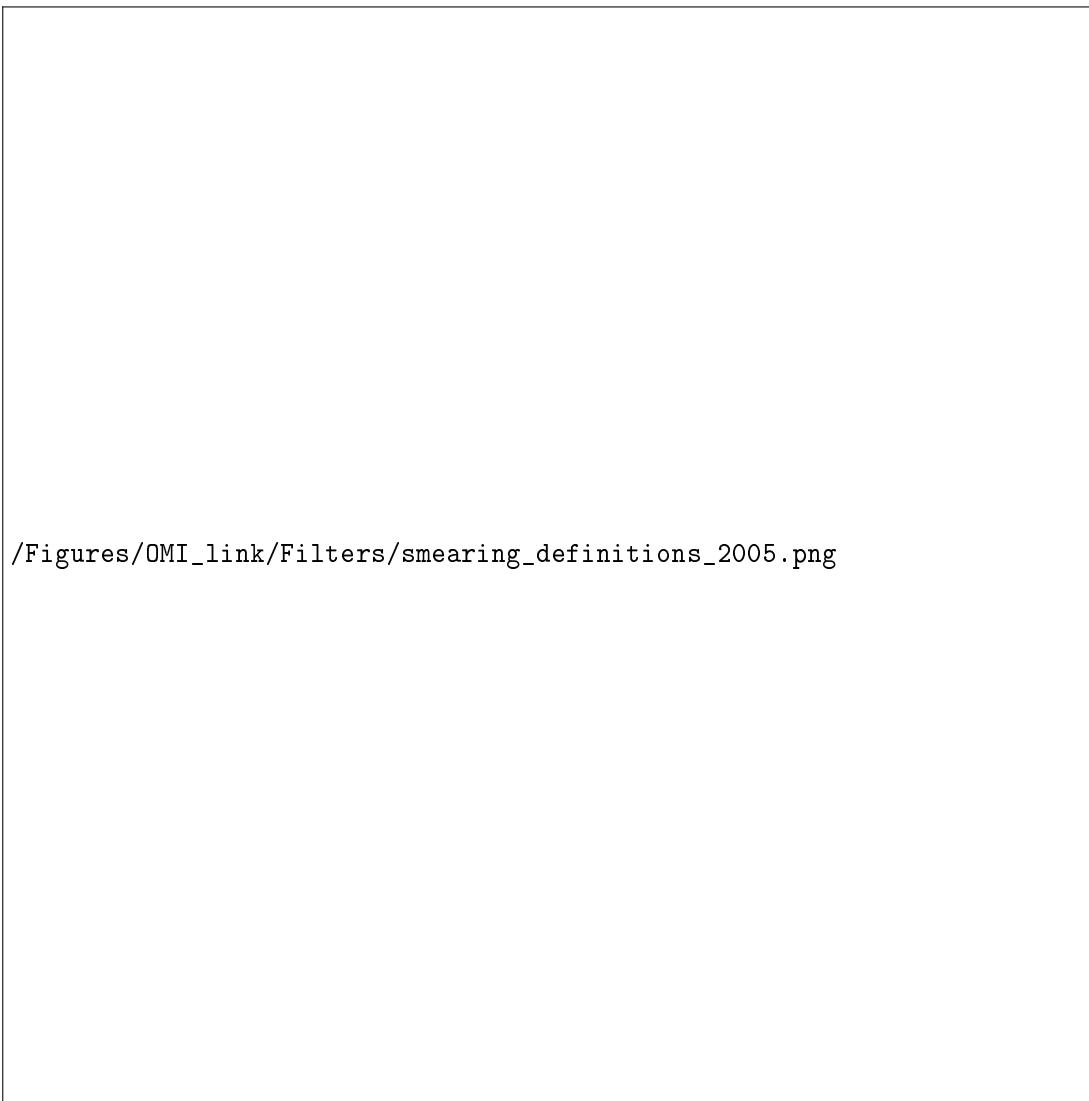
3.2.7.2 Sensitivity to smearing

Smearing can be dependent on local or regional weather patterns, as greater wind speeds will reduce the time any emitted compound stays within the local grid box. As such smearing can vary greatly both spatially and temporally. Smearing is also sensitive to time of day, season, and latitude, as higher lower insolation results in slower photolysis. Figure 3.7 shows smearing in summer and winter. The smearing filter is more active in winter, especially at higher latitudes. The filter removes more data on coastal grid squares as they are more affected by winds and transport. Additionally, any grid square with low isoprene emissions will be more sensitive to transport, as the signal is lower. During summer data-loss from smearing is only minor (todo: Y%); however, data-loss peaks in winter (TODO X %), especially in higher southern latitudes.



Figures/OMI_link/GC/HCHO_lifetime.png

FIGURE 3.6: Monthly area averaged HCHO lifetime (τ in hours), with IQR shaded. Solid lines show lifetime assuming yield is 0.2, and 0.4 (higher and lower lines respectively). Coloured by regions shown in Figure 3.12.



/Figures/OMI_link/Filters/smearing_definitions_2005.png

FIGURE 3.7: Seasonally averaged smearing (\hat{S} , see text) in summer (DJF) and winter (JJA) from 2005. Diamonds represent grid squares which have had at least 1 day removed due to the smearing filter, and red crosses show where more than 45 days ($\sim 50\%$) have been removed.

When limiting smearing (\hat{S}) to within 800-4600 s, GEOS-Chem correlations between isoprene emissions and HCHO columns improve marginally and not uniformly (Figure 3.3). Where smearing is prevalent, the relationship between a priori emissions and HCHO columns may already be weak due to low actual emissions or unsuitable meteorological conditions. Loss of data due to filtering is handled by using multiple years of data for any affected grid square month as follows. S and associated regression coefficients (r) are calculated monthly. If $r < 0.4$ then the regression is calculated using multiple years of data for that month. This multiple year regression is discarded completely if it also has a coefficient of $r < 0.4$, leaving no slope for the grid square for the month.

3.2.7.3 Smearing length scale

The expected horizontal transport (prior to reaction) of a precursor can be calculated using the smearing length (Palmer2003). The distance travelled (L) downwind by a precursor (i) before forming HCHO can be estimated through:

$$L_i = \frac{U}{k_i - k_{HCHO}} \ln \left(\frac{k_i}{k_{HCHO}} \right)$$

where U is wind-speed. Palmer2003 further define a smearing length scale $L_{s,i}$ as the distance downwind where most of the precursor is completely transformed into HCHO. Here we use the simplification of $L_{s,isop} \approx \frac{U}{k_{HCHO}}$ as isoprene loss rates (2 h^{-1}) exceed those of HCHO ($0.25\text{--}0.67 \text{ h}^{-1}$) by a factor of 3-8. These numbers come from assuming HCHO lifetime of 1.5-4 h, and isoprene lifetime of 30 m TODO: cite. Throughout most of the year, over most of Australia, monthly averaged wind speeds do not exceed 20 km h^{-1} (http://www.bom.gov.au/jsp/ncc/climate_averages/wind-velocity/ [accessed Feb., 2019]), although daily wind speeds will be highly variable. This means that a reasonable upper limit for the smearing length scale is $20/0.25 = 80 \text{ km}$ over Australia. Grid boxes used in top-down estimation of isoprene emissions describe rectangles with both side lengths approximately 200 km, so generally I do not expect to be overly impacted by smearing. The estimated loss rate of HCHO in GEOS-Chem is up to three times higher in summer and along the north and eastern regions associated with denser forest regions, when compared against other regions. This is largely due to loss rates being proportional to concentrations, and leads to less smearing sensitivity over areas of high isoprene emissions and HCHO concentrations.

GEOS-Chem daily averaged HCHO lifetime (τ) is calculated using daily averaged surface loss rates (L_{HCHO}) and concentrations of HCHO:

$$\tau = \frac{[HCHO]}{L_{HCHO}}$$

The expected lifetime of HCHO is determined by assuming loss is linear (first order) and dividing grid box daily averaged concentrations of GEOS-Chem HCHO ($[HCHO]$ in molecules cm^{-3}) by their modelled losses (L_{HCHO} in molecules $\text{cm}^{-3} \text{ s}^{-1}$). For each grid square over Australia this daily averaged surface lifetime in summer (Jan., Feb.) and winter (Jul., Aug.) is shown in Figure 3.8. Additionally lifetimes coloured by location (dots in the top right panel) are shown over time in the bottom panel. Note

that this figure shows the daily averaged lifetime, as loss rate diagnostics are only available as a daily average. Loss rates would maximise at midday (13:00-14:00 LT), along with HCHO concentrations, which makes the estimate shown here an upper limit. This figure highlights the seasonal nature of HCHO lifetime, although midday numbers are expected to have less seasonality than shown here, since midday lifetime will be less affected by how long the daylight lasts. Another highlighted issue is the potential latitudinal dependence of HCHO lifetimes, since there is less total insolation leading to lower HCHO loss rates at higher latitudes. The overall takeaway is that the accuracy and utility of any top down HCHO precursor estimation technique will be limited by both season and potentially latitude. These limitations are due to both data availability (as satellite HCHO uncertainty increases at high zenith angles) and spatial smearing (due to HCHO lifetimes that increase with reduced insolation and temperature).

3.2.7.4 NO_x dependence

NO_x concentration directly affects the fate of VOCs in the atmosphere, influencing HCHO production by isoprene. In low NO_x environments, reported HCHO yields from isoprene are around 0.2 - 0.3 C per C (or 100-150 molar %), while in high NO_x environments this value becomes two to three times higher (**Palmer2003; Wolfe2016**). Some values for HCHO yield from prior literature are shown in Table 3.2.

The effect of NO₂ on smearing can be seen in Figure 3.9. This plot shows how smearing over Australia compares to satellite NO₂, with smearing distributions binned by NO₂ both with and without filtering for smearing. At lower NO₂ levels the smearing is often 2-4 orders of magnitude above the upper threshold. This abruptly decreases at around 5×10^{14} molec cm⁻² NO₂. There is also a higher number of data points below the lower threshold before that same NO₂ level, showing that transport is a bigger issue at $\text{NO}_2 < 5 \times 10^{14}$ molec cm⁻². Due to the coarse resolution of the model, many regions with low NO₂ also have low isoprene emission levels, and this filtering does not remove too many data points from biogenic-dominated regions.

3.2.8 Running GEOS-Chem using a posteriori emissions

After creating our a posteriori isoprene emissions estimate, we run GEOS-Chem again with biogenic emissions scaled to match the new estimate. This is performed by applying a seasonal scaling factor α , based on the multi-year monthly average difference between midday a priori and a posteriori emissions at $2^\circ \times 2.5^\circ$ horizontal resolution. α is the ratio between the multi-year averaged monthly emissions from GEOS-Chem E_{GC} and the a posteriori E_{OMI} :

$$\alpha = \frac{E_{OMI}}{E_{GC}} \quad (3.12)$$

This seasonal scaling retains shorter time-scale variability and meteorological dependencies, while ensuring the multi-year monthly averaged emissions match. Initially α is uniformly set to 1 globally. Where top down emissions exist and E_{GC} is non-zero, we can set α using Equation 3.12. α is applied through the emissions module in GEOS-Chem where isoprene emissions are calculated. First all the new midday (13:00-14:00 LT) emissions (per grid box) are combined forming a multi-year monthly mean,



/Figures/OMI_link/Filters/tau_2005.png

FIGURE 3.8: Top left, right: summer (Jan., Feb.) and winter (JJA) averaged daily surface HCHO lifetime (τ). Bottom panel: τ over the year, coloured by location (see dots in top right panel).

TABLE 3.2: Isoprene to HCHO yields and lifetime.

HCHO Yield (molar %)	Lifetime	NO _x background	Source
315±50		High	a
285±30		High	a
225	35 min	High	b
150		Low	b
150		Low	d
450		High	d
235		1 ppbv	e
150		0.1 ppbv	e

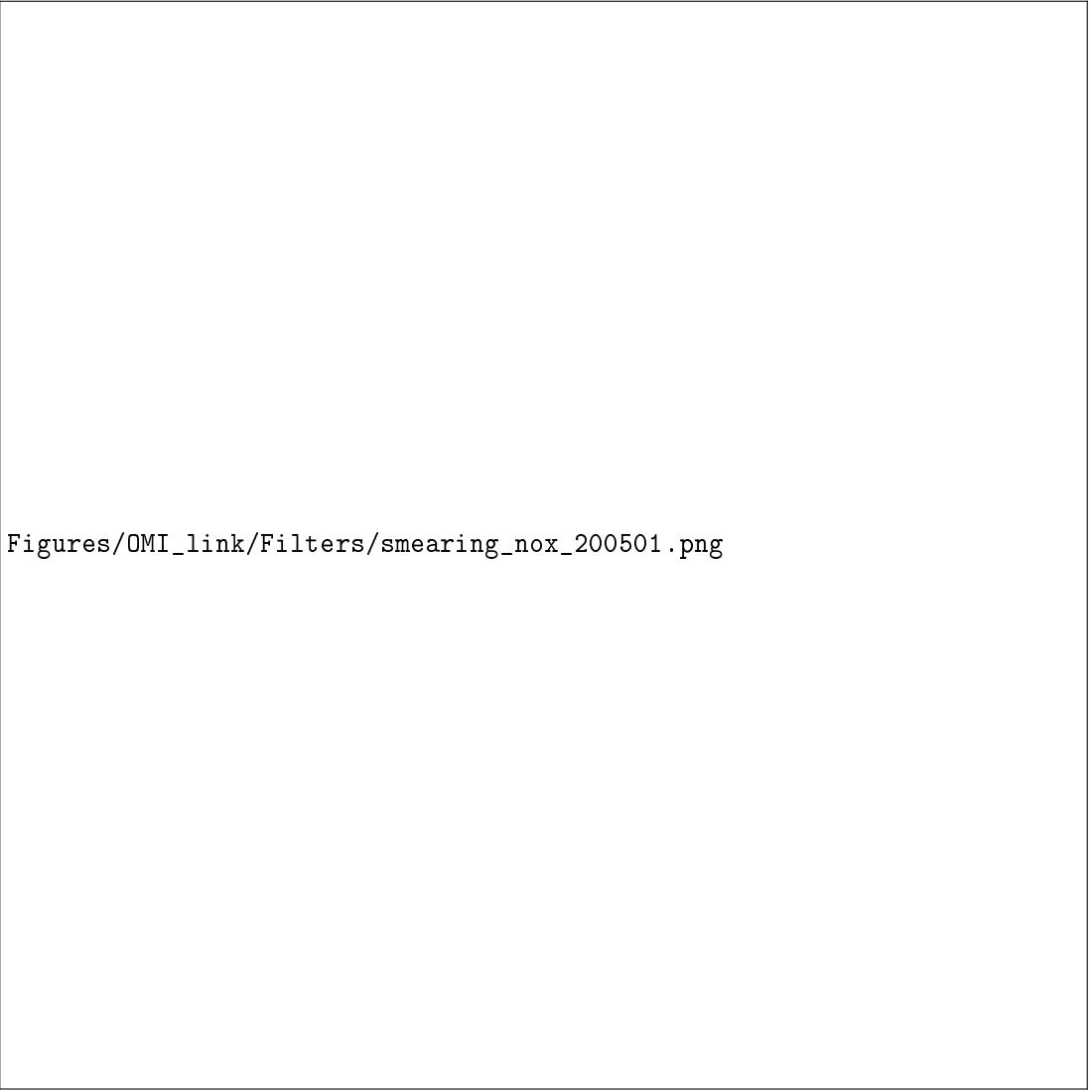
a **AtkinsonArey2003**: Table 2, Yield from Isoprene reaction with OH, two values are from two referenced papers therein.

b **Palmer2003**: lifetimes assume [OH] is 1e15 mol cm⁻³.

c **(Lee2006)**: Calculated through change in concentration of parent and product linear least squares regression. Estimates assume 20° C conditions.

d **Wolfe2016**: “prompt yield”: change in HCHO per change in isop₀. $[isop]_0 = [isop] \exp(k_1[OH]t)$; where k_1 is first order loss rate. Effectively relates HCHO abundance with isoprene emission strength.

e **Dufour2009**: One-day yields from oxidation modelled by CHIMERE, using MCM reference scheme.



Figures/OMI_link/Filters/smearing_nox_200501.png

FIGURE 3.9: Top left: OMNO2d tropospheric NO_2 columns (NO_2 : molec cm^{-2}) averaged into $2^\circ \times 2.5^\circ$ horizontal bins for Jan, 2005. Right: Scatter plot of NO_2 against smearing calculations from GEOS-Chem (\hat{S}), with points above and below the smearing threshold range of 900–5200 s coloured red and blue respectively. Points are binned by NO_2 with and without the smearing filter applied (orange and magenta respectively). Overplotted is the mean and standard deviation (error bars) within each bin. Due to the logarithmic Y scale we only show the positive direction of standard deviations for unfiltered data. Bottom left: Daily NO_2 scattered against smearing with (magenta) and without (orange) applying the smearing filter. This plot is a zoomed out version of the right panel.

which can be compared to the a priori equivalent. Missing values for α when E_{GC} are zero are a negligible issue since the dominant discrepancies between estimates occurs during summer when high emission rates are overestimated. Figure 3.10 shows α for the average January and June over 2005-2012, along with a time series of E_{GC} and E_{OMI} and α calculated for Sydney, and their multi-year seasonal average.

Top-down emission rates calculated in this work are in units of atom C cm⁻² s⁻¹. In order to calculate the emissions in kg, each grid square is multiplied by its area, and then daily emissions are assumed to follow a sine wave peaking at the midday value. Figure 3.11 shows how the daily approximation of total emitted isoprene per grid square is calculated. Daytime hours are estimated per month, from 14 hrs (Jan) to 10 hr (Jul) (<https://en.wikipedia.org/wiki/Daytime>). This approximation is required since OMI observations occur at midday, when isoprene emissions are at their diurnal peak.

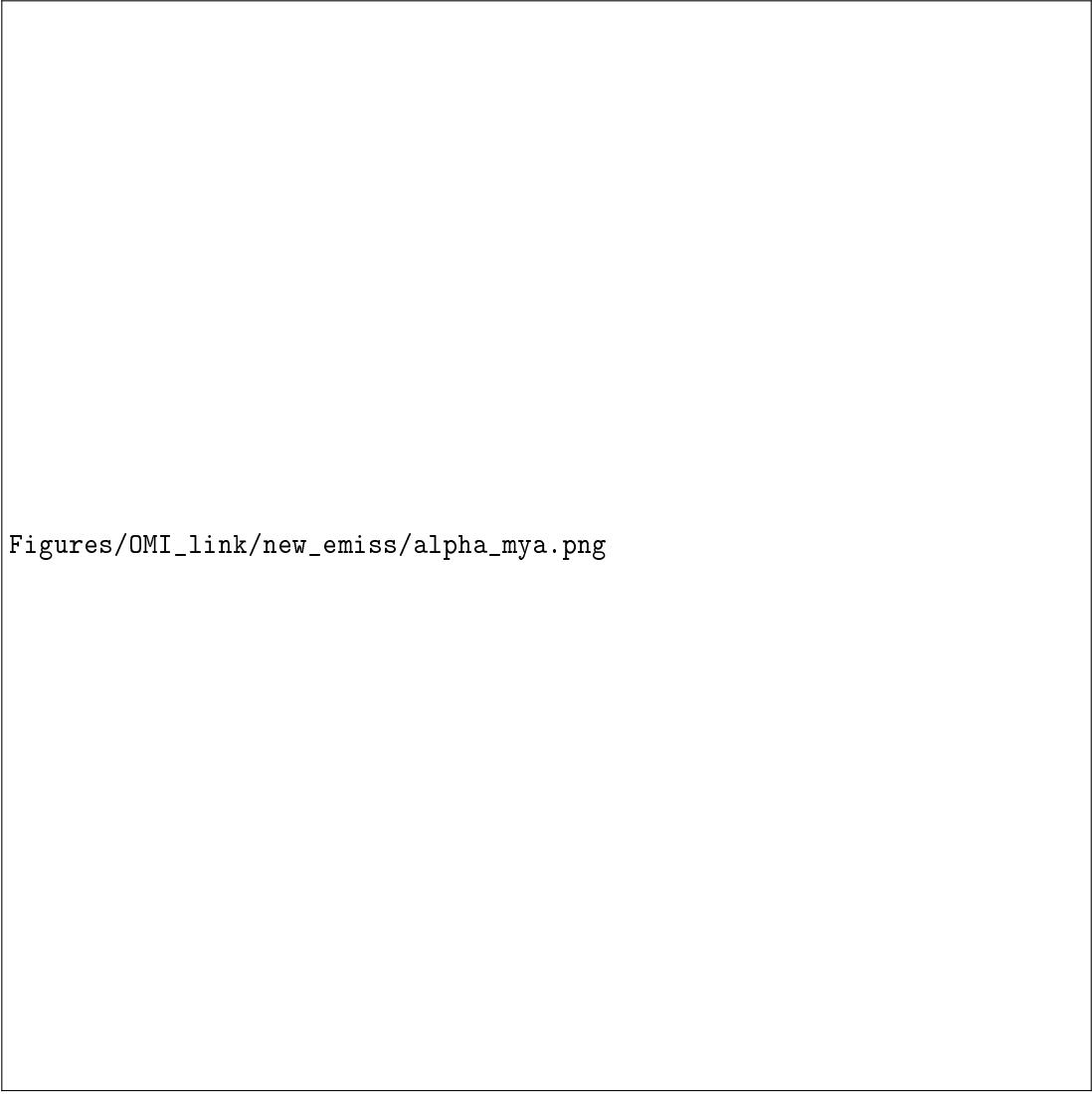
3.3 Results

Australia is roughly 7.7×10^6 km², with heterogeneous environmental conditions. The results presented in this section are frequently split into five regions that are differentiated by colour, as shown in Figure 3.12.

3.3.1 A posteriori emissions

Figure 3.13 shows a priori emissions over Australia and the a posteriori emissions calculated using OMI. The figure shows the time series (daily midday averages) their absolute differences (a priori - a posteriori) over 2005 TODO: Update plot to MYA. The seasonal cycle and the difference between the two metrics is clear, with much higher spring and summer averages in the a priori. Additionally, in the northern region, the a posteriori does not show significant emissions from mid March through to November which are seen in a priori emissions. This overestimate by the a priori may be caused by some mixture of overly high emission factors and high sensitivity to temperatures during the dry season. January and February also show a low feature in the a posteriori data, which may be due to filtering (TODO check data count), or may have been caused by meteorological factors (TODO Check if drought or fires occurred). This figure is repeated using monthly medians (of the daily midday estimates) in Figure 3.14 over 8 years from 2005 to 2012. Over the entirety of Australia the seasonal cycle of emissions is shown to be overestimated by the a priori. The difference between a priori and a posteriori estimates differs between regions, and some noise can be seen in the monthly time series due to data loss (TODO: check data counts). While most regions show similar overestimates, the northern region of Australia follows a different cycle of bias. Northern Australia appears to be overestimated throughout the year, however the bias tends to be lower in early summer when the a priori is similar to the a posteriori. One potential reason is the onset of the wet season (November-April) that disrupts the ecosystems and limits satellite measurement through increased cloud coverage.

TODO Figure shows the satellite data counts over time and summer vs winter maps from the OMHCORP dataset, along with uncertainty in each region



Figures/OMI_link/new emiss/alpha_mya.png

FIGURE 3.10: Row 1: α for the average January (left) and June (right) over 2005-2012. Row 2: E_{GC} (magenta, left axis), E_{OMI} (cyan, left axis), and α (black, right axis) multi-year monthly averages calculated for Sydney. Row 3: Monthly averages of the same terms in Row 2.

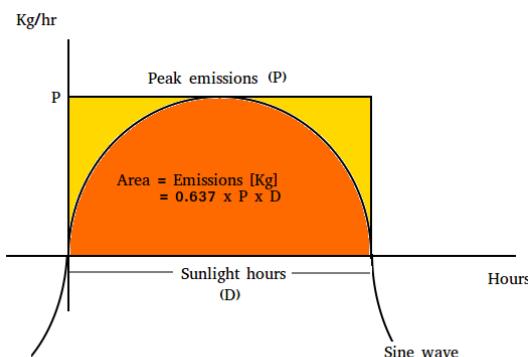
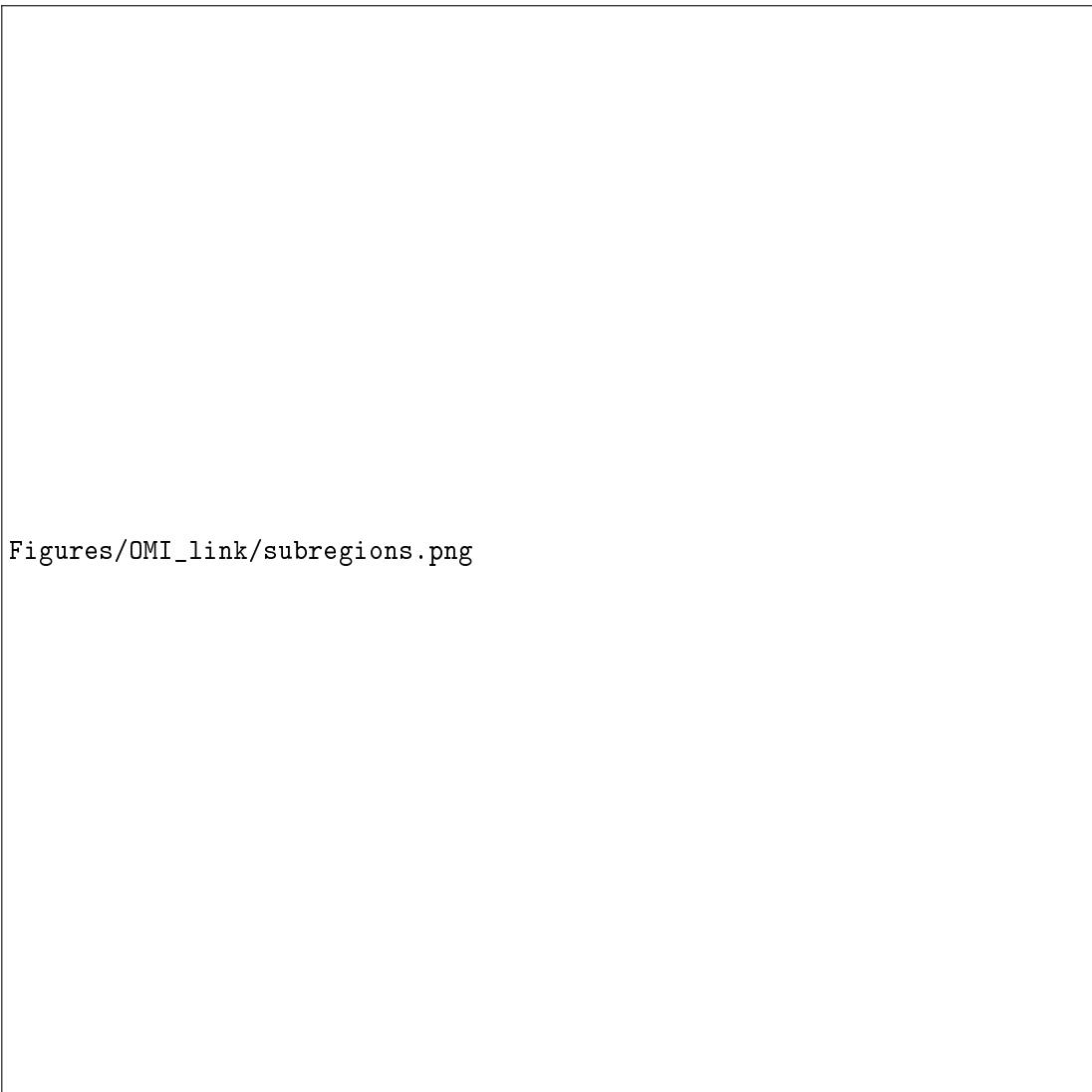


FIGURE 3.11: Total daily isoprene emissions (in kg) is represented by the area under the sine wave.

Figure 3.15 shows the multi-year monthly mean and inter-quartile range of daily midday isoprene emissions estimates, averaged regionally. Months outside of May to August show the a posteriori lower than the a priori. The most overlap is seen in the south-eastern region, and both eastern regions show the highest variance, potentially due to diversity within the regions and anthropogenic influence. Here the difference is only prominent during February, TODO check if this may be because of data point counts. TODO: Which regions show best correlation? Worst? Why might that be, and what does it mean?

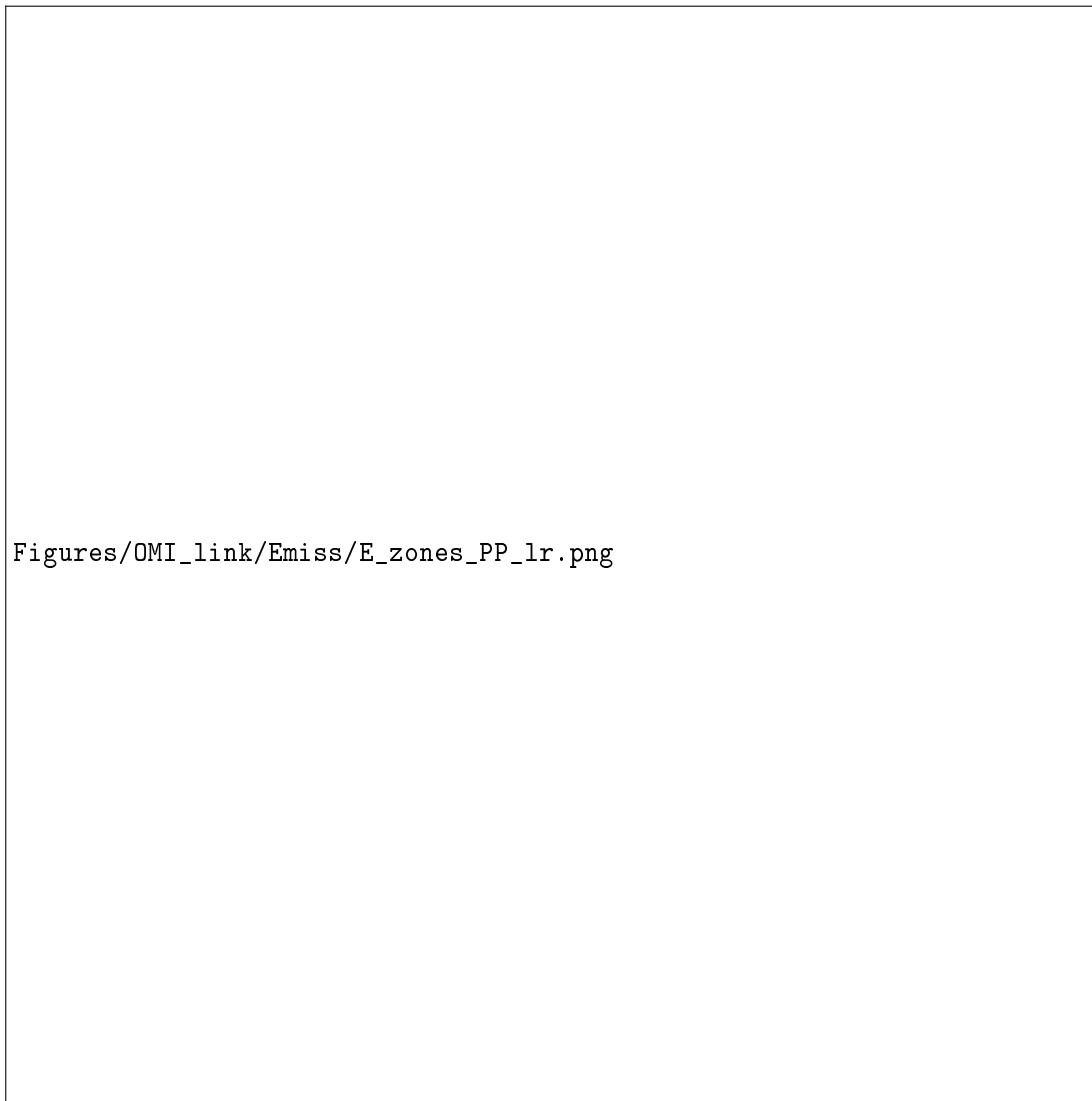
Figures 3.16 and 3.17 show how the distributions of a posteriori emissions compare to a priori emissions in each region during summer months (DJF) with zeros removed from both distributions. Figure 3.16 shows the daily midday distributions, binned hexagonally and coloured to show data-point frequency. There is only weak correlation between daily estimates, and the distributions are quite different with more variance shown by the a priori. This is likely due to the filtering applied to satellite data (e.g., whenever cloud coverage exceeds 40%) which reduces the number of days of a posteriori emission calculations. Figure 3.17 displays the regressions between monthly averages of the same data. In the monthly averages a stronger correlation is apparent, with regression coefficients ranging from 0.48 to 0.78 across regions. The portion of this correlation due to seasonality is examined by re-running the regression after deseasonalising both data sets by subtracting the multi-year monthly average from each dataset. TODO: run regression after removing seasonal average and see how r and slope are affected.

When comparing GEOS-Chem (MEGAN) to the a posteriori calculated using our top-down inversion, we find a decrease of from 45.4 Tg yr^{-1} to 16.5 Tg yr^{-1} ($\sim 66\%$). Table 3.3 compares annual Australian isoprene emissions from this work to previously published values. Our a posteriori estimate of 19 Tg yr^{-1} suggests isoprene emissions may be lower than any bottom up estimates, and is close to the top-down estimate of Bauwens2016 of 26 Tg yr^{-1} . Figure 3.18 shows how this decrease is distributed spatially, with E_{GC} and E_{OMI} in Tg yr^{-1} calculated as a multi-year mean. Across all of Australia we see large reductions of total emissions using the new top-down estimate.



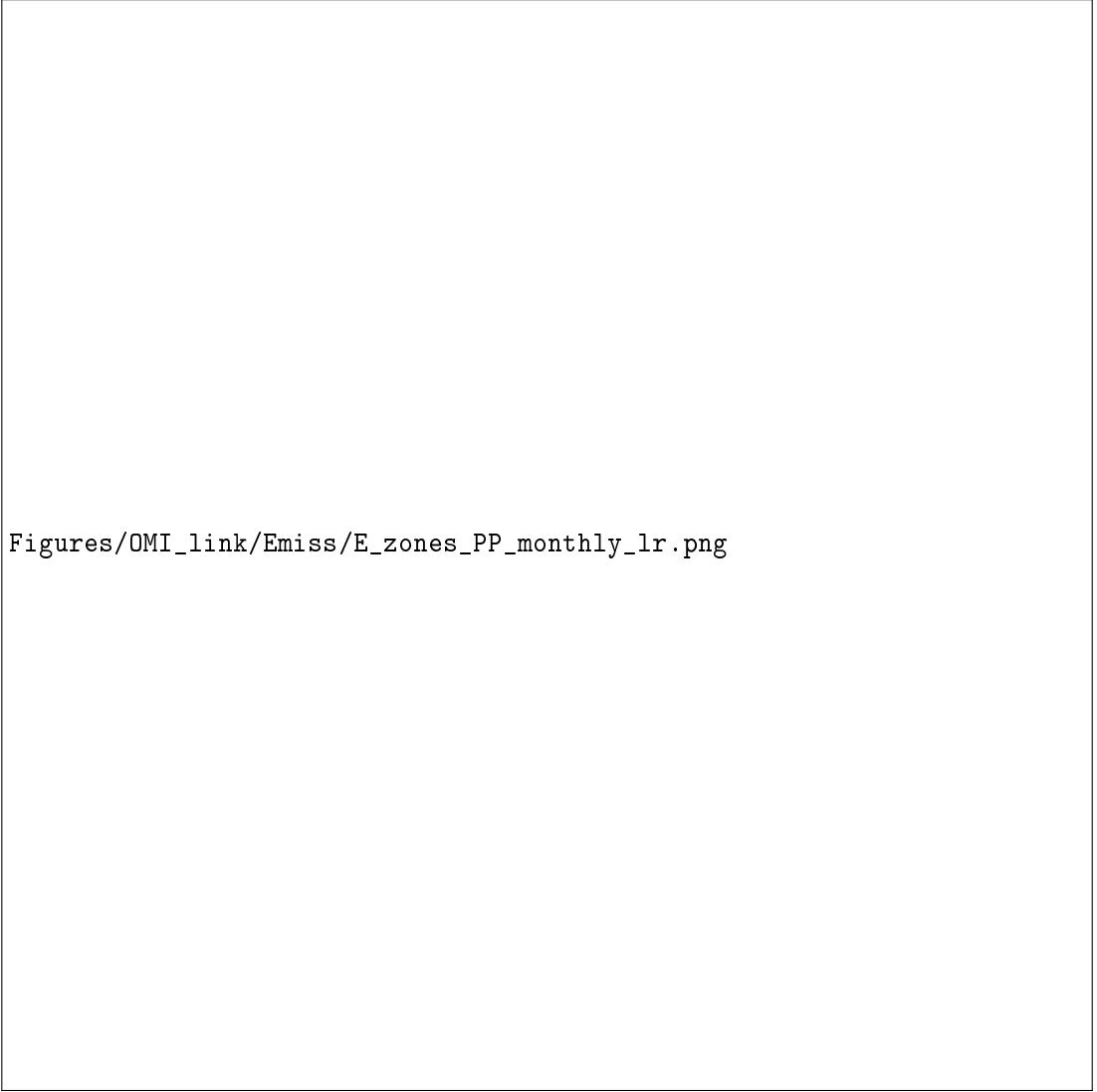
Figures/OMI_link/subregions.png

FIGURE 3.12: Sub-regions used in subsequent figures. Australia-wide averages will be black or grey, while averages from within the coloured rectangles will match the colour shown here.



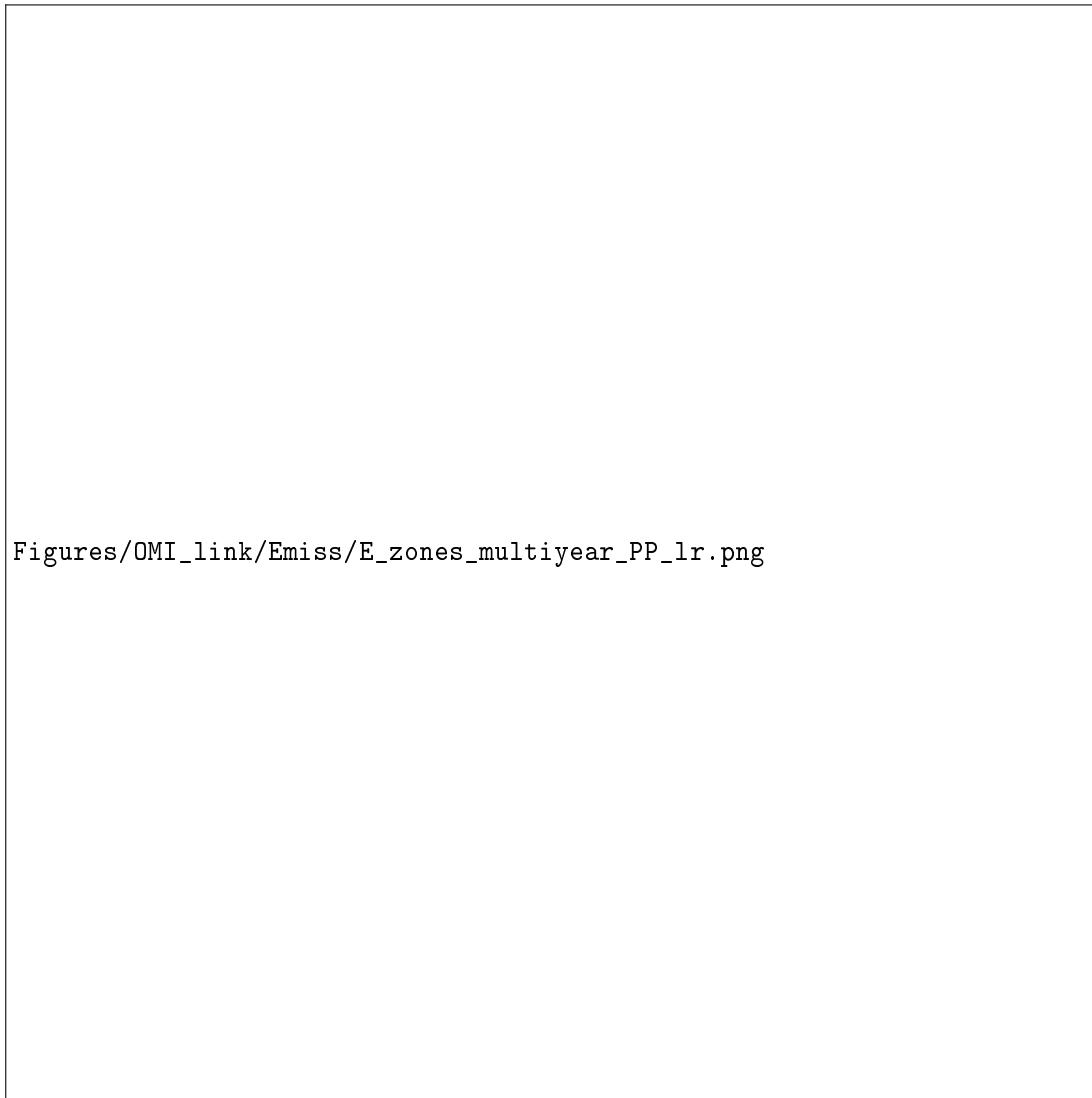
Figures/OMI_link/Emiss/E_zones_PP_lr.png

FIGURE 3.13: Biogenic emissions of isoprene from GEOS-Chem (running MEGAN), which is our a priori in column 1, and the emissions calculated using the OMI top down inversion, which is our a posteriori in column 2. Row 1: time series of daily midday averages for all of Australia along with several subregions (shown in row 1). The black lines and grey areas show the Australian mean and inter-quartile range respectively, while the coloured lines show the mean within subregions as shown in Figure 3.12. The overall regional averages are shown at the end of the time series by horizontal lines. Row 2: Absolute differences: a priori - a posteriori. TODO: remove spines along top and middle of top row



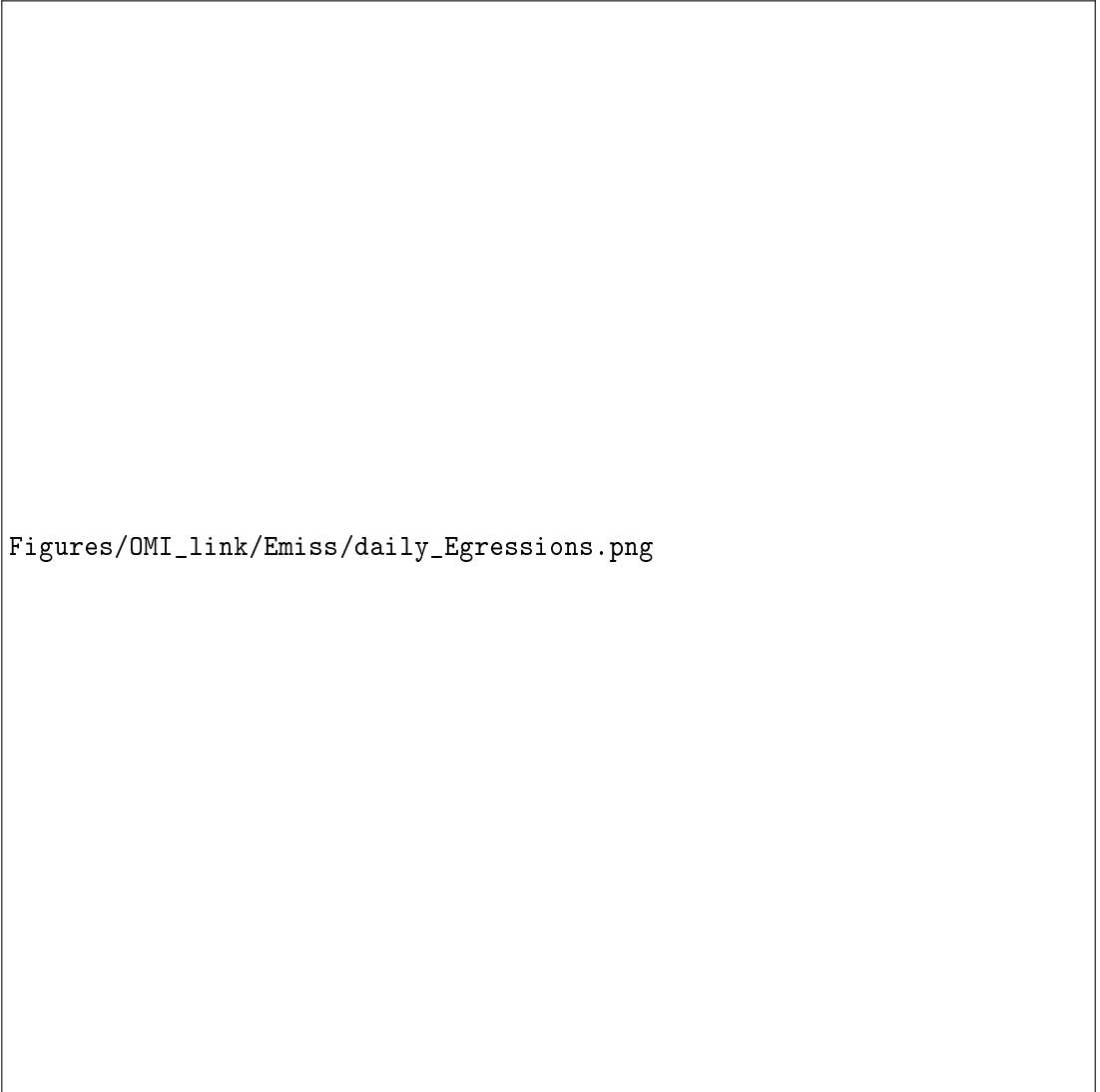
Figures/OMI_link/Emiss/E_zones_PP_monthly_lr.png

FIGURE 3.14: As Figure 3.13 using monthly medians of the daily mid-day emissions estimates.



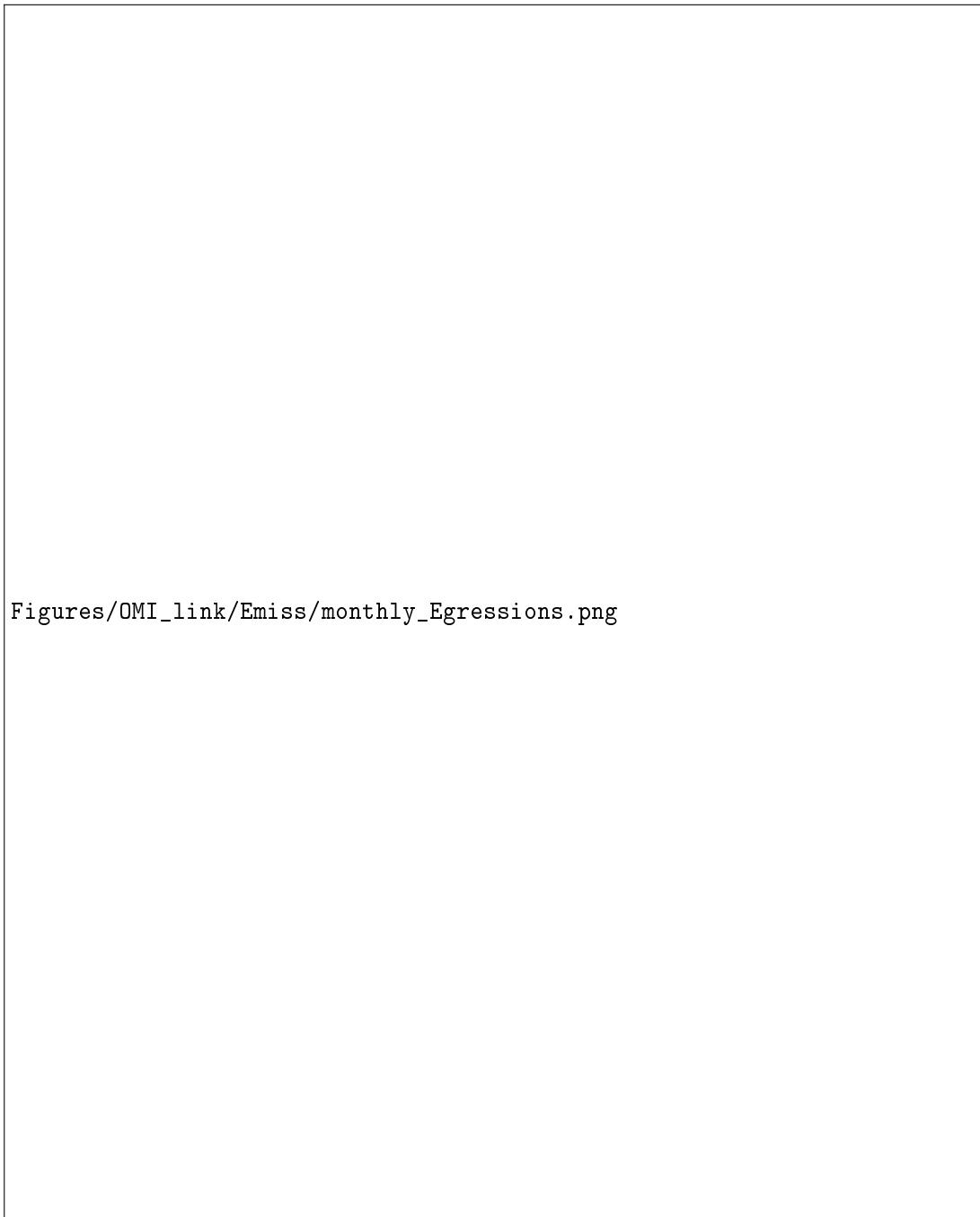
Figures/OMI_link/Emiss/E_zones_multiyyear_PP_lr.png

FIGURE 3.15: The multi-year monthly mean (lines) and IQR (shaded) of midday (13:00-14:00 LT) isoprene emissions estimates. A priori emissions are shown by the dashed lines and hatched shaded areas show the IQR. A posteriori emissions are shown using the solid lines, with IQR shown by unhatched shaded areas. Colours denote the region over which the monthly average was taken, as in Figure 3.12.



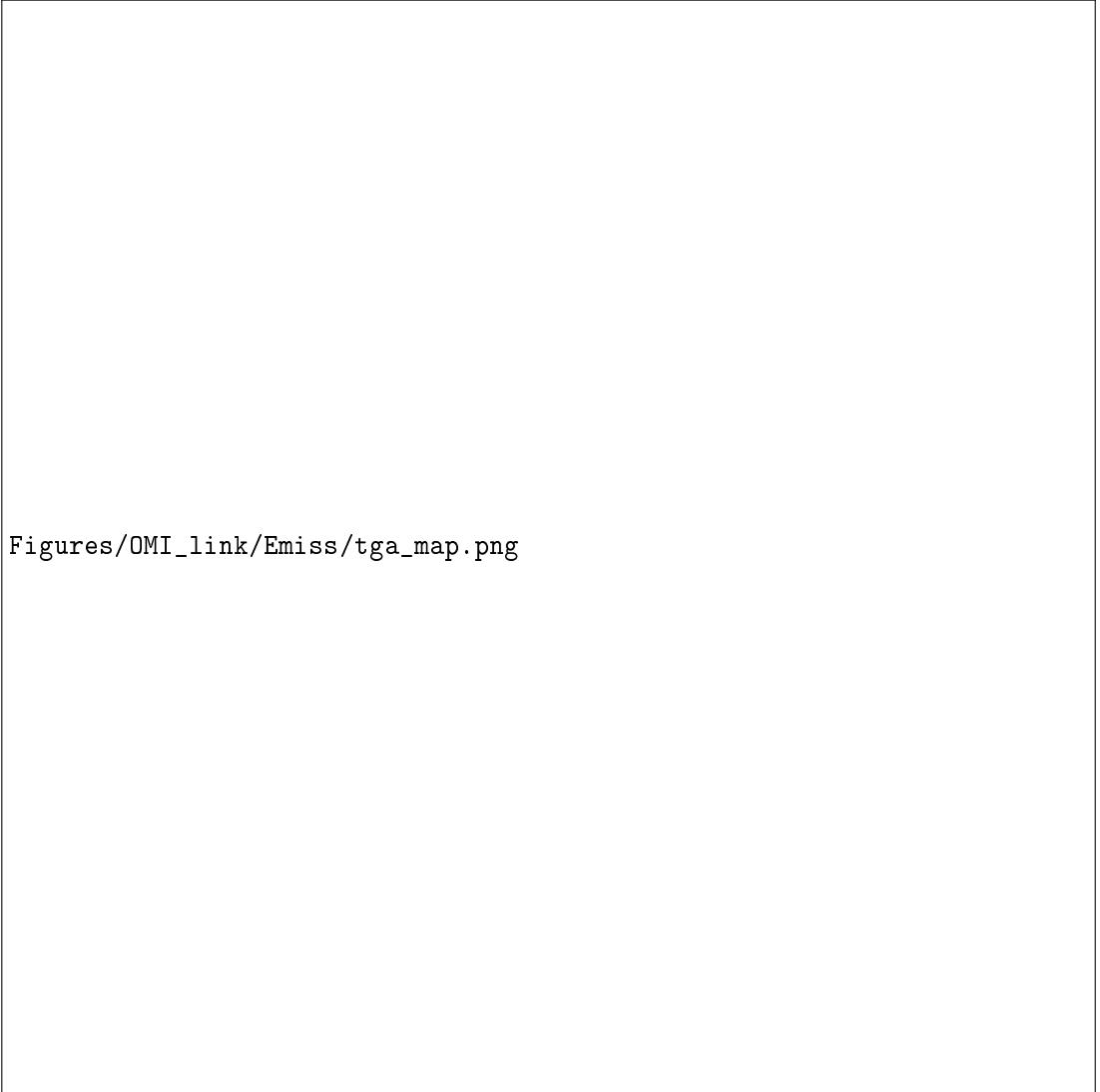
Figures/OMI_link/Emiss/daily_Egressions.png

FIGURE 3.16: Scatter plot (binned hexagonally with darker colours representing higher data-point frequency) along with the distributions of MEGAN (y axis) and the top down estimate (x axis). This figure is based on summer (DJF) midday values over multiple years. Coloured by regions shown in Figure 3.12.



Figures/OMI_link/Emiss/monthly_Egressions.png

FIGURE 3.17: Scatter plot of a priori emissions against a posteriori using monthly averaged grid squares as regression datapoints. Data points are created using monthly averages (of midday emissions) for each grid box for each month of summer (DJF) within each region shown. Multiple years of data are used, meaning if a region has 10 grid boxes, the 8 years of data will add up to $10 \text{ boxes} \times 3 \text{ months} \times 8 \text{ years} = 240$ data points minus filtered and zero emission squares. Plots are coloured by regions matching those shown in Figure 3.12. The linear best fit regression is inset into each plot along with the line equation and regression coefficient.



Figures/OMI_link/Emiss/tga_map.png

FIGURE 3.18: Top row: multi-year mean emissions in Tg yr^{-1} from E_{GC} (GEOS-Chem; running MEGAN) and E_{OMI} (top-down emissions) respectively. E_{OMI} uses an assumed sinusoidal daily cycle, with daylight hours prescribed for each month: see Section 3.2.6). Bottom left and right shows the absolute and relative differences respectively.

TABLE 3.3: Isoprene emissions (Tg/yr) from Australia

Estimate	source	year(s)	notes
43(2)	bottom-up	2005-2010	(a) from this thesis
19(2)	top-down	2005-2010	(b) from this thesis
~ 80	bottom-up	1980-2010	(c) Sindelarova2014
26-94	both	2005-2013	(d) Bauwens2016

- a: GEOS-Chem with MEGAN diagnostics based on 3-hourly averages
 b: Based on daily peak emissions integrated over a sinusoidal daily curve
 c: MEGAN run using MERRA meteorology.
 d: Range shown here based on 3 different models and one top-down inversion

3.3.2 Modelled impacts of reduced isoprene emissions

This section uses GEOS-Chem to determine how the improvements to biogenic isoprene emissions impact subsequent atmospheric chemistry and composition. A posteriori emissions are implemented in GEOS-Chem as described in Section 3.2.8. Outputs from the scaled GEOS-Chem run using the a posteriori will be denoted by a superscript α . For example, column HCHO from GEOS-Chem before and after scaling are denoted Ω_{GC} and Ω_{GC}^α respectively.

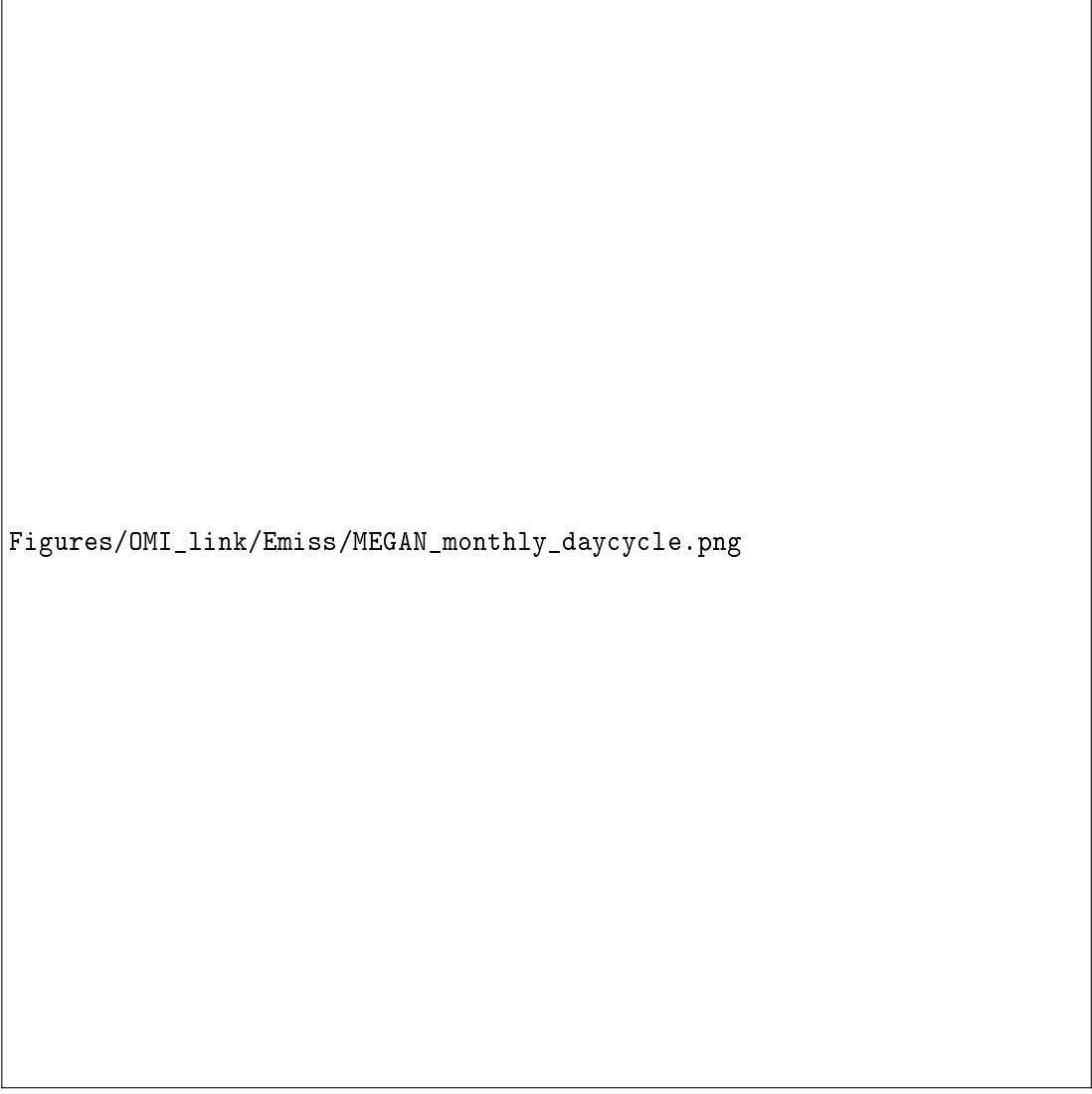
Figure 3.19 shows the multi-year monthly mean daily cycle of isoprene emissions from GEOS-Chem, along with the top-down midday estimate, for an example grid box located 2.5° west of Sydney. This box is typical of our south-eastern region and is not heavily impacted by the anthropogenic filter. TODO: Analysis of this and lead into more analysis of differences

3.3.2.1 Implications for HCHO

After running GEOS-Chem with isoprene emission scaled to match the a posteriori, we first compare simulated Ω_{GC} and Ω_{GC}^α to Ω_{OMI} over January in 2005 (TODO: update to use all summers and winters) in Figure 3.20. In every region, Ω_{GC}^α is closer to Ω_{OMI} with biases decreasing from 50-120% down to 30-50%. These Ω_{OMI} drive the creation of Ω_{GC}^α , and as we expect the relationship is much improved. The remaining differences are due to temporal averaging of the applied scaling factor α .

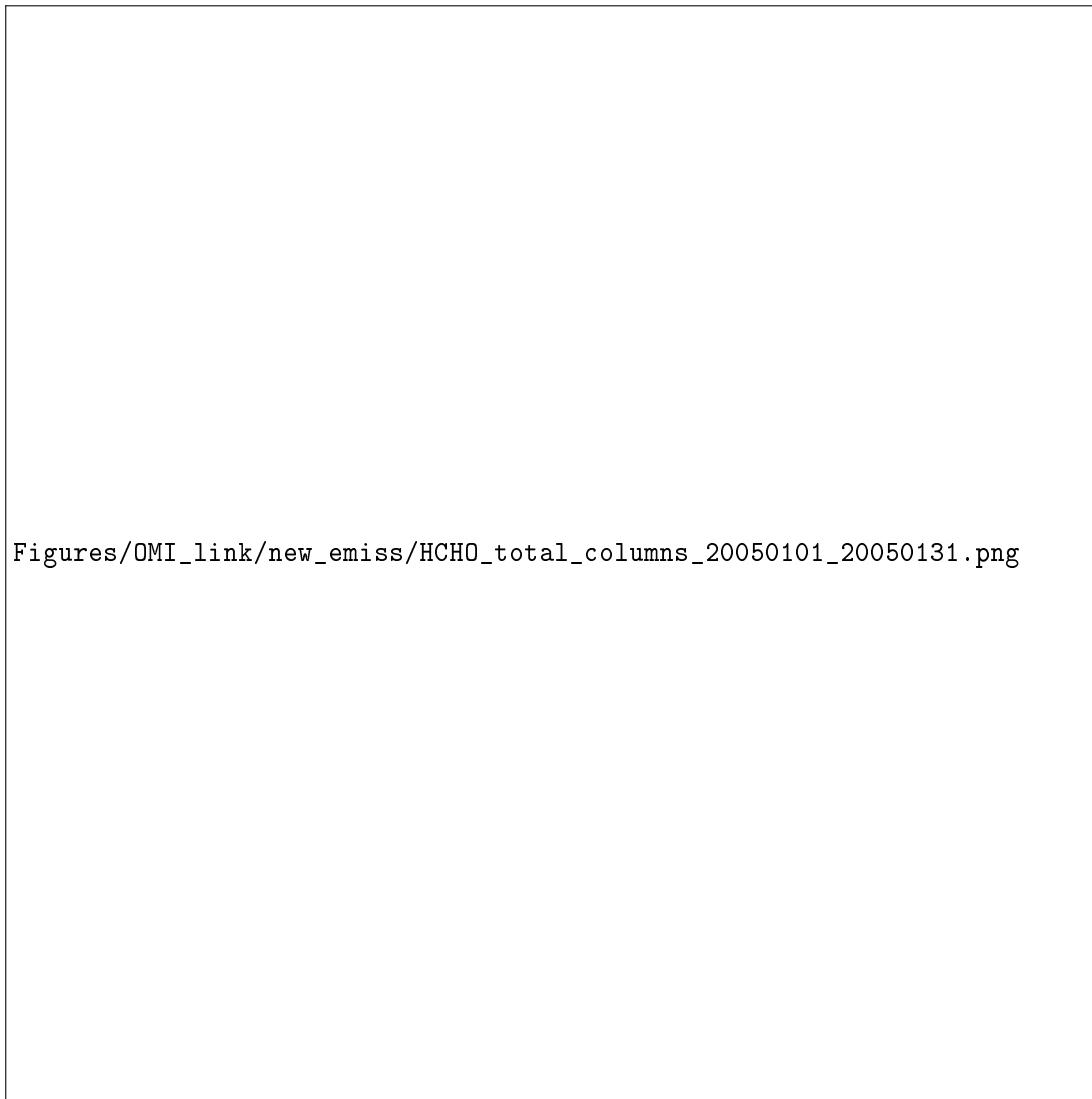
I also examine the effects of scaling isoprene emissions on the correlation and variances between modelled and satellite based HCHO columns. Figure TODO: shows the regressions between Ω_{GC} , Ω_{GC}^α , and Ω_{OMI} , for summer and winter between for each region. Figure TODO shows variances in each season for each region in Ω_{GC} , Ω_{GC}^α , and Ω_{OMI} . TODO: Analysis of plots

Wollongong FTIR measurements (see Section 2.2.3.4) provide total column amounts TODO: get these and check process. This is the only non-satellite long-term measurement record of total column HCHO available in Australia and we use it here to examine trends and seasonality. The time series for HCHO is shown in figure TODO,



Figures/OMI_link/Emiss/MEGAN_monthly_daycycle.png

FIGURE 3.19: The diurnal cycle of MEGAN emissions averaged by month over 1, Jan, 2005 to TODO, 2013 are shown with lines, while top-down emissions estimates are shown with plus symbols. MEGAN emissions are estimated hourly per $2^\circ \times 2.5^\circ$ horizontal grid box, shown here are the area averages (denoted by colour, see Figure 3.12). Top-down estimates are similarly grouped by colour, and shown at the 13:00 LT mark for each month.



Figures/OMI_link/new emiss/HCHO_total_columns_20050101_20050131.png

FIGURE 3.20: Daily mean total column HCHO amounts from GEOS-Chem with (new emissions run) and without (tropchem run) a posteriori scaled isoprene emissions, along with the original OMI HCHO columns. Each row shows the average over regions in Figure 3.12.

along with GEOS-Chem output before and after updating isoprene emissions. Modelled HCHO in this figure only shows the grid square containing Wollongong, which is the same grid square that contains Sydney and a large amount of rural and forested area. This makes direct comparison between GEOS-Chem output and the FTIR suffer from representational error. TODO: discuss plot here

TODO: Figure showing campaign data against model and recalculated model.

Figure TODO shows how Wollongong FTIR HCHO profiles compare against GEOS-Chem modelled HCHO before and after isoprene scaling is performed. The regressions show improvement/no improvement TODO

Figure todo shows how the longer term FTIR HCHO measurements compare against modelled HCHO before and after changing isoprene emissions

3.3.2.2 Implications for ozone

Isoprene oxidation and can eventually lead to ozone formation, especially when isoprene enriched air masses mix with polluted urban air masses that contain high NO_x concentrations. Figure 5.3 shows surface level (up to ~ 150 m altitude) ozone concentrations over 2005 before and after scaling modelled isoprene emissions. Reducing isoprene emissions lowers surface ozone concentrations by TODO: XX to YY % in summer, and XX to YY % in winter. The direct correlation between reduced emissions and surface ozone (Figure TODO) shows that reduced emissions directly relate to reduced ozone concentrations, and using multiple years enhances/reduces the relationship. A look at the regression using only summer data reveals a stronger/weaker relationship, analyse per region (TODO).

TODO Figure showing isoprene vs ozone time series and correlation.

The relatively clean atmosphere over the majority of Australia is likely to be NO_x limited outside of major population centres. This means that cities may have higher sensitivity to changes in modelled isoprene emissions than shown here, since we are examining large area averages which are mostly non-urban. TODO: Notes about ozone exposure from State of the Environment report 2016.

TODO: Figure shows modelled surface ozone concentrations and their differences between model runs over an average summer (DJF). Figure TODO shows the same over an average winter (JJA). Figure TODO compares the summer differences to surface NO emissions (modelled), and total column NO (satellite).

3.3.2.3 Trends

Figure 3.21 shows monthly deseasonalised a priori and a posteriori midday emission anomalies for each region. First the emissions are spatially averaged within each region to form a daily time series of midday emission rates. This is averaged into monthly data, and then the multiyear monthly mean is subtracted to form the anomaly time series. Any anomaly greater than three standard deviations from the mean is removed (crosses in Figure 3.21). An ordinary least squares linear regression is then performed to look for any significant trend. A trend is considered significant if the p-value from a Wald test (roughly equivalent to a t-test) is less than 0.05. The midday emissions show no significant trends over the 8 year period from 2005-2012. The same process is repeated for surface ozone, surface HCHO, and surface NO₂, with anomaly series

shown in Figures 3.22 to 3.24. None of these metrics shows any significant changes in trend due to scaling isoprene emissions. Some trends are no longer significant in surface HCHO, however none of the trends change sign. Since the scaling factor α is applied to each grid square seasonally and not changing each year, changing trends are not expected.

3.3.3 Comparison with in situ measurements

TODO: Analyse comparison of grid box with campaigns of measurements

Comparison between ground-based measurements and large ($2^\circ \times 2.5^\circ$) averaged grid squares suffers from representation error. Figure 3.25 shows the SPS and MUMBA measurement sites, along with the extent of the $2^\circ \times 2.5^\circ$ relevant GEOS-Chem grid box. The grid box is the area over which GEOS-Chem outputs are averaged, a rectangle with edge lengths of roughly 200 km². The urban footprint of Sydney and Wollongong can be seen, along with some ocean, forest, and rural regions, which will all affect the model output and other calculations averaged here. Due to high uncertainty in components of the top-down emissions estimate, temporal resolution is also limited. MUMBA, SPS1 and SPS2 each provide relatively on the order of one month of hourly or daily datapoints, which are compared in this section against surface level concentrations from GEOS-Chem before and after scaling the biogenic emissions.

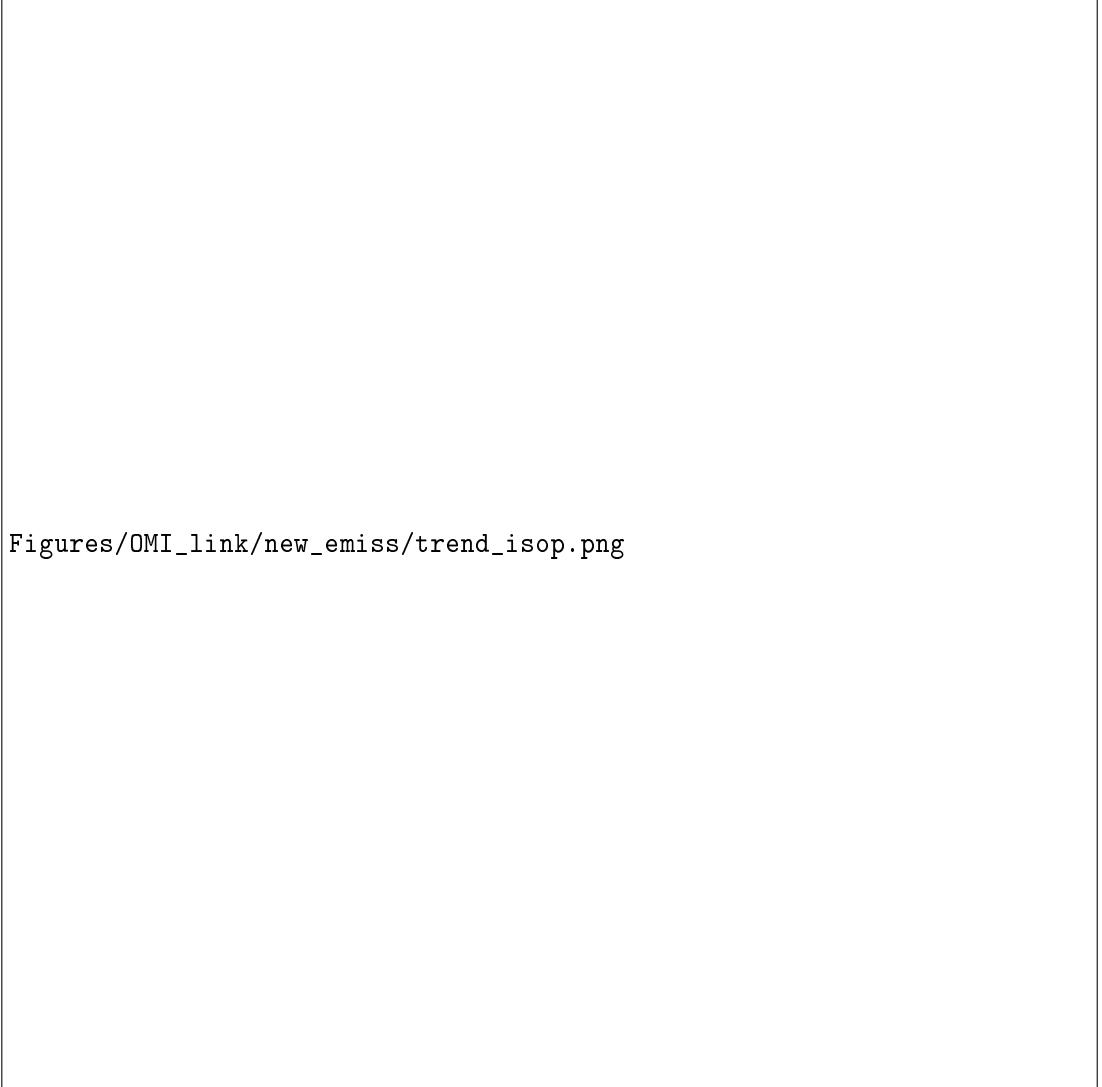
Figure 3.26 shows GEOS-Chem output in the grid square containing Sydney overlaid on SPS measurement data. Superficially the comparison is not too bad between these two datasets. TODO: mean midday quantities for matching months. This figure shows at a glance how the measurements compare to modelled data, and also the limited nature of the temporal and spatial scale of available measurements. TODO: Add scaled output to plot, and discuss The SPS data is point-source and taken during the daytime when isoprene is higher, so it is very likely that GEOS-Chem HCHO and isoprene output is in fact too high since the daily average should not match the peak of the measurements.

MUMBA data captured concentrations at the surface in Wollongong, and is compared in figure TODO against simulated HCHO concentrations in the first level (~0 - 150 metres) over corresponding dates. GEOS-Chem midday concentrations are shown, along with Wollongong bottom-level FTIR measurements to compare against the MUMBA data. TODO Discussion of comparison.

3.4 Uncertainty

Uncertainties introduced through the inversion process are difficult to adequately quantify. We identify the uncertainties in the linear regression used to relate HCHO to isoprene emissions, as well as in the satellite data product used as the basis for the inversion; however, these uncertainties lack verification against measurements. Even as this top-down inversion attempts to work around the lack of measurement over Australia, it suffers from the lack of independent observations against which it can be verified.

This section identifies and quantifies the overall uncertainties of calculating isoprene emissions using OMI HCHO observations and the GEOS-Chem model in the



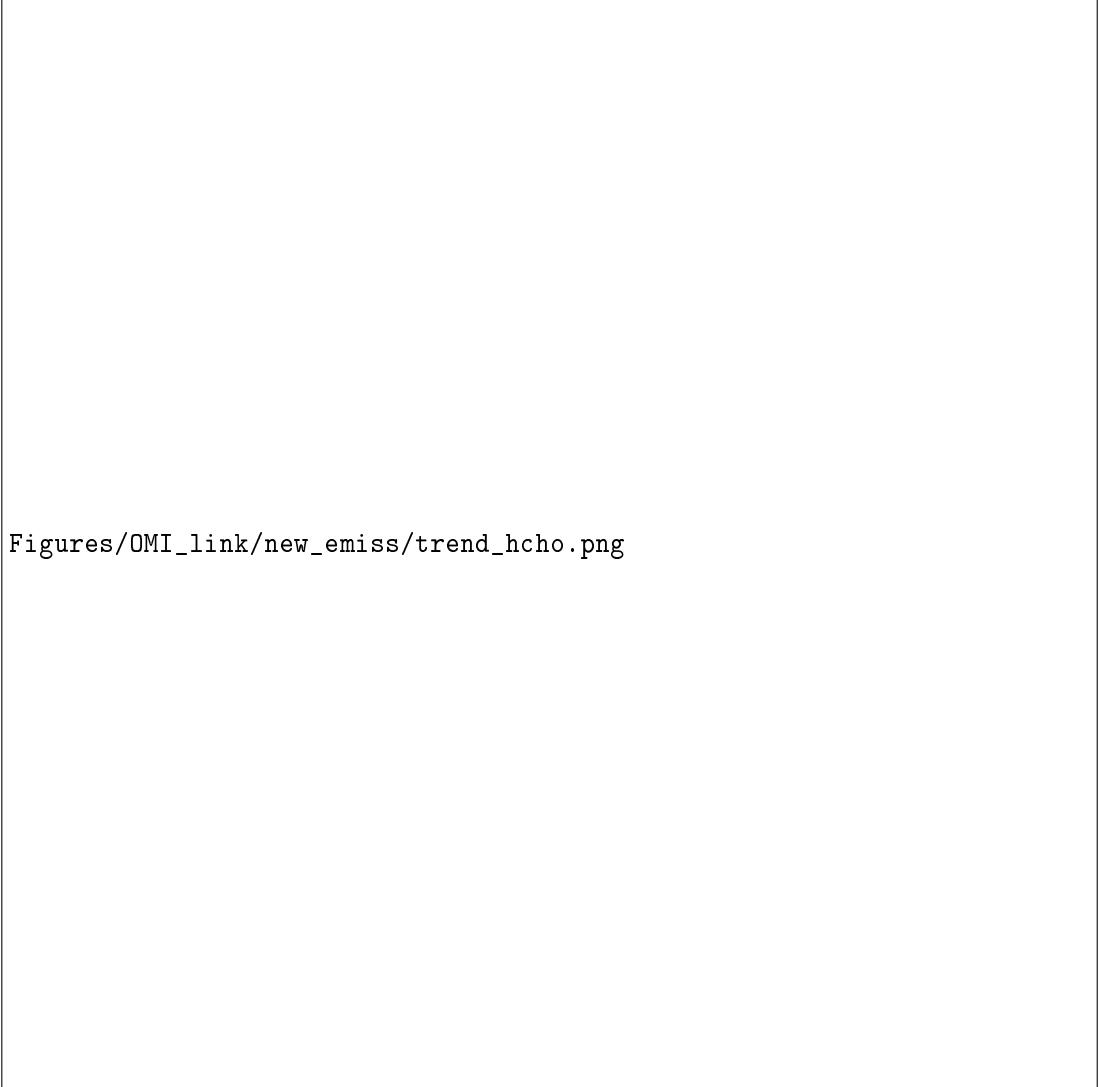
Figures/OMI_link/new emiss/trend_isop.png

FIGURE 3.21: A priori (row 1) and a posteriori (row 2) emissions anomaly from multiyear monthly mean, split by region (see Figure 3.12).



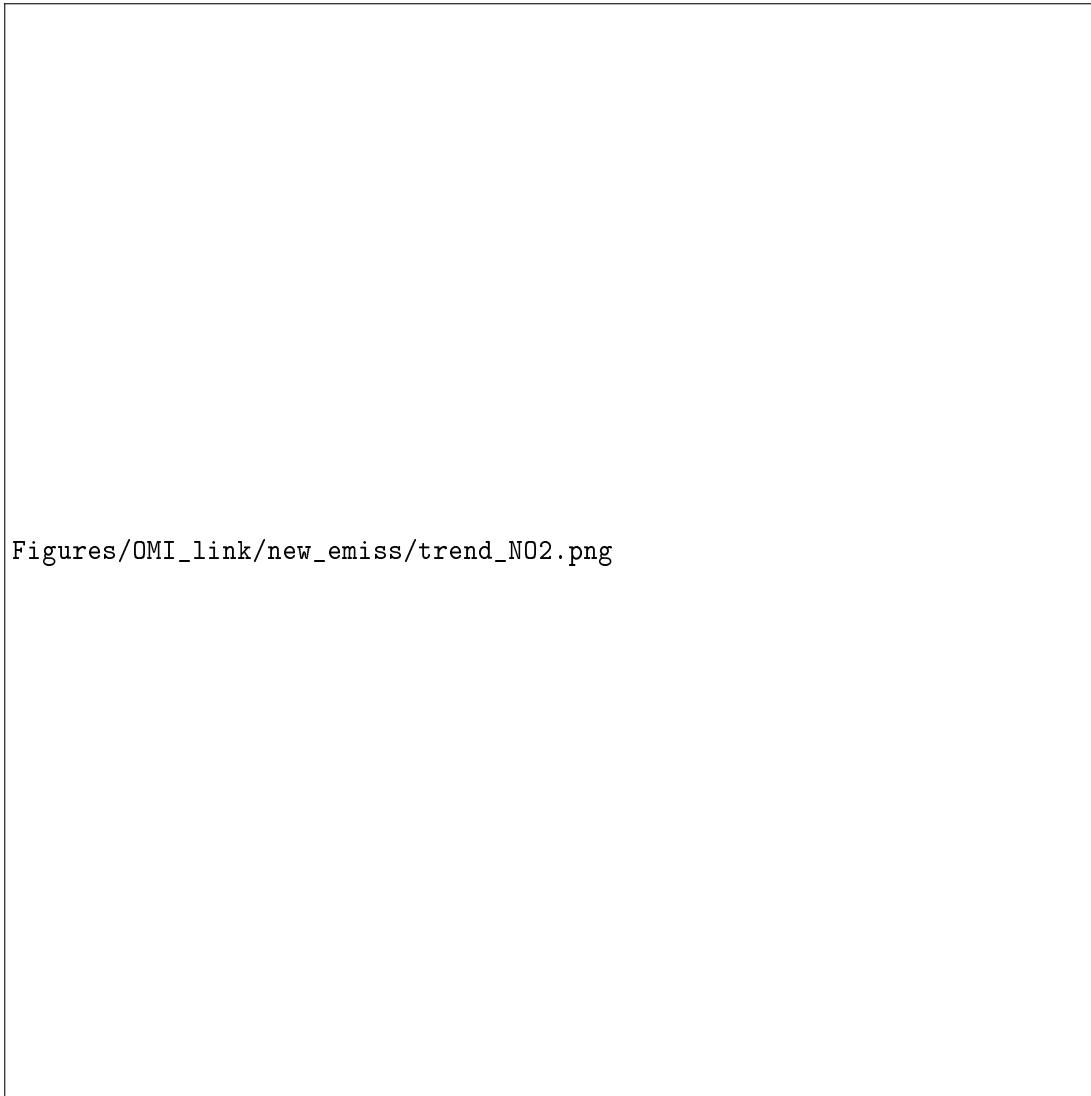
Figures/OMI_link/new emiss/trend_03.png

FIGURE 3.22: A priori (row 1) and a posteriori (row 2) surface ozone anomaly from multiyear monthly mean, split by region (see Figure 3.12).



Figures/OMI_link/new emiss/trend_hcho.png

FIGURE 3.23: A priori (row 1) and a posteriori (row 2) surface HCHO anomaly from multiyear monthly mean, split by region (see Figure 3.12).



Figures/OMI_link/new emiss/trend_N02.png

FIGURE 3.24: A priori (row 1) and a posteriori (row 2) surface NO_2 anomaly from multiyear monthly mean, split by region (see Figure 3.12).



FIGURE 3.25: grid box at $2^\circ \times 2.5^\circ$ containing SPS, and MUMBA campaign data.



Figures/OMI_link/GC_vs_Campaigns_20050131-20050601.png

FIGURE 3.26: Comparison between GEOS-Chem HCHO concentrations in the grid square containing Sydney for the duration of the SPS 1 and 2 campaigns

TABLE 3.4: Uncertainty estimates.

Region	Summer			Winter		
	ΔE_{OMI}	$\Delta \Omega_{OMI}$	ΔS	ΔE_{OMI}	$\Delta \Omega_{OMI}$	ΔS
Aus						
SE						
NE						
...						

top-down method used in this chapter. The major source of uncertainty lies in TODO, which is due to TODO. Uncertainty in satellite HCHO is seasonally dependent, with better signal during the summer. Reliable OMI measurements are less frequent (so uncertainty is higher) at high solar zenith angles, which occurs at higher latitudes and during winter. Table 3.4 shows the estimated uncertainty calculated in this work in summer and winter over each region described by Figure 3.12. The process (or methods) used to calculate each of these uncertainties is described in the following subsections.

3.4.1 Top down emissions

Important factors in the calculation of isoprene emissions using OMI HCHO include the modelled a priori, the relationship between HCHO and isoprene, and satellite measurements. Uncertainty in each of these terms is quantified before being combined in quadrature to give the uncertainty estimate of the a posteriori. Additional biases may arise due to the filters applied to satellite data and model output, and where possible these are assessed.

The final determination of top-down emissions comes from Equation 3.8, repeated here:

$$E_{OMI} = \frac{\Omega_{OMI} - \Omega_{OMI,0}}{S}$$

Assuming each term is independent, we use the following quadrature rules to estimate random error in E_{OMI} :

$$z = x + y : \Delta z = \sqrt{(\Delta x)^2 + (\Delta y)^2} \quad (3.13)$$

$$z = x/y : \Delta z = z \sqrt{\left(\frac{\Delta x}{x}\right)^2 + \left(\frac{\Delta y}{y}\right)^2} \quad (3.14)$$

Which leads to the uncertainty estimation for our a posteriori emissions as follows

$$\Phi \equiv \Omega_{OMI} - \Omega_{OMI,0} \\ \Delta\Phi = \sqrt{(\Delta\Omega_{OMI})^2 + (\Delta\Omega_{OMI,0})^2} \quad (3.15)$$

$$\Delta E_{OMI} = E_{OMI} \times \sqrt{\left(\frac{\Delta\Phi}{\Phi}\right)^2 + \left(\frac{\Delta S}{S}\right)^2} \quad (3.16)$$

To quantify ΔE_{OMI} we determine the uncertainty in underlying terms: ΔS , $\Delta \Omega_{OMI}$, and $\Delta \Omega_{OMI,0}$. For ΔS ($\Omega_{GC} = S \times E_{GC} + \Omega_{OMI,0}$ from equation 3.7) we examine related GEOS-Chem output in Section 3.4.2, since S comes directly from the monthly linear regression of modelled isoprene emissions and column HCHO. Uncertainty in terms Ω_{OMI} , and $\Omega_{OMI,0}$ come from the OMHCHO slant columns (SC), the background error, and the AMF calculated to transform them into vertical columns ($\Omega = \frac{SC}{AMF}$ from equation 2.8). Section 3.4.3 describes these calculations. The overall summary of uncertainty is presented in Section 3.4.6.

3.4.2 Model Uncertainty

E_{OMI} depends partly on the product it is trying to improve, as modelled yield is based on GEOS-Chem run with MEGAN emissions. The uncertainty in the RMA regression slope between model HCHO and emissions (Ω_{OMI} and E_{GC} respectively) is used to estimate ΔS in Equation 3.16. This is a simple method of approximating the uncertainty of this term, only accounting for monthly uncertainty of the slope calculation. It does not take into account uncertainty in the underlying model, nor uncertainties arising from temporal or spatial resolution issues, which are much more difficult to quantify.. Figure TODO shows the multi-year mean summer and winter ΔS over Australia. Analysis todo For comparison Palmer2006 found $\frac{\Delta S}{S}$ to be 30%, after comparison with another chemical model and in situ measurements. To improve understanding of uncertainty in S would require further analysis of GEOS-Chem yield over Australia, including how it responds to environmental and meteorological parameters, and how representative this modelled quantity is compared to measurements.

Filtering for spatial smearing (see Section 3.2.7) reduces the number of data points making up our regression slope S . The process generally improves the linear relationship, and where it does not a multi-year average (or in the worst cases no value at all) is used in lieu of monthly S data. The thresholds for the smearing filter are based on literature based overseas, a better approximation of HCHO lifetimes and yields as well as properly accounting for seasonality and regional NO_x concentrations would improve the filter. The filtering process is accounted for in the calculation of ΔS , and the overall effect on uncertainty is limited, so we do not analyse it further.

Model biases are not analysed in this thesis, except to note that they would impact both preliminary OMI calculations and the modelled slope. Insufficient independent measurements in Australia mean that quantified uncertainties would require unreasonable extrapolation to apply at a national scale.

3.4.3 Satellite Uncertainty

There are three main sources of error in the satellite HCHO columns:

Fitting error from the OMI retrieval

the uncertainty in the DOAS technique used to estimate HCHO concentrations.
Fitting error is provided in the OMHCHO product.

Fitting error represents

Uncertainty in AMF calculations

Air mass factors model how the satellite instrument is vertically sensitive to slant path measurements, and uncertainties arise predominantly from uncertain cloud parameters (Palmer2006).

AMF uncertainty can be determined through comparison of GEOS-Chem output to independently measured HCHO columns. Here we use 30% as a rough estimate of error in this term (**Palmer2006**), since measurements over Australia are lacking.

Uncertainty of HCHO background

OMI Vertical columns are corrected using background (or reference sector) measurements. This is to account for instrument degradation, and adds some uncertainty to the column. In this work vertical columns are corrected using reference sector measurements combined with modelled HCHO, which is described in Chapter 2 (Section 2.6.5).

Uncertainty from these three sources is required to give a representation of the confidence in vertical column amounts. These sources of error can be reduced through spatial and temporal averaging, as they are assumed to be unbiased. We use the uncertainty in each pixel averaged over our $2^\circ \times 2.5^\circ$ grid squares for each day or month, and divided by the square root of the number of pixels.

Fitting error is the error in the slant column (ΔSC), which is combined with our assumed relative AMF error ($\frac{\Delta AMF}{AMF}$) of 30%, and background error. $\Delta\Omega_{OMI}$ and $\Delta\Omega_{OMI,0}$ are calculated using the quadrature rules as follows:

$$\begin{aligned}\Omega &= \frac{SC - RSC}{AMF} \\ \Delta(SC - RSC) &= (SC - RSC) \sqrt{\left(\frac{\Delta SC}{SC}\right)^2 + (0.3)^2} \\ \Delta\Omega &= \Omega \sqrt{\left(\frac{\Delta(SC - RSC)}{(SC - RSC)}\right)^2 + (0.3)^2}\end{aligned}\quad (3.17)$$

The RSC term comes from Chapter 2 Equation 2.23 for calculating the corrected Ω .

Figure TODO shows the multi-year monthly average of the satellite uncertainty for each sub-region and all of Australia. TODO analysis.

Uncertainty in satellite HCHO ($\Delta\Omega$) from literature and calculated here is listed in Table 3.5. **DeSmedt2012** found satellite HCHO uncertainty to be 30 – 40% for the GOME-2 instrument by combining slant column systematic and random errors. For mid latitude winters they saw an excess of 60% uncertainty. OMI measurements will have similar amounts of uncertainty; however, the array of detectors provide more pixels which can be averaged to reduce this uncertainty.

In order to calculate the bias or systematic error, an understanding of biases in the underlying terms is required, since there is little in the way of comparable measurements. OMHCHO has been seen to underestimate observed HCHO by up to 40%, although this amount is regionally dependent (**Zhu2016; DeSmedt2015; Barkley2013**). TODO: list regions and their overestimates. Satellite HCHO may also suffer from ~ 13% overestimation when taking a monthly average due to only measuring on relatively cloud-free days (**Surl2018**). Since our a posteriori is linearly related to the satellite HCHO, any bias is directly transferred. We note therefore that our product may be biased by up to 40% due to satellite bias. This may be complicated further if

TABLE 3.5: Uncertainties in satellite total column HCHO.

uncertainty	location	source	notes
40%	North America	(a)	mostly due to cloud interference
26%	North America	(b)	with cloud fraction less than 20%
30-40%	global	(c)	GOME-2 instrument
> 60%	Mid-latitude	(c)	GOME-2 instrument in winters
X%-Y%	Australia	this work	monthly uncertainty at $2^\circ \times 2.5^\circ$

a: **Millet2006; Palmer2006**

b: **Millet2008**

c: **DeSmedt2008; DeSmedt2012**

the satellite bias over Australia does not match the bias over the remote Pacific at corresponding latitudes. However we can not quantify bias over Australia due to insufficient measurements. GEOS-Chem biases would affect the recalculation of HCHO, but they are assumed to be minimal in this thesis. Since we lack sufficient suitable measurements to estimate satellite bias over Australia, we assume bounds on potential satellite underestimation of up to 40%.

3.4.3.1 Fitting error from the OMI retrieval

Provided with the OMI product is the measurement of uncertainty in each pixel, calculated by the Smithsonian Astrophysical Observatory from the backscattered solar radiation fit (**Abad2015; Abad2016**).

3.4.3.2 Uncertainty in AMF calculations

Palmer2006 calculate the error in AMF through combining estimates of error in the UV albedo database ($\sim 8\%$), model error based on in situ measurements, cloud error (20 – 30%) (**Martin2003**), and aerosol errors (< 20%), totalling AMF error of around $\sim 30\%$ (calculated in quadrature). Compare this error estimate with that of **Curci2010**, where the error in AMF calculations and background columns are respectively found to be 30% and 15% based on their analysis of CHIMERE. **Millet2008** also examine this uncertainty and determine an overall uncertainty (1σ) of 25 – 27% in HCHO vertical columns with calculated AMFs where cloud fraction < 0.2 .

3.4.3.3 Uncertainty of HCHO background

One method of finding background satellite uncertainty involves using the variance over the remote Pacific (15°S to 15°N , and 180°W to 120°W) (**DeSmedt2012; DeSmedt2015**). In the remote Pacific, it can be assumed that HCHO variations are weak, with concentrations remaining steady in the short term (~ 1 month). Since monthly averaged reference sector HCHO is used as a vertical column correction, monthly variance in remote ocean HCHO can be used as an estimate of the background uncertainty. Monthly variance of GEOS-Chem vertical columns averaged over the remote pacific (15°S to 15°N , and 180°W to 120°W) is from X to Y TODO, and looking at just the Australian

latitudes (10°S to 50°S) the range is XX to YY TODO. For simplicity, uncertainty from the HCHO background is assumed to be constant in this work. Further simplification is made by assuming that this background error does not change when calculating a new reference sector correction (*RSC*, see Section 2.6.5).

Uncertainty in the *RSC* term is assumed to be 30%, based on the monthly variance seen at Australian latitudes and prior work (Curci2010). This is near to the 15% background error assumed in Curci2010 following (Dufour2009). Error in the *RSC* is based on this background uncertainty, following the equivalence of Equation 2.23 ($VCC = \frac{(SC - RSC)}{AMF}$) and Equation 2.24 ($VCC = \frac{(SC - SC_0)}{AMF} + \Omega_{GC,0}$).

3.4.4 Sensitivity to vertical column recalculation

Here we examine the sensitivity of the top-down isoprene emission estimations (E_{OMI}) to the AMF recalculation method. The a posteriori emissions change linearly with recalculated vertical columns, which are calculated in three different ways: using product provided AMF (AMF_{OMI}), recalculating AMF shape factors but keeping the original scattering weights (AMF_{GC}), or recalculating both shape factors and scattering weights (AMF_{PP}).

Figure todo shows the emissions over Australia averaged within January 2005. TODO: analysis of differences

TODO: Figure goes here.... A priori emissions and a posteriori estimates using satellite pixels recalculated using the different AMFs. Row 1: calculation using corrected vertical columns that implement the original OMI AMF. Row 2: as row 1, updating shape factors to use GEOS-Chem output. Row 3: as rows 1 and 2, using the code from Paul Palmer's group to completely recalculate the AMF.

Figure todo shows emissions over time from a single grid square, estimated by MEGAN (black) and the three top-down estimates, using $2^\circ \times 2.5^\circ$ horizontal resolution.

3.4.5 Sensitivity to filtering

Figure TODO shows emissions estimates for January 2005 with and without filtering for anthropogenic and pyrogenic influences TODO: update plot and Analysis

3.4.6 Summary

Figure TODO shows the average summer and winter random uncertainty over Australia, along with a time series of total emissions and the error bars using monthly averaged data and MEGAN emissions for reference. TODO brief analysis of spatial uncertainty stuff

3.5 Conclusions and implications

TODO: Brief summary of results and implications...

Our top down biogenic isoprene emission estimate shows that MEGAN overestimates emissions in summer by up to a factor of 5. The overestimation is spatially

and temporally diverse and leads to model overestimation of HCHO. By applying a monthly gridded scaling factor based on the multi-year monthly difference model emissions and our top down a posteriori, HCHO bias can be improved from 50-120% down to 30-50% (TODO: update that number for summer season rather than the 2005 January it is based on).

Global and Australian emissions estimates for isoprene still range widely due to the emission models sensitivities to inexact parameters including leaf area indices, plant functional type emission factors, and meteorological factors, with the full story behind the decrease seen here in a posteriori emissions not yet clear. It is likely a mixture of poorly defined emission factors (Emmerson2016), unaccounted for soil moisture (Sindelarova2014) and poorly understood responses to meteorological stresses in Australian forests (TODO: cite papers talking about clay and poor moisture responses). Sindelarova2014 see a 41 Tg yr^{-1} decrease in Australia when introducing soil moisture parameterisation.

Model overestimation of biogenic isoprene emissions also lowers simulated surface ozone concentrations by several percent. This improves our understanding of biogenic emissions impact on surface ozone levels, and could improve certainty in simulated ozone exceedances.

What else?

Chapter 4

Stratospheric ozone intrusions

4.1 Foreword

This chapter is unchanged (excepting titles, labels, and numbering) from work published by me into the atmospheric chemistry and physics journal: **Greenslade2017**.

4.2 Introduction

Tropospheric ozone constitutes only 10% of the total ozone column but is an important oxidant and greenhouse gas which is toxic to life, harming natural ecosystems and reducing agricultural productivity. Over the industrial period, increasing tropospheric ozone has been estimated to exert a radiative forcing (RF) of 365 mWm^{-2} (**Stevenson2013**), equivalent to a quarter of the CO₂ forcing (**IPCC_Chapter2**). While much tropospheric ozone is produced photochemically from anthropogenic and natural precursors, downward transport from the ozone-rich stratosphere provides an additional natural source of ozone that is particularly important in the upper troposphere (**Jacobson2000**). The contribution of this source to overall tropospheric ozone budgets remains uncertain (**Skerlak2014**), especially in the southern hemisphere (SH). Models show that stratospheric ozone depletion has propagated to the upper troposphere (**Stevenson2013**). However, work based on the Southern Hemisphere Additional OZonesonde (SHADOZ) network suggests stratospheric mixing may be increasing upper tropospheric ozone near southern Africa (**Liu2015; Thompson2014**). Uncertainties in the various processes which produce tropospheric ozone limit predictions of future ozone-induced radiative forcing. Here we use a multi-year record of ozonesonde observations from sites in the southern hemisphere extra-tropics, combined with a global model, to better characterise the impact of stratospheric ozone on the tropospheric ozone budget in the southern hemisphere.

Stratosphere-to-troposphere transport (STT) primarily impacts the ozone budget in the upper troposphere but can also increase regional surface ozone levels above the legal thresholds set by air quality standards (**Danielson1968; Lelieveld2009; Lefohn2011; Langford2012; Zhang2014; Lin2015**). In the western US, for example, deep STT events during spring can add 20-40 ppbv of ozone to the ground-level ozone concentration, which can provide over half the ozone needed to exceed the standard set by the U.S. Environmental Protection Agency (**Lin2012; Lin2015**). Another hotspot for STT is the Middle East, where surface ozone exceeds values of 80 ppbv in summer, with a stratospheric contribution of 10 ppb (**Lelieveld2009**). Estimates of the overall contribution

of STT to tropospheric ozone vary widely (**Galani2003; Stohl2003; Stevenson2006; Lefohn2011**). Early work based on two photochemical models showed that 25-50% of the tropospheric ozone column can be attributed to STT events globally, with most contribution in the upper troposphere (**Stohl2003**). In contrast, a more recent analysis of the Atmospheric Chemistry and Climate Model Inter-comparison Project (ACCMIP) simulations by **Young2013** found that STT is responsible for $540 \pm 140 \text{ Tg yr}^{-1}$, equivalent to $\sim 11\%$ of the tropospheric ozone column, with the remainder produced photochemically (**Monks2015**). This wide range in model estimates exists in part because STT is challenging to accurately represent, and finer model resolution is necessary to simulate small scale turbulence. Observation-based process studies are therefore key in determining the relative frequency of STT events, with models then able to quantify STT impact over large regions. Ozonesondes are particularly valuable for this purpose as they provide multi-year datasets with high vertical resolution.

Lower stratospheric and upper tropospheric ozone concentrations are highly correlated, suggesting mixing across the tropopause mainly associated with the jet streams over the Atlantic and Pacific oceans (**Terao2008**). Extra-tropical STT events most commonly occur during synoptic-scale tropopause folds (**Sprenger2003; Tang2012; Frey2015**) and are characterised by tongues of high potential vorticity (PV) air descending to lower altitudes. As these tongues become elongated, filaments disperse away from the tongue and mix irreversibly into the troposphere. STT can also be induced by deep overshooting convection (**Frey2015**), tropical cyclones (**Das2016**) and mid-latitude synoptic scale disturbances (**Stohl2003; Mihalikova2012**). STT events have been observed in tropopause folds around both the polar front jet (**Vaughan1994; Beekmann1997**) and the subtropical jet (**Baray2000**). The summertime pool of high tropospheric ozone over the eastern Mediterranean (EM) is mainly attributed to the downward ozone transport, as a result of the enhanced subsidence (**Zanis2014**) and the tropopause fold activity (**Akritidis2016**) over the region. The EM exhibits a summer maximum of subsidence, which sits between 20°E and 35°E and 31°N to 39°N , while zonally most subtropical tropopause folds occur during winter (**Tyrlis2014**). They are also observed near cut-off lows (**Price1993; Wirth1995**), so both regional weather patterns and stratospheric mixing are important to understand for STT analysis.

Stratospheric ozone intrusions undergo transport and mixing, with up to half of the ozone diffusing within 12 hours following descent from the upper troposphere (**Trickl2014**). The long range transport of enhanced ozone can be facilitated by upper tropospheric winds, with remarkably little convective mixing, as shown by **Trickl2014** who measure STT air masses two days and thousands of kilometres from their source. **Cooper2004** shows how STT advection can transport stratospheric air over long distances, with a modelled STT event spreading from the northern Pacific to the East coast of the USA over a few days.

The strength (ozone enhancement above background levels), horizontal scale, vertical depth, and longevity of these intruding ozone tongues vary with wind direction and strength, topography, and season. While the frequency, seasonality, and impacts of STT events have been well described in the tropics and northern hemisphere (NH), observational estimates from the SH extra-tropics are noticeably absent in the literature. The role of STT in the SH remains highly uncertain due to the more limited data availability compared to the NH and the temporal sparsity of these datasets (**Mze2010**;

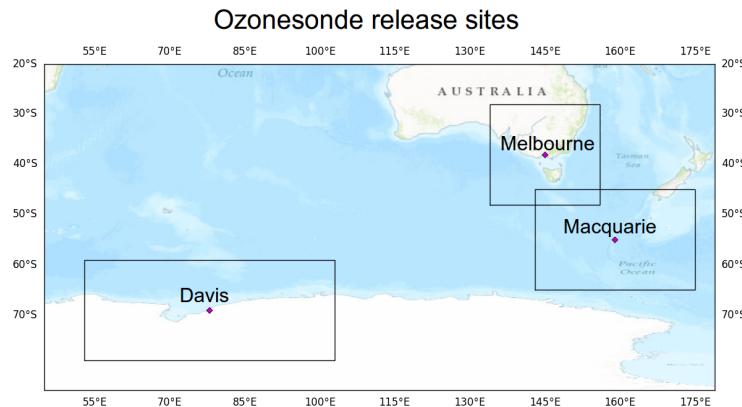


FIGURE 4.1: Ozonesonde release sites and the regions used to examine STT effect on tropospheric ozone levels.

Thompson2014; Liu2015).

Here, we characterise the seasonal cycle of STT events and quantify their contribution to the SH extra-tropical tropospheric ozone budget using nearly a decade of ozonesonde observations from three locations around the Southern Ocean spanning latitudes from 38°S-69°S. In Section 4.3 we describe the observations and methods used to identify STT events and to relate STT occurrence to meteorological events. We show how possible biomass burning smoke plume influence is detected and handled, and we introduce the GEOS-Chem model which is used for ozone flux estimation. Within Section 4.4 we show the seasonality, altitude, depth, and frequency of detected STT events, along with a comparison of our findings to other literature where possible. In Section 4.5 we analyse how well GEOS-Chem captures the tropospheric ozone seasonality and quantity near our three sites. In Section 4.6 an extrapolation of the STT detection frequencies along with GEOS-Chem monthly tropospheric ozone columns is used to estimate STT ozone flux near our three sites. We also compare and contrast our results against relevant literature. Finally, in Section 4.7 we examine in detail the uncertainties involved in our STT event detection technique and ozone flux estimations.

4.3 Data and Methods

4.3.1 Ozonesonde record in the Southern Ocean

Ozonesondes provide a high vertical resolution profile of ozone, temperature, pressure, and humidity from the surface and up to 35 km. In the troposphere, the ozonesondes generally perform 150-300 measurements. Ozone mixing ratio is quantified with an electrochemical concentration cell, using standardised procedures when constructing, transporting, and releasing the ozonesondes (<http://www.ndsc.ncep.noaa.gov/organize/protocols/appendix5/>). Ozonesondes are estimated to provide around 2% precision in the stratosphere, which improves at lower altitudes, and ozonesondes have been shown to be accurate to within 5% when the correct procedures are followed (Smit2007).

TABLE 4.1: Number of sonde releases at each site over the period of analysis.

Site	Total Releases	Monthly Releases (J, F, M, ...)	Date Range
Davis	240	11, 12, 13, 12, 17, 31, 29, 28, 32, 28, 15, 12	2006/04/13 - 2013/11/13
Macquarie Island	390	32, 31, 45, 28, 34, 33, 28, 35, 29, 33, 31, 31	2004/01/20 - 2013/01/09
Melbourne	456	31, 38, 40, 38, 41, 36, 38, 39, 46, 40, 38, 31	2004/01/08 - 2013/12/18

Ozonesondes are launched approximately weekly from Melbourne (38° S, 145° E), Macquarie Island (55° S, 159° E) and Davis (69° S, 78° E), as shown in Fig. 4.1. Melbourne is a major city in the south east of Australia, and may be affected by anthropogenic pollution in the lower troposphere. Macquarie Island is isolated from the Australian mainland, situated in the remote Southern Ocean and unlikely to be affected by any local pollution events. Davis is on the coast of Antarctica and also unlikely to experience the effects of anthropogenic pollution.

For this study, we use the 2004-2013 data for Melbourne and Macquarie Island and the 2006-2013 data for Davis because both ozone and geopotential height (GPH) are available from the World Ozone and Ultraviolet Data Centre archived data in these periods. At Davis, ozonesondes are launched twice as frequently during the ozone hole season and preceding months (June-October) as at other times of year (Alexander2013). A summary of ozonesonde releases at each site can be seen in Table 4.1.

Characterisation of STT events requires a clear definition of the tropopause. Two common tropopause height definitions are the standard lapse rate tropopause (WMO1957) and the ozone tropopause (Bethan1996). The lapse rate tropopause is defined as the lowest altitude where the lapse rate (vertical gradient of temperature) is less than $2^{\circ}\text{C km}^{-1}$, provided the lapse rate averaged between this altitude and 2 km above is also below $2^{\circ}\text{C km}^{-1}$. The ozone tropopause is defined as the lowest altitude satisfying the following three conditions for the ozone mixing ratio (OMR) (Bethan1996):

1. Vertical gradient of OMR is greater than 60 ppb km^{-1} ;
2. OMR is greater than 80 ppb; and
3. OMR exceeds 110 ppb between 500 m and 2000 m above the altitude under inspection (modified to between 500 m and 1500 m in the Antarctic, including the site at Davis).

The ozone tropopause may misdiagnose the real tropopause altitude during stratosphere-troposphere exchange; however, it is useful at polar latitudes in winter, where the lapse-rate definition may result in artificially high values for tropopause height (Bethan1996; Tomikawa2009; Alexander2013). We require lapse rate defined tropopauses to be at a minimum of 4 km altitude. Another commonly used tropopause definition is determined with the use of PV (dynamical tropopause). In the extra-tropics the isosurface where $\text{PV} = 2 \text{ PVU}$ ($1 \text{ PVU} = 10^{-6} \text{ m}^2 \text{ s}^{-1} \text{ K kg}^{-1}$) is often used to define the

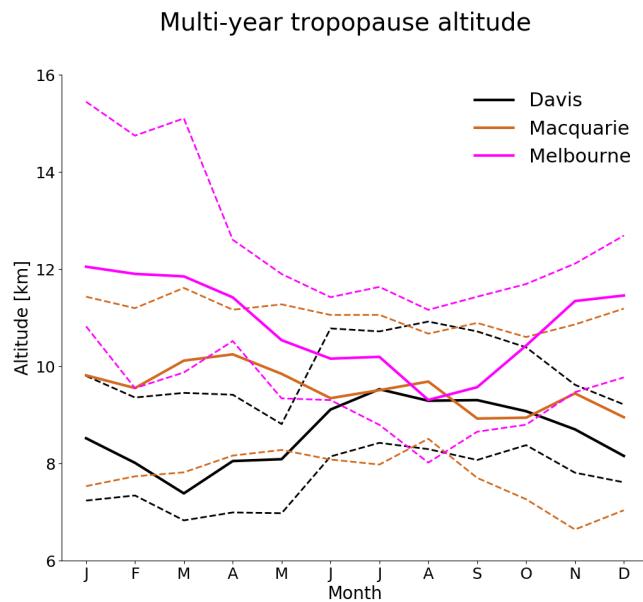


FIGURE 4.2: Multi-year monthly median tropopause altitude (using the ozone defined tropopause) determined from ozonesondes measurements at Davis (2006-2013), Macquarie Island (2004-2013), and Melbourne (2004-2013) (solid lines). Dashed lines show the 10th to the 90th percentile of tropopause altitude for each site.

tropopause, allowing the 3D representation of tropopause folds and other tropopause features in a sufficiently resolved model (**Skerlak2014; Tyrlis2014**). The PV is not calculable using the ozonesonde measurements alone, so in this work the ozone tropopause is used when determining STT events or measured tropopause altitude.

Figure 4.2 shows the monthly median ozone tropopause altitudes at each location (solid lines). At Melbourne, the tropopause altitude displays a seasonal cycle with maximum in summer and minimum in winter. This seasonality is missing at Macquarie Island and almost reversed at Davis, which has a minimum during autumn and maximum from winter to spring. Tropopause altitude decreases with latitude from 9-14 km at Melbourne (38° S) to 7-9 km at Davis (69° S).

Figure 4.3 shows multi-year averaged ozone mixing ratios measured by ozonesonde over the three stations. Over Melbourne, increased ozone extending down through the troposphere is apparent from December to March and from September to November. The increased tropospheric ozone in these months is due to STT (in summer), and possible biomass burning influence (in spring), both discussed in more detail in the following sections. Over Davis and Macquarie Island, tropospheric ozone is higher between March and October, although the seasonal differences are small compared to those at Melbourne. The seasonality shown in Fig. 4.3 for Davis is consistent with remote free tropospheric photochemistry determined by solar radiation availability and temperature, resulting in higher ozone in winter (**Lelieveld2000**). NO₂ stratospheric observations have been conducted in the Southern hemisphere at Lauder, Macquarie

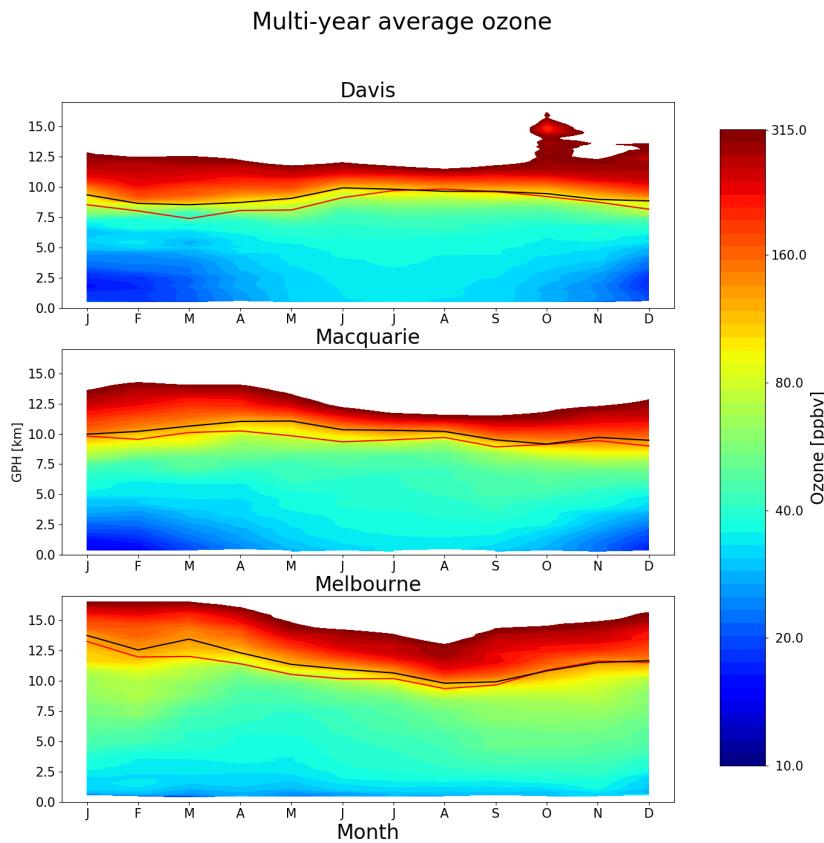


FIGURE 4.3: Multi-year mean seasonal cycle of ozone mixing ratio over Davis, Macquarie Island, and Melbourne as measured by ozonesondes. Measurements were interpolated to every 100 m and then binned monthly. Black and red solid lines show median ozone and lapse-rate defined tropopause altitudes (respectively), as defined in the text.

Island and Arrival Heights (**Struthers2004**) which displays a winter minima in seasonality consistent with an ozone maxima. Influence from the ozone hole can be seen over Davis in October, with relatively low ozone levels extending up 5-6 km into the stratosphere.

4.3.2 Model description

To provide regional and global context to the ozonesonde observations, we use the GEOS-Chem version 10-01 global chemical transport model (**Bey2001**), which simulates ozone along with more than 100 other trace gases throughout the troposphere and stratosphere. Stratosphere-troposphere coupling is calculated using the stratospheric unified chemistry extension (UCX) (**Eastham2014**). Transport is driven by assimilated meteorological fields from the Goddard Earth Observing System (GEOS-5) maintained by the Global Modeling and Assimilation Office (GMAO) at NASA. Ozone precursor emissions are from the Model of Emissions of Gases and Aerosols from

Nature (MEGAN) version 2.1 (**Guenther2012**) for biogenic emissions, the Emissions Database for Global Atmospheric Research (EDGAR) version 4.2 for anthropogenic emissions, and the Global Fire Emissions Database (GFED4) inventory (**Giglio2013**) for biomass burning emissions. Our simulation was modified from the standard v10-01 to fix an error in the treatment of ozone data from the Total Ozone Mapping Spectrometer (TOMS) satellite used to calculate photolysis (see http://wiki.seas.harvard.edu/geos-chem/index.php/FAST-JX_v7.0_photolysis_mechanism#Fix_for_TOMS_to_address_strange_cycle_in_OH_output.).

Our simulations span 2005-2012 (following a 1-year spin-up) with horizontal resolution of 2° latitude by 2.5° longitude and 72 vertical levels from the surface to 0.01 hPa. The vertical resolution is finer near the surface at ~ 60 m between levels, spreading out to ~ 500 m near 10 km altitude. For comparison to the ozonesonde observations, the model state was saved every 6 hours within the grid boxes containing each site. When comparing against ozonesondes, GEOS-Chem UTC+0 time samples are used for all sites. This means that the simulated ozone profiles are analysed at local times of 7AM for Davis, and 11AM for Macquarie Island and Melbourne. GEOS-Chem uses the tropopause height provided by GEOS-5 meteorological fields, which are calculated using a lapse-rate tropopause definition using the first minimum above the surface in the function $0.03 \times T(p) - \log(p)$, with p in hPa (**Rienecker2008**).

4.3.3 Characterisation of STT events and associated fluxes

We characterise STT events using the ozonesonde vertical profiles to identify tropospheric ozone enhancements above a local background (in moles per billion moles of dry air, referred to here as ppb). The process is illustrated in Figure 4.4 on an example ozone profile. First, the ozone vertical profiles are linearly interpolated to a regular grid with 20 m resolution from the surface to 14 km altitude. Small vertical-scale fluctuations in ozone, which are captured by the high-resolution ozonesondes, can be regarded as sinusoidal waves superimposed on the large vertical scale background tropospheric ozone. As such, the interpolated profiles are bandpass-filtered using a fast Fourier transform (**Press1992**) to retain these small vertical scales, between 0.5 km and 5 km (removing low and high frequency perturbations). The high frequency perturbations are removed as they may represent noise in the measurements. The perturbations with scales longer than 5 km represent the vertical gradient of ozone concentration from the surface to the stratosphere. In what follows, these filtered vertical profiles are referred to as perturbation profiles.

For an event to qualify as STT, a clear increase above the background ozone level is needed, as a bandpass filter leaves us with enhancements minus any noise or seasonal scale vertical profile effects. We next use all the perturbation profiles at each site to calculate the 95th percentile perturbation value for the site. The threshold is calculated from all the interpolated filtered values between 2 km above the surface and 1 km below the tropopause. This is our threshold for tropospheric ozone perturbations, and any profiles with perturbations exceeding this value in individual ozonesondes are classified as STT events. STT events at altitudes below 4 km are removed to avoid surface pollution, and events within 0.5 km of the tropopause are removed to avoid false positives induced by the sharp transition to stratospheric air. We note that this ozone detection methodology detailed above does not allow us to resolve STT events

Ozone at Melbourne on 2004/01/08

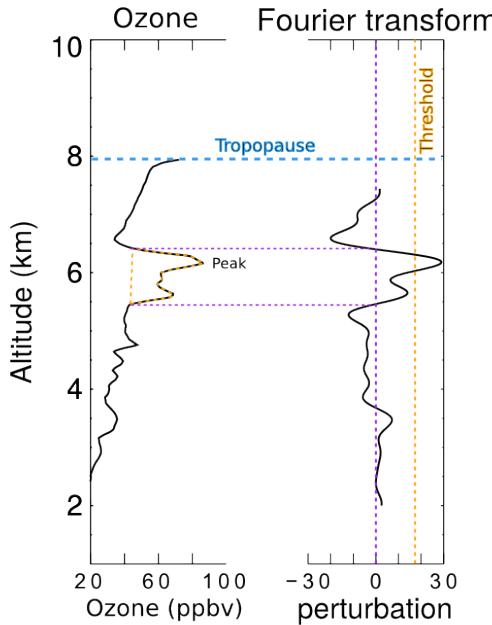


FIGURE 4.4: An example of the STT identification and flux estimation methods used in this work. The left panel shows an ozone profile from Melbourne on 8 January 2004 from 2 km to the tropopause (blue dashed horizontal line). The right panel shows the perturbation profile created from bandpass filtering of the mixing ratio profile. The STT occurrence threshold calculated from the 95th percentile of all perturbation profiles is shown as the orange dashed line, and the vertical extent of the event is shown with the purple dashed lines (see details in text). The ozone flux associated with the STT event is calculated using the area outlined with the orange dashed line in the left panel.

where the ozone flux is spread diffusely across the troposphere without a peak-like structure in the ozonesonde profile. In other words, STT events which might have occurred some distance and time away from the location of the ozonesonde profiles may not be readily detected using the high vertical resolution, but infrequent, ozonesonde launches.

We define the ozone peak as the altitude where the perturbation profile is greatest between 2 km from the surface and 0.5 km below the tropopause. The STT event is confirmed if the perturbation profile drops below zero between the ozone peak and the tropopause, as this represents a return to non-enhanced ozone concentrations. Alternatively, the STT event is also confirmed if the OMR between the ozone peak and the tropopause drops below 80 ppb and is at least 20 ppb lower than the OMR at the ozone peak. If neither of these conditions are met, the profile is rejected as a non-event. This final step removes near-tropopause anomalies for which there is insufficient evidence of detachment from the stratosphere. Vertical ozone profiles recorded by ozonesondes are highly dependent on the time of launch ([Sprenger2003](#)), and it cannot be guaranteed that detected ozone enhancements are fully separated from the stratosphere,

although this method minimises that risk by removing detected events too near the tropopause.

We estimate the ozone flux into the troposphere associated with each event by integrating the ozone concentration enhancement vertically over the altitude range for which an STT event is identified (i.e. enhancement near the ozone peak over which the perturbation profile is greater than zero). This estimate is conservative because it does not take into account any ozone enhancements outside of the detected peak that may have been caused by the STT, and also ignores any enhanced ozone background amounts from synoptic-scale stratospheric mixing into the troposphere.

Our method differs somewhat from that used by [Tang2010](#) to detect STT events from ozonesonde measurements. Their definition is based on subjective analysis of sondes released from 20 stations ranging in latitude from 35° S to 40° N. They identify an STT event if, starting from 5 km altitude, ozone exceeds 80 ppb and then within 3 km decreases by 20 ppb or more to a value less than 120 ppb. Their technique would miss many events due to the lower ozone concentrations found in the cleaner Southern Hemisphere.

4.3.4 Biomass burning influence

The STT detection algorithm described in Sect. [4.3.3](#) assumes all ozone enhancements are caused by stratospheric intrusions. In some cases, however, these perturbations may in fact reflect ozone production in lofted smoke plumes. Biomass burning in southern Africa and South America has previously been shown to have a major influence on atmospheric composition in the vicinity of our measurement sites ([Oltmans2001](#); [Gloudemans2006](#); [Edwards2006](#)), particularly from July to December ([Pak2003](#); [Liu2017](#)). On occasion, smoke plumes from Australian and Indonesian fires can also reach the mid-high southern latitudes, as seen from satellite measurements of carbon monoxide (CO) discussed below.

Large biomass burning events emit substantial quantities of ozone precursors, some of which are capable of being transported over long distances and driving ozone production far from the fire source ([Jaffe2012](#)). Ozone production from biomass burning is complex and affected by photochemistry, fuel nitrogen load, and time since emission, among other factors. While ozone production occurs in some biomass burning plumes, this is not always the case; therefore ozone perturbations detected during transported smoke events may or may not be caused by the plume. For this reason all detected STT events which could be caused by smoke plumes are flagged, following the procedure outlined below. Calculations of seasonality, and ozone flux do not include flagged events, however they are included in summary plots in this work.

Possible biomass burning influence is identified using satellite observations of CO from the AIRS (Atmospheric Infra-red Sounder) instrument on board the Aqua satellite ([AIRS3STD](#)). CO is emitted during incomplete combustion and is an effective tracer of long-range transport due to its long lifetime ([Edwards2003](#); [Edwards2006](#)). In the Southern Hemisphere, biomass burning is the primary source of CO, making CO a good proxy for fire plumes ([Sinha2004](#); [Mari2008](#)). To identify possible biomass burning influence, AIRS vertical column CO is visually inspected for all dates with detected STT events. Smoke plumes are diagnosed over areas with elevated CO columns ($\sim 2 \times 10^{18}$ molecules cm^{-2} or higher), and any sonde-detected STT event that occurs

near (within \sim 150 km of) a smoke plume is flagged. Removal of these detections reduces the yearly estimated ozone flux by $\sim 15\%$ at Macquarie Island and $\sim 20\%$ at Melbourne.

All days with detected STT events were screened, with the exception of one event during which there were no available AIRS data (January 2010). We find that biomass burning may have influenced 27 events over Melbourne and 21 events over Macquarie Island. These events are flagged in the following sections, and are not used in our calculation of total STT flux. All of the flagged events except for two occurred during the SH burning season (July to December). No events at Davis were seen to be influenced by smoke transport.

4.3.5 Classifying synoptic conditions during STT events

Synoptic scale weather patterns are examined using data from the European Centre for Medium-range Weather Forecasts (ECMWF) Interim Reanalysis (ERA-I) (Dee2011). This is done using the ERA-I data products over the three sites on dates matching the detected STT events. We use the ERA-I 500 hPa data to subjectively classify the events based on their likely meteorological cause, by visually examining each date where an event was detected. During STT occurrence, the upper troposphere is typically characterised by nearby cyclones, cut-off lows, or cold fronts. Over Melbourne and Macquarie Island, we find that frontal and low pressure activity are prevalent during STT events (see Sect. 4.4). Over Davis, the weather systems are often less clear, however we see a higher portion of probable cut-off lows. The stratospheric polar vortex may create tropopause folds without other sources of upper tropospheric turbulence such as low pressure fronts or cyclones (Baray2000; Sprenger2003; Tyrlis2014). Cut-off low pressure systems can be seen clearly in synoptic scale weather maps as regions with lowered pressure and cyclonic winds. Low pressure fronts in the higher southern latitudes travel from west to east and lower the tropopause height. We examine two cases in detail to illustrate the relationship between synoptic-scale conditions and STT events over Melbourne. These are included in a supplementary document (Fig. S2 and S3) which show an archetypal cut-off low and low pressure front. To detect cut-off low pressure systems we look for cyclonic winds and a detached area of low pressure within ~ 500 km of a site on days of event detection. For low pressure fronts we look for low pressure troughs within ~ 500 km. Frontal passage is a known cause of STT as stratospheric air descends and streamers of ozone-rich air break off and mix into the troposphere (Sprenger2003).

4.4 STT event climatologies

Figure 4.5 shows the seasonal cycles of STT frequency at Davis, Macquarie Island, and Melbourne. Frequency is determined as detected event count divided by total launched ozonesondes for each month. STT events in Figures 4.5-4.8 are coloured based on the meteorological classification described in Sect. 4.3.5, with events classified as either low pressure fronts (“frontal”, dark blue), cut-off low pressure systems (“cutoff” teal), or indeterminate (“misc”, cyan). Events that may have been influenced

TABLE 4.2: Total number of ozonesonde detected STT events, along with the number of events in each category (see text).

Site	Events	Cut-offs	Frontals	Misc	Fire
Davis	80	44	19	17	0
Macquarie Island	105	19	31	34	21
Melbourne	127	28	31	41	27

by transported smoke plumes (Sect. 4.3.4) are shown in red. Ozonesonde releases are summarised in Table 4.1 and detected event counts are summarised in Table 4.2.

There is an annual cycle in the frequency of STT events (Fig. 4.5) with a summertime peak at all three sites. This summertime peak is due to a prevalence of summer low-pressure storms and fronts, which increase turbulence and lower the tropopause (Reutter2015). At Davis, there are more STT detections during winter relative to our other sites, which may be due to the polar vortex and its associated lowered tropopause and increased turbulence. STT events associated with cut-off low pressure systems are more prevalent during summer, while STT events associated with frontal passage occur throughout the year. The high frequency of STT ozone enhancements is comparable to the > 25% frequencies seen over Turkey and east of the Caspian sea in an ERA-I analysis performed by Tyrlis2014.

The SH summer maximum we see for STT ozone flux can also be seen in Fig. 16 of Skerlak2014, which shows seasonal flux over the southern ocean, although this is less clear over Melbourne. This seasonality is not clear in the recent ERA-Interim tropopause fold analysis performed by Skerlak2015, where a winter maximum of tropopause fold frequency ($\sim 0.5\%$ more folds in winter) over Australia can be seen to the north of Melbourne. Their work seems to show slightly higher fold frequencies over Melbourne in summer (Skerlak2015), however not to the same extent that our summer peak suggests. Their winter maximum is in the subtropics only - from around 20° S to 40° S, which can be seen as the prevalent feature over Australia in their Fig. 5. Wauben1998 look at modelled (CTM driven by ECMWF output) and measured ozone distributions and find more SH ozone in the lower troposphere during austral winter, however they note that the ECMWF fields are uncertain here again due to lack of measurements. Their work shows a generally cleaner lower troposphere in the SH summer but this can not be construed to suggest more or less STT folds in either season. Sprenger2003 examine modelled STT folds using ECMWF output over March 2000 - April 2001, and show that for this year there is a clear austral winter maximum, again over the 20° S to 40° S band. The winter maximum does not include Melbourne, or the southern ocean, which explains why we see a seasonality not readily evident in these global-scale studies.

The measurement sites are not in the regions which have a clear winter maximum seen in Sprenger2003, and the large scale winter maximum shown by all three studies seems to be dominated by the system in that region. The seasonality of our three sites is not driven by the larger STT system seen over the southern Indian ocean and middle Australia which dominates prior analysis near or over Australia.

To examine the robustness of the distributions shown in Fig. 4.5, we developed an alternative assessment of the seasonal occurrence of STT events, with results shown

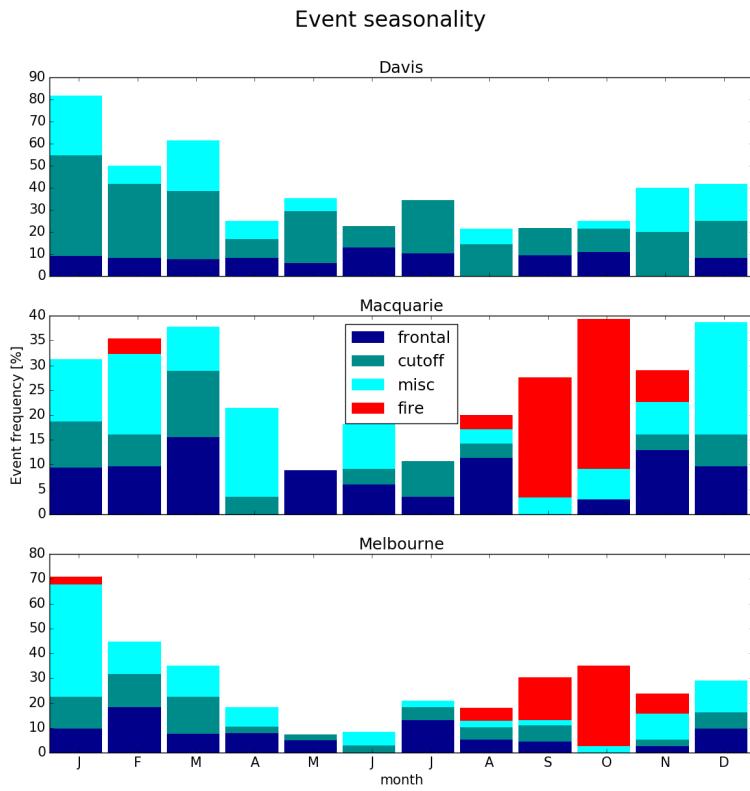


FIGURE 4.5: Seasonal cycle of STT event frequency at Davis (top), Macquarie Island (middle), and Melbourne (bottom). Events are categorised by associated meteorological conditions as described in the text, with low pressure fronts (“frontal”) in dark blue, cut-off low pressure systems (“cutoff”) in teal, and indeterminate meteorology (“misc”) in cyan. Events that may have been influenced by transported smoke plumes are shown in red (see text for details).

in Fig. 4.6. Here STT occurrence is evaluated by consideration of the square of the dry Brunt-Väisälä frequency (N^2) at the heights of the ozone tropopause (z_{OT}) and lapse rate tropopause (z_{LRT}) in each ozonesonde profile that has been binned to 100 m resolution. We use N^2 to assess atmospheric stability, which is normally distinctly higher in the stratosphere than in the troposphere, and assume that the vertical temperature gradients within the intrusion respond most rapidly to transported heat, which is an additional characteristic of stratospheric air. N^2 is evaluated using 250 m resolution data (to smooth variability in the vertical gradient of potential temperature that is due to small temperature fluctuations likely associated with gravity waves). The altitude binning chosen is a compromise between vertical resolution and the level of variability in N^2 introduced by temperature gradients associated with perturbations from gravity waves and changes near the lapse rate tropopause, and is the minimum that produces a robust seasonal distribution. We define STT as having taken place if $N^2(z_{OT}) > N^2(z_{LRT})$ and $z_{OT} < z_{LRT}$; in this way the characteristically higher static stability and ozone concentration of stratospheric intrusion is regarded as being retained

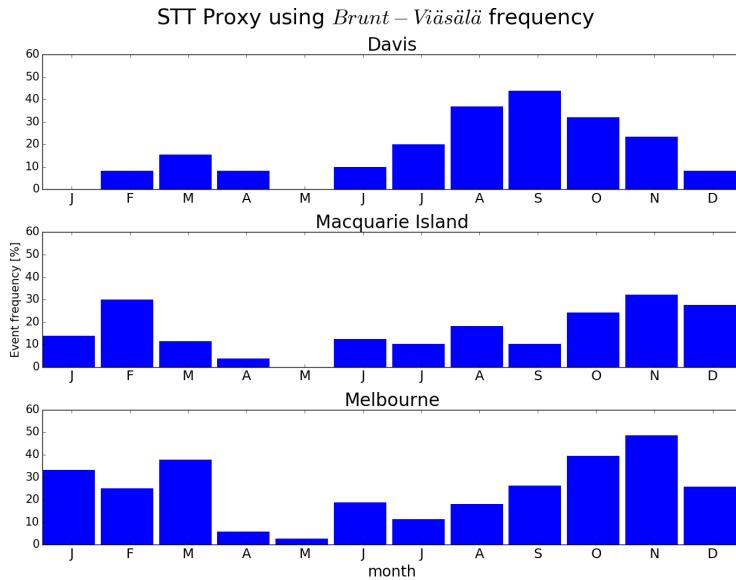


FIGURE 4.6: Seasonal distribution of STT events using the alternative STT proxy, obtained from consideration of the static stability at the ozone and lapse rate tropopauses, for Davis (2006-2013), Macquarie Island (2004-2013), and Melbourne (2004-2013).

as it penetrates below the lapse rate tropopause. The seasonal distributions shown for the three stations in Fig. 4.6 are generally similar to those shown in Fig. 4.5 (although detected events are less frequent), with the main exception that very few events are identified with the alternative method at Davis in the first half of the year. For our STT proxy, we only detect intrusions where the lowest altitude of the intrusion satisfies the ozone tropopause definition. During summer and autumn, the vertical ozone gradients at Davis are weaker compared with the other seasons, and the detected ozone tropopause tends to lie above the lapse rate tropopause potentially reducing the ability to identify STT events based on the definition of our proxy.

Figure 4.7 shows the altitudes of detected events, based on the altitude of peak tropospheric ozone (local maximum ozone within enhancement altitude) in the ozonesonde profile. STT event peaks most commonly occur at 6–11 km above Melbourne and anywhere from 4–9 km at Davis and Macquarie Island. There is no clear relationship between meteorological conditions and event altitude, which may reflect the fact that the ozonesondes observe a snapshot of an event at different stages of its life cycle.

Figure 4.8 shows the distance from the event peak to the ozone defined tropopause, referred to as event depth. The majority of STT events occur within 2.5 km of the tropopause at Davis and Macquarie Island. Over Melbourne, the event depth is more spread out, with peak ozone enhancement generally occurring up to 6 km below the tropopause. Again, there is no clear relationships between meteorological conditions and event depth.

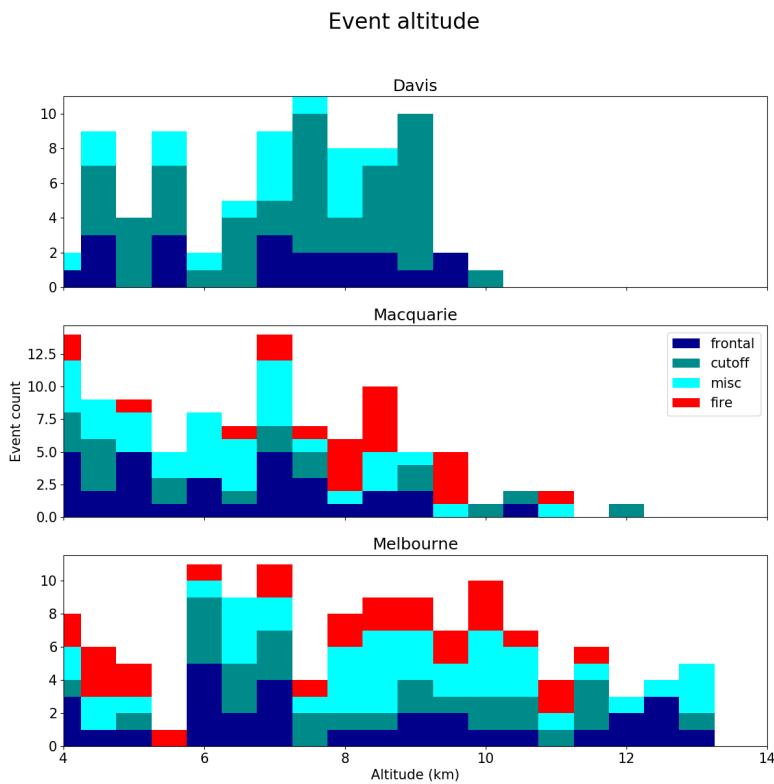


FIGURE 4.7: The distribution of STT events' altitudes at Davis (top), Macquarie Island (middle), and Melbourne (bottom), determined as described in the text. Events are coloured as described in Fig. 4.5.

4.5 Simulated ozone columns

Figure 4.9 compares the time series of tropospheric ozone columns (Ω_{O_3}) in molecules cm^{-2} simulated by GEOS-Chem (red) to the measured tropospheric ozone columns (black). GEOS-Chem outputs ozone density (molecules cm^{-3}), and height of each simulated box, as well as which level contains the tropopause, allowing modelled Ω_{O_3} to be calculated as the product of density and height summed up to the box below the tropopause level. In both observations and model, the maximum ozone column at Melbourne occurs in austral summer, with a minimum in winter, while Macquarie Island and Davis show the opposite seasonality.

GEOS-Chem provides a reasonable simulation of the observed seasonality and magnitude of Ω_{O_3} . Reduced major axis regression of observed versus simulated Ω_{O_3} gives a line of best fit with slopes of 1.08 for Davis, 0.99 for Macquarie Island, and 1.34 for Melbourne. The model is only partially able to reproduce the variability in the observations, with paired r^2 values of 0.38 for Davis, 0.18 for Macquarie Island, and 0.37 for Melbourne. Much of the variability is driven by the seasonal cycle, and after removing this effect (by subtracting the multi-year monthly means), the r^2 values decrease to 0.07, 0.11, and 0.30 respectively, although the slope improves at Melbourne

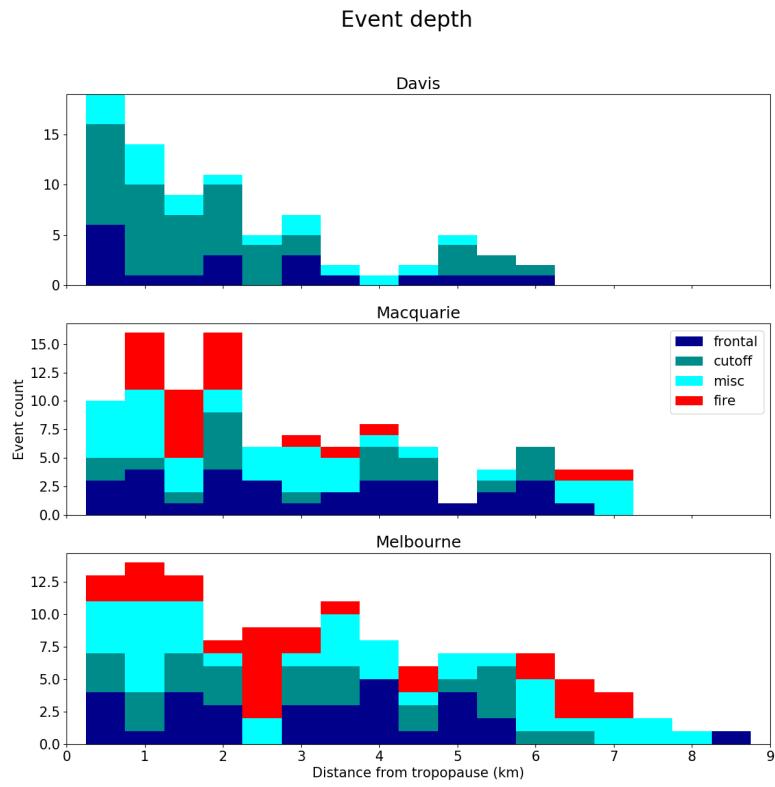


FIGURE 4.8: The distribution of STT events' depths, defined as the distance from the event to the tropopause, at Davis (top), Macquarie Island (middle), and Melbourne (bottom), determined as described in the text.

Events are coloured as described in Fig. 4.5.

to 1.08.

Figure 4.10 shows the observed and simulated ozone profiles at all sites, averaged seasonally. The model generally underestimates ozone in the lower troposphere (up to 6 km) over Davis, although this bias is less pronounced during summer. Over Melbourne, ozone in the lower troposphere is well represented, but the model overestimates ozone from around 4 km to the tropopause. Over Macquarie Island we see model overestimation of ozone above 4 km, as well as underestimated ozone in the lower troposphere, suggesting that this region is influenced by processes seen at both of our other sites. Also shown is the mean tropopause height simulated by the model (horizontal dashed red line), which is always higher than the observed average, although this difference is not statistically significant. The effect of local pollution over Melbourne during austral summer (DJF) can be seen from the increased mean mixing ratios and enhanced variance near the surface. The gradient of the O₃ profiles is steeper in the measurements than the model, at all sites during all seasons. Recently **Hu2017** examined GEOS-Chem ozone simulations and found a similar overestimation of upper troposphere ozone in the mid southern latitudes when using the GEOS5 meteorological fields.

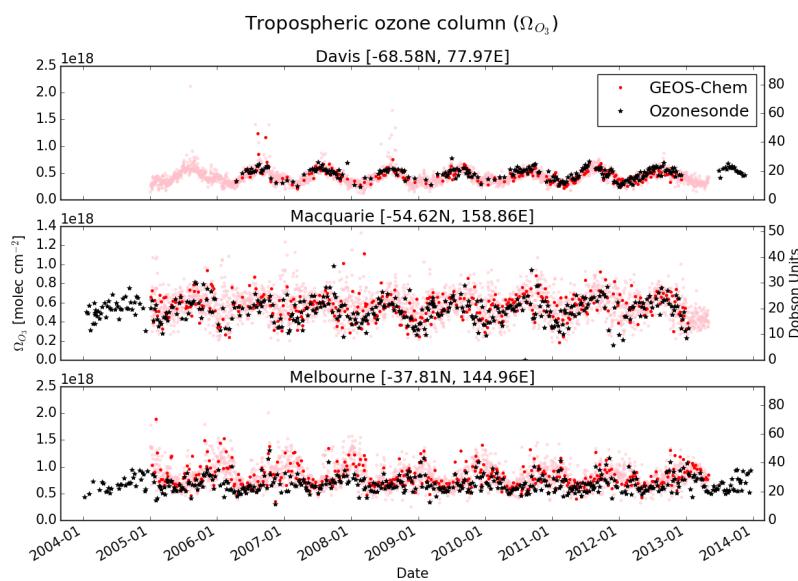


FIGURE 4.9: Comparison between observed (black) and simulated (pink, red) tropospheric ozone columns (Ω_{O_3} , in molecules cm^{-2}) from 1 January 2004 to 30 April 2013. For the model, daily output is shown in pink, while output from days with ozonesonde measurements are shown in red. For each site, the model has been sampled in the relevant grid square.

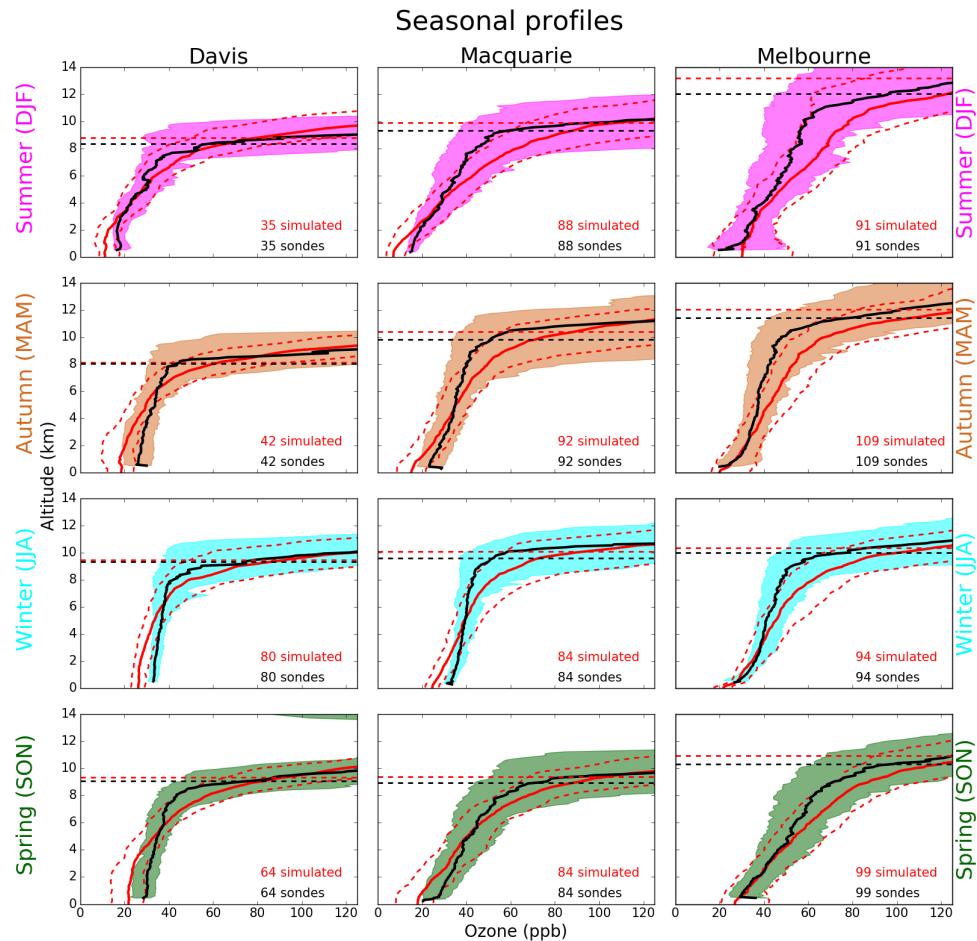


FIGURE 4.10: Observed and simulated tropospheric ozone profiles over Davis, Macquarie Island, and Melbourne, averaged seasonally. Model medians (2005-2013 average) are shown as red solid lines, with red dashed lines showing the 10th and 90th percentiles. Ozonesonde medians (over each season, for all years) are shown as black solid lines, with coloured shaded areas showing the 10th and 90th percentiles. The horizontal dashed lines show the median tropopause heights from the model (red) and the observations (black).

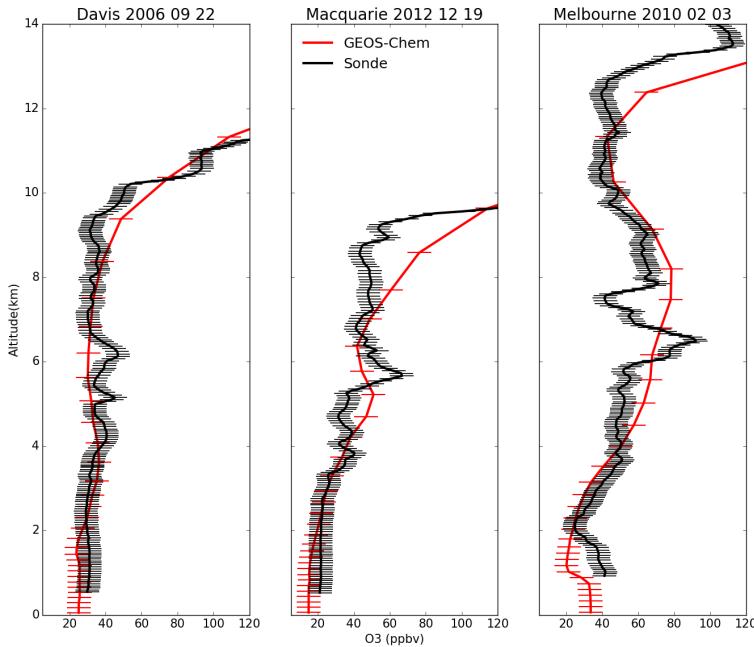


FIGURE 4.11: Example comparisons of ozone profiles from ozonesondes (black) and GEOS-Chem (red) from three different dates during which STT events were detected from the measurements. The dates were picked based on subjective visual analysis. The examples show the best match between model and observations for each site. GEOS-Chem and ozonesonde pressure levels are marked with red and black dashes respectively.

Figure 4.11 compares modeled (red) and observed (black) ozone profiles on three example days when STT events were detected using the ozonesondes. The figures show the profile for each site with the closest (qualitative) match between model and observations. The resolution (both vertical and horizontal) of GEOS-Chem in the upper troposphere is too low to consistently allow detection of STTs, although in a few cases (e.g., Melbourne in Fig. 4.11) it appears that the event was large enough to be visible in the model output.

4.6 Stratosphere-to-troposphere ozone flux from STT events

4.6.1 Method

We quantify the mean stratosphere-to-troposphere ozone flux due to STT events at each site based on the integrated ozone amount associated with each STT event (see Sect. 4.3.3). Events that may have been influenced by transported biomass burning are excluded from this calculation. Our estimate provides a preliminary estimate of how much ozone is transported from the stratosphere by the events detected by our method. The estimate is conservative for several reasons: it ignores secondary ozone

peaks which may also be transported from the stratosphere, it ignores potential ozone enhancements which may have dispersed and increased the local background mixing ratio, and any influence from STT events nearby which may also increase the local background ozone.

Observed tropospheric columns are calculated from the ozonesondes by calculating the ozone number density (molecules cm⁻³) using measured ozone partial pressure (P_{O_3}) and integrating vertically up to the tropopause:

$$\Omega_{O_3} = \int_0^{TP} \frac{P_{O_3}(z)}{k_B \times T(z)} dz$$

where z is the altitude (GPH), TP is the altitude at the tropopause, T is the temperature, and k_B is the Boltzmann constant.

Three regions are used to examine possible STT flux over a larger area using modeled tropospheric ozone concentrations. The regions are shown in Fig. 4.1. The regions are centred at each site, plus or minus ten degrees latitude, and plus or minus 25, 16, and 11 degrees longitude for Davis, Macquarie Island, and Melbourne respectively. These boundaries approximate a rectangle centred at each site with ~ 2000 km side lengths, covering ~ 4.4 , 4.6 , and 4.8 million square km, for Davis, Macquarie Island, and Melbourne respectively.

To determine the ozone column attributable to STT, we determine monthly averaged STT impact (I ; fraction of tropospheric ozone sourced from the stratosphere as shown above) and the monthly mean tropospheric ozone column (from the GEOS-Chem multi-year mean, Ω_{O_3}) over the regions described above. This can be expressed simply as the STT flux per event (flux_{*i*} in each month: $\text{flux}_i = \Omega_{O_3} \times I$). Next we determine how many events are occurring per month by assuming only one event can occur at one time, and that no event is measured twice. These assumptions allow a simple estimate of events per month from the relatively sparse dataset and should hold true as long as our regions of extrapolation are not too large. The (*P*)robability of any sonde launch detecting an event is calculated as the fraction of ozonesonde releases for which an STT event was detected, calculated for each month. We assume events last N days, then find how many events per month we expect by multiplying the days in a month by P and dividing by this assumed event lifetime. For example if we expect to see an event 25% of the time in a month, and events last one day, we expect one event every four days (~ 7.5 events in that month) whereas if we expect events to last a week then we would expect \sim one event in that month. This leads us to multiply our flux_i by P , and then by the term M ($M = \frac{\text{days per month}}{N}$) determined by our assumed event lifetime in order to determine monthly STT ozone flux.

The longevity of ozone events is very difficult to determine, and we have chosen 2 days as a representative number based on several examples in Lin2012 where intrusions were seen to last from 1-3 days (occasionally longer) and an analysis of one large event by Cooper2004 showing that most of the ozone had dispersed after 48 hours. Worth noting is the recent work of Trickl2014, where intrusions are detected > 2 days and thousands of kilometres away from their initial descent into the troposphere over Greenland or the Arctic. In those regions with high wind shear, mixing appears to be slower, which allows ozone intrusions to be transported further without dissipating into the troposphere. Relative uncertainty in our M term is set to 50%, as we assume

TABLE 4.3: Seasonal STT ozone contribution in the regions near each site, in $\text{kg km}^{-2} \text{ month}^{-1}$. In parentheses are the relative uncertainties.

Region	DJF	MAM	JJA	SON
Davis	54.5 (102%)	47.7 (97%)	30.7 (114%)	18.8 (127%)
Macquarie Island	61.3 (85%)	70.7 (91%)	17.9 (139%)	7.7 (229%)
Melbourne	96.7 (103%)	88.6 (89%)	26.7 (102%)	21.4 (109%)

these synoptic events to generally last from 1-3 days.

4.6.2 Results

The top panel of Fig. 4.12 shows the STT ozone enhancements, based on a vertical integration of the ozone above baseline levels for each ozonesonde where an event was detected. The area considered to be ‘enhanced’ ozone is outlined with yellow dashes on the left panel of Fig. 4.4. We find that the mean ozone flux associated with STT events is $\sim 0.5\text{--}2.0 \times 10^{16} \text{ molecules cm}^{-2}$. The bottom panel shows the mean fraction of total tropospheric column ozone (calculated from ozonesonde profiles) attributed to stratospheric ozone intrusions at each site for days when an STT event occurred. First the tropospheric ozone column is calculated, then the enhanced ozone column amount is used to determine the relative increase. At all sites, the mean fraction of tropospheric ozone attributed to STT events is $\sim 1.0\text{--}3.5\%$. On three separate days over Macquarie and Melbourne, this value exceeds 10%.

The upper panels in figures 4.13-4.15 show the factors I , P , and Ω_{O_3} which are used along with the assumed event lifetime to estimate the STT flux. The tropospheric ozone and area of our region is calculated using the output and surface area from GEOS-Chem over our three regions. The lower panel of these figures show the results of the calculation when we choose two days for our flux estimation, with dotted lines showing the range of flux calculated if we assume events last from one day to one week. The seasonal cycle of ozone flux is mostly driven by the P term, which peaks in the SH summer over all three sites. Total uncertainty (shaded) is on the order of 100% (see Sect. 4.7.2). We calculate the annual amount based on the sum of the monthly values. The regions over Davis, Macquarie Island, and Melbourne have estimated STT ozone contributions of $\sim 5.7 \times 10^{17}$, $\sim 5.7 \times 10^{17}$, and $\sim 8.7 \times 10^{17}$ molecules $\text{cm}^{-2} \text{ a}^{-1}$ respectively, or equivalently $\sim 2.0, 2.1$, and 3.3 Tg a^{-1} .

4.6.3 Comparison to literature

Skerlak2014 show an estimate of roughly 40 to $150 \text{ kg km}^{-2} \text{ month}^{-1}$ in these regions, over all seasons (see Fig. 16, 17 in their publication) while we estimate from 0 to $180 \text{ kg km}^{-2} \text{ month}^{-1}$ STT impact, following a seasonal cycle with the maximum in austral summer. We estimate higher maximum flux over Melbourne, (178 , and $150 \text{ kg km}^{-2} \text{ month}^{-1}$ in January and February) than in either Davis ($89 \text{ kg km}^{-2} \text{ month}^{-1}$ in March) or Macquarie Island ($68 \text{ kg km}^{-2} \text{ month}^{-1}$ in January). Our calculated seasonal contributions, along with total uncertainty are shown in Table 4.3.

This result disagrees with several other studies which have found STT ozone fluxes in the SH extra-tropics are largest from autumn or winter to early spring. Roelofs1997

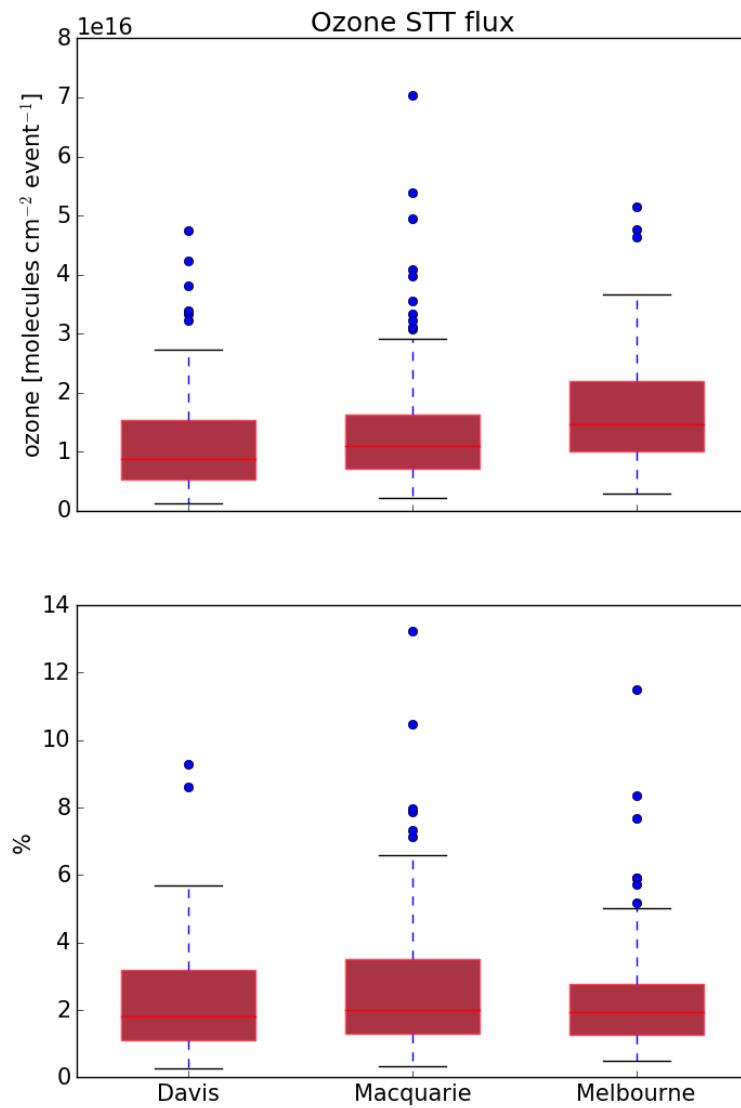


FIGURE 4.12: Top panel: tropospheric ozone attributed to STT events. Bottom panel: percent of total tropospheric column ozone attributed to STT events. Boxes show the inter-quartile range (IQR), with the centre line being the median, whiskers show the minimum and maximum, circles show values which lie more than 1.5 IQR from the median. Values calculated from ozonesonde measurements as described in the text.

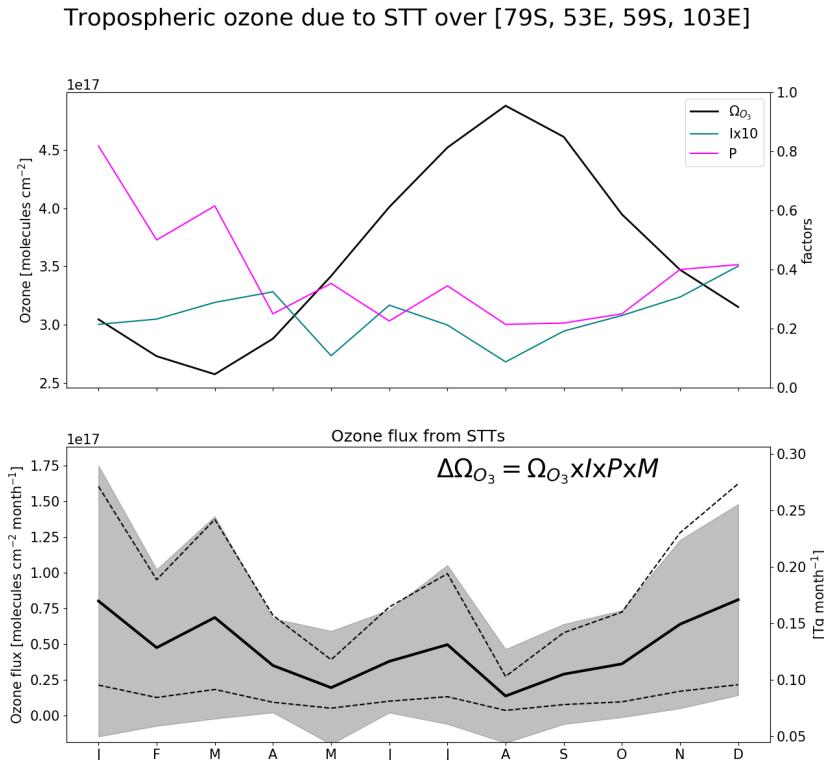


FIGURE 4.13: (Top) Tropospheric ozone, (I)impact per event, and (P)probability of event detection per sonde launch, averaged over the region above Davis. The tropospheric ozone column Ω_{O_3} (black, left axis) is from GEOS-Chem, while the STT probability P (magenta, right axis) and impact I (teal, right axis) are from the ozonesonde measurements. The STT impact is multiplied by ten to better show the seasonality. (Bottom) Estimated contribution of STT to tropospheric ozone columns over the region, with uncertainty (shaded area) estimated as outlined in Sect. 4.7. The black line shows STT ozone flux if event lifetime is assumed to be two days, with dashed lines showing the range of flux estimation if we assumed events lasted from one day to one week.

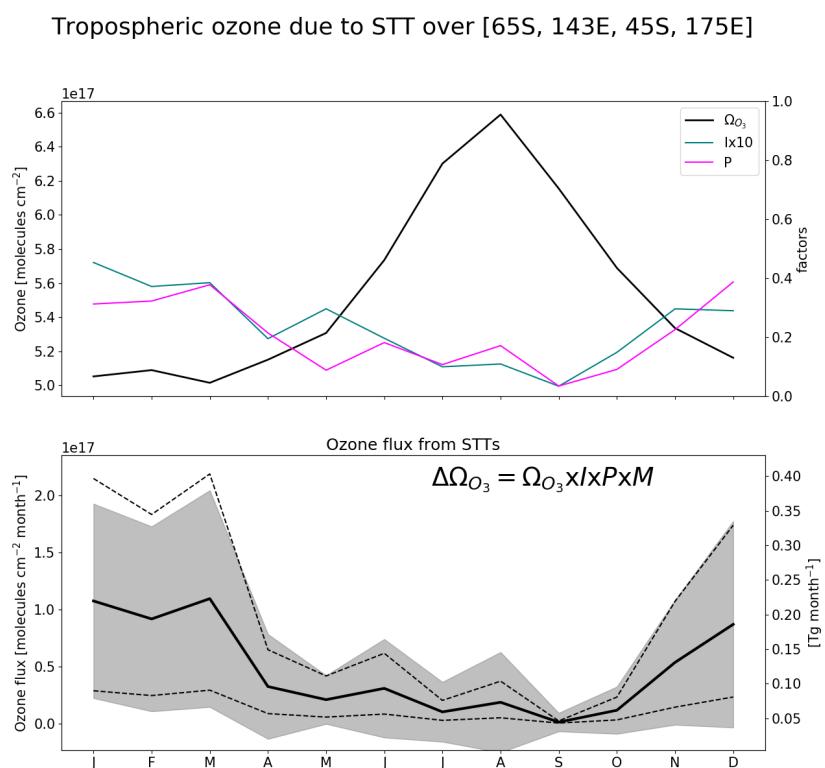


FIGURE 4.14: As described in 4.13, for the region containing Macquarie Island.

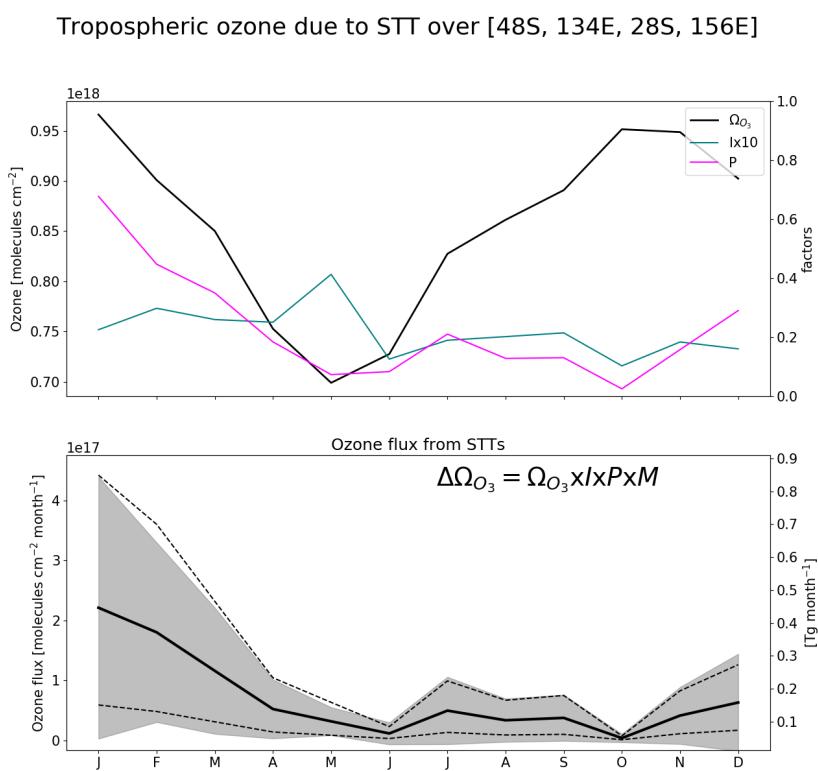


FIGURE 4.15: As described in 4.13, for the region containing Melbourne.

used a model carrying a tracer for stratospheric ozone to estimate STT impacts. They see higher SH tropospheric ozone concentrations, as well as STT flux, in the SH winter. Our model also shows ozone column amounts peaking in winter, however flux is maximised in summer due to our detected event frequencies. **Elbern1998** examine STT using ECMWF data for prior to 1996, using PV and Q-vectors to determine STT frequency and strength, and suggest fewer fold events in the SH occur from December to February. **Olsen2003** used PV and winds from the GEOS reanalysis combined with ozone measurements from the TOMS satellite to estimate that the ozone flux between 30° S and 60° S is 210 Tg yr^{-1} , with the maximum occurring over SH winter. **Liu2017** model the upper tropospheric ozone and its source (emissions/lightning/stratospheric) over the Atlantic ocean between 30° S and 45° S, and suggest that most of this is transported from the stratosphere from March to September, which is when the subtropical jet system is strongest.

The disagreements largely reflect the difference between point source based estimates and zonally averaged estimates, as the meteorological behaviour at our three sites is not the same as the system that dominates the southern hemisphere in general. As detailed in Sect. 4.4, the maximum STT influx which occurs during SH winter is almost entirely due to the dominant STT system which occurs annually over the southern Indian ocean and middle of Australia. It is difficult to compare remote ozonesonde datasets with area averaged models or re-analyses based on non-co-located measurements (such as ERA).

4.7 Sensitivities and limitations

4.7.1 Event detection

Our method uses several subjectively-defined quantities in the process of STT event detection. Here we briefly discuss these quantities and the sensitivity of the method to each. Using the algorithm discussed in Sect. 4.3.3, we detect 80 events at Davis, 105 (21 fire influenced) events at Macquarie Island, and 127 (27 fire influenced) events at Melbourne.

The cut-off threshold (defined separately for each site) is determined from the 95th percentile of the ozone perturbation profiles between 2 km above the earth's surface and 1 km below the tropopause. We use the 95th percentile because at this point the filter locates clear events with fewer than 5% obvious false positive detections. Event detection is sensitive to this choice; for example, using the 96th, and 97th percentile instead decreased detected events by 2, 9 (2,10%) at Davis, 13, 31 (11, 28%) at Macquarie Island, and 8, 24 (6, 18%) at Melbourne. Event detection is therefore also sensitive to the range over which the percentile is calculated. This range was chosen to remove anomalous edge effects of the Fourier bandpass filter and to discount the highly variable ozone concentration which occurs near the tropopause.

Ozone enhancements are only considered STT events if they occur from 4 km altitude up to 500 m below the tropopause. This range removes possible ground pollution and events not sufficiently separated from the stratosphere, while still capturing many well-defined events that occur within 1 km of the tropopause. An example of

a well-defined event that occurs within 1 km of the tropopause is shown in the supplementary (Fig. S2). However, STT events which reach below 4 km are physically possible and we may have some false negative detections due to the altitude restricted detections.

4.7.2 Flux calculations

Flux is calculated as $I \times P \times M \times \Omega_{O_3}$, with each term calculated as described in Sect. 4.6.1. The uncertainty is determined using the standard deviation of the product, with variance calculated using the variance of a product formula, assuming that each of our terms is independent:

$$\text{var}(\Pi_i X_i) = \Pi_i (\text{var}(X_i) + E(X_i)^2 - (\Pi_i E(X_i))^2)$$

The standard deviations for the I and Ω_{O_3} terms are calculated over the entire dataset. These terms are considered to be homoskedastic (unchanging variance over time). Uncertainty in assumed event lifetime is set at 50%, as we believe it is reasonable to expect events to last 1-3 days. P is the probability of any ozonesonde detecting an event, and is assumed to be constant (for any month). The overall uncertainty as a percentage is shown in parentheses in Table 4.3, these values are on the order of 100%, largely due to relative uncertainty in the I factor which ranges from 50-120% for each month.

Small changes in the region don't have a large affect on the per area flux calculations: increasing or decreasing the regions by 1° on each side ($\sim 10\%$ change in area) change the resulting flux by $\sim 1\%$. However due to the large portion of winter STT events being flagged due to potential smoke plume influence, a significant change in the yearly flux is seen when we don't remove these events. Without removing smoke flagged events we see an increase in estimated yearly flux of $\sim 1.1, 2.1 \times 10^{17}$ molecules $\text{cm}^{-2} \text{yr}^{-1}$ (which is a change of $\sim 15, 20\%$), over Macquarie Island and Melbourne respectively.

Considering the I factor, as discussed in here and in Sect. 4.7, there are several uncertainties in our method that are likely to lead to a low bias, such as the conservative estimate of flux within each event. Although there is little available data on SH ozone events for us to compare against, consider Terao2008, who estimated that up to 30–40% of the ozone at 500 hPa was transported from the stratosphere, in the northern hemisphere.

Our STT event impact estimates have some sensitivity to our biomass burning filter: including smoke-influenced days increases the mean per area flux by 15-20%. Although events which are detected near fire smoke plumes are removed, some portion of these could be actual STTs. The change in our P parameter when we include potentially smoke influenced events leads to a yearly estimated STT of 11×10^{17} molecules $\text{cm}^{-2} \text{yr}^{-1}$ over Melbourne, which suggests that up to 2.1×10^{17} molecules $\text{cm}^{-2} \text{yr}^{-1}$ ozone enhancement could be caused by smoke plume transported precursors. This is a potential area for improvement, as a better method of determining smoke influenced columns would improve confidence in our estimate.

Other possibly important uncertainties in our calculation of STT flux which we don't cover are listed here. Filtering events which occur within 500 m of the tropopause

may also lead to more false negatives. This could also cause lower impact estimates due to only measuring ozone enhancements which have descended and potentially slightly dissipated. On the other hand we have no measure of how often the detached ozone intrusion reascends into the stratosphere, which would lead to a reduced stratospheric impact. The estimated tropospheric ozone columns modelled by GEOS-Chem may be biased, for instance Hu2017 suggest that in general GEOS-Chem (with GEOS-5 met. fields) underestimates STT, with $\sim 360 \text{ Tg a}^{-1}$ simulated globally, compared to $\sim 550 \text{ Tg a}^{-1}$ observationally constrained. Transport uncertainty is very difficult to estimate with the disparate point measurements; it's possible that detected events are (at least partially) advected out of the analysis regions, which would mean we overestimate the influx into the region, and it is also possible that we are influenced by STT events outside the regions of analysis. Uncertainty in event longevity is set to 50%, however this implies a very simplistic model of event lifetimes. A great deal of work could be done to properly model the regional event lifetimes, however this is beyond the scope of our work.

Uncertainties in STT ozone flux detection are ($\sim 100\%$), and could be directly improved with larger or longer datasets. Possible parameterisations and an improved model of event lifetime could also improve the confidence in our estimate of event impacts, as well as allowing fewer assumptions.

4.8 Conclusions

Stratosphere-to-troposphere transport (STT) can be a major source of ozone to the remote free troposphere, but the occurrence and influence of STT events remains poorly quantified in the southern extra-tropics. Ozonesonde observations in the SH provide a satellite-independent quantification of the frequency of STT events, as well as an estimate of their impact and source. Using almost ten years of ozonesonde profiles over the southern high latitudes, we have quantified the frequency, seasonality, and altitude distributions of STT events in the SH extra-tropics. By combining this information with ozone column estimates from a global chemical transport model, we provided a first, conservative estimate of the influence of STT events on tropospheric ozone over the Southern Ocean.

Our method involved applying a bandpass filter to the measured ozone profiles to determine STT event occurrence and strength. The filter removed seasonal influences and allowed clear detection of ozone-enhanced tongues of air in the troposphere. By setting empirically-derived thresholds, this method clearly distinguished tropospheric ozone enhancements that are separated from the stratosphere. Our method is sensitive to various parameters involved in the calculation; however, for our sites we saw few false positive detections of STT events.

Detected STT events at three sites spanning the SH extra-tropics (38°S , 55°S , and 69°S) showed a distinct seasonal cycle. All three sites displayed a summer (DJF) maximum and an autumn to winter (AMJJA) minimum, although the seasonal amplitude was less apparent at the Antarctic site (Davis) as events were also detected regularly in winter and spring (likely due to polar jet stream-caused turbulence). Analysis of ERA-Interim reanalysis data suggested the majority of events were caused by turbulent weather in the upper troposphere due to low pressure fronts, followed by cut-off low

pressure systems. Comparison of ozonesonde-measured ozone profiles against those simulated by the GEOS-Chem global chemical transport model showed the model is able to reproduce seasonal features but does not have sufficient vertical resolution to distinguish STT events.

By combining the simulated tropospheric column ozone from GEOS-Chem with ozonesonde-derived STT estimates, we provide a first estimate of the total contribution of STT events to tropospheric ozone in these southern extra-tropical regions. We estimate that the ozone enhancement due to STT events near our three sites ranges from 300-570 kg km⁻² a⁻¹, with seasonal maximum in SH summer.

Estimating STT flux using ozonesonde data alone remains challenging; however, the very high vertical resolution provided by ozonesondes suggests they are capable of detecting STT events that models, re-analyses, and satellites may not. Further work is needed to more accurately translate these ozonesonde measurements into STT ozone fluxes, particularly in the SH where data are sparse and STT is likely to be a major contributor to upper tropospheric ozone in some regions. More frequent ozonesonde releases at SH sites could facilitate development of better STT flux estimates for this region.

4.9 Contributions and Acknowledgements

JWG wrote the algorithms, ran the GEOS-Chem simulations, performed the analysis and led the writing of the paper under the supervision and guidance of SPA, RS, and JAF. AK contributed the Davis ozonesonde data and performed the analysis of the alternate STT proxy. All authors contributed to editing and revising the manuscript.

Data availability. All GEOS-Chem model output and the ozonesonde observational data are available from the authors upon request.

We thank Dr. Sandy Burden for help clarifying some of the uncertainties involved in methods within this work. We also thank Dr. Clare Paton-Walsh, who identified the need to account for smoke-influenced events, and provided discussions on how to go about doing such. Ozonesonde data comes from the World Ozone and Ultraviolet Data Centre (WOUDC). The ERA-Interim data were downloaded from the ECMWF website following registration. This research was undertaken with the assistance of resources provided at the NCI National Facility systems at the Australian National University through the National Computational Merit Allocation Scheme supported by the Australian Government. This work was supported through funding by the Australian Government's Australian Antarctic science grant program (FoRCES 4012), the Australian Research Council's Centre of Excellence for Climate System Science (CE110001028), the Commonwealth Department of the Environment ozone summer scholar program. This research is supported by an Australian Government Research Training Program (RTP) Scholarship.

Chapter 5

Summary and Concluding remarks

In this thesis I aimed to improve understanding of natural contributions to ozone over Australia and the southern ocean. Chapter 1 Introduction and Literature Review examined processes leading to ozone creation in the troposphere, primarily isoprene emissions and subsequent chemistry. The secondary cause of ozone, stratosphere to troposphere transport (STT), is also outlined with associated processes and causes discussed. A summary of how the lack of information available over Australia affects modelling and forecasting is provided. Methodologies, tools and data-sets used throughout the thesis are detailed in Chapter 2 Data and Modelling . Chapter 3 Biogenic Isoprene emissions in Australia showed how models are misrepresenting isoprene emissions by a large margin in Australia, along with the description and implementation of a relatively simple method of improvement. Chapter 4 Stratospheric ozone intrusions provided an estimate of stratospheric ozone influx to the troposphere, along with potential classifications and seasonality.

5.1 Outcomes

I aim to recalculate satellite vertical columns of HCHO using updated model a priori information. HCHO from OMI on board the AURA satellite is examined and recalculated in Chapter 2, where influence of the air mass factor (AMF) is discussed in detail. This AMF was created using an older version of GEOS-Chem with out-of-date HCHO chemistry. Recalculation, binning, and analysis of the satellite HCHO vertical columns is performed using GEOS-Chem v10.01, outlined in Section 2.6.1. This is performed using my own partial AMF recalculation code, along with full recalculation code set up in collaboration with Prof. Palmer and Dr. Surl. Subsequent use of the data makes further use of the recalculation as a basic sensitivity analysis of top-down emissions estimations to a priori satellite information.

I aim to determine biogenic isoprene emissions in Australia using a top-down inversion of satellite HCHO, through an estimated yield from GEOS-Chem. In Chapter 3 I determine the linear relationship between total column HCHO and biogenic isoprene emissions over Australia using the global CTM GEOS-Chem. Applying this relationship to satellite based HCHO measurements creates the desired top-down isoprene emissions estimate. This process is described in Section 3.2, and requires intensive filtering of both satellite data (to exclude non-biogenic HCHO sources) and

model yield (to minimise spatial smearing). The uncertainty and limitations of top-down estimates due to satellite and model uncertainty, along with temporal and horizontal resolution of available data is examined in Section 3.4. Furthermore, this top-down estimation is used to scale a new simulation of HCHO, and O₃ over Australia, described in Section 3.2.8.

I aim to improve understanding of ozone transported to the troposphere from the stratosphere in Australia and the southern ocean. In Chapter 4 the seasonal cycle of STT events is characterised, and their contribution to the SH extra-tropical tropospheric ozone budget is quantified using GEOS-Chem to estimate ozone flux extrapolated from three measurement sites. Causal climatology and event seasonality are examined in Section 4.4. STT detection frequencies and modelled tropospheric ozone columns are used to estimate STT ozone flux near three sites in Section 4.6. Findings are compared against relevant literature, and the uncertainties involved in STT event detection and ozone flux estimation are studied.

In the following sections, summaries of results over Australia are often split into regions shown in Figure 5.1.

5.2 Isoprene emissions

Being a major driver of continental boundary ozone production, a new isoprene emissions estimate is created using OMI satellite measurements of HCHO. This estimate is compared against the bottom-up estimate from GEOS-Chem (running MEGAN). Generally months outside of May to August show the a posteriori lower than the a priori, except in the south eastern portion of Australia where only February is notably reduced. MEGAN isoprene emissions appear to be overestimated in Austral summer, suggesting poorly understood emission factors for Australian forests.

5.3 Ozone over Australia

Ozone production in the troposphere is a complex process involving various compounds. Tropospheric ozone is enhanced through VOC chemistry, stratospheric transport, and pollution. The first two of these processes are highly uncertain, with few studies performed on either topic within Australia. Of these ozone sources, VOC chemistry uncertainty is dominated by poor understanding of biogenic emissions (mainly) of isoprene. Emissions of isoprene are globally modelled at $\sim 465\text{--}500 \text{ Tg C yr}^{-1}$ (Guenther2006; Messina2016). However these appear to be overestimated in Australia, as seen here with top-down estimates at $\sim 19 \text{ Tg C yr}^{-1}$ compared to $\sim 43 \text{ Tg C yr}^{-1}$ using GEOS-Chem (which implements MEGAN).

TODO: ozone concentrations using emissions from MEGAN/TOP-Down, also how these compare to estimated STT: Tropospheric ozone production is estimated using GEOS-Chem before and after scaling isoprene emissions based on top-down satellite data. Figure 5.2 shows Australian summer surface level ozone with and without scaling isoprene emissions to match the multi-year top-down estimation. Tropospheric column ozone reduces very minimally (< 1%), however surface ozone drops $\sim 5\%$. This suggests that impacts from reduced isoprene emission are mostly important at the surface, and outweighed by other processes at higher altitudes. The



Figures/OMI_link/subregions.png

FIGURE 5.1: Sub-regions used in subsequent figures. Australia-wide averages will be black or grey, while results from within the coloured rectangles will match the colours shown here.

TABLE 5.1: Isoprene emissions from MEGAN and top-down estimation in Tg a^{-1} , along with ozone tropospheric column amounts in $\text{O}_3 \text{ cm}^{-2} \times 10^{17}$.

Metric	AUS	SEA	NEA	NA	SWA	MID
MEGAN	43(2)	blah				
Ozone	9.70	11.17	11.03	11.19	11.69	9.09
Top-Down	19(2)					
Ozone	9.64	11.11	10.99	11.12	11.63	9.02

Standard deviations shown in parenthesis.

results are tabulated for each of five regions, and all of Australia, and shown here in Table 5.1. Ozone production from isoprene emissions form an extra layer on top of anthropogenic ozone. A 5% mean difference in surface ozone levels will directly affect exposure levels and impacts air quality estimates.

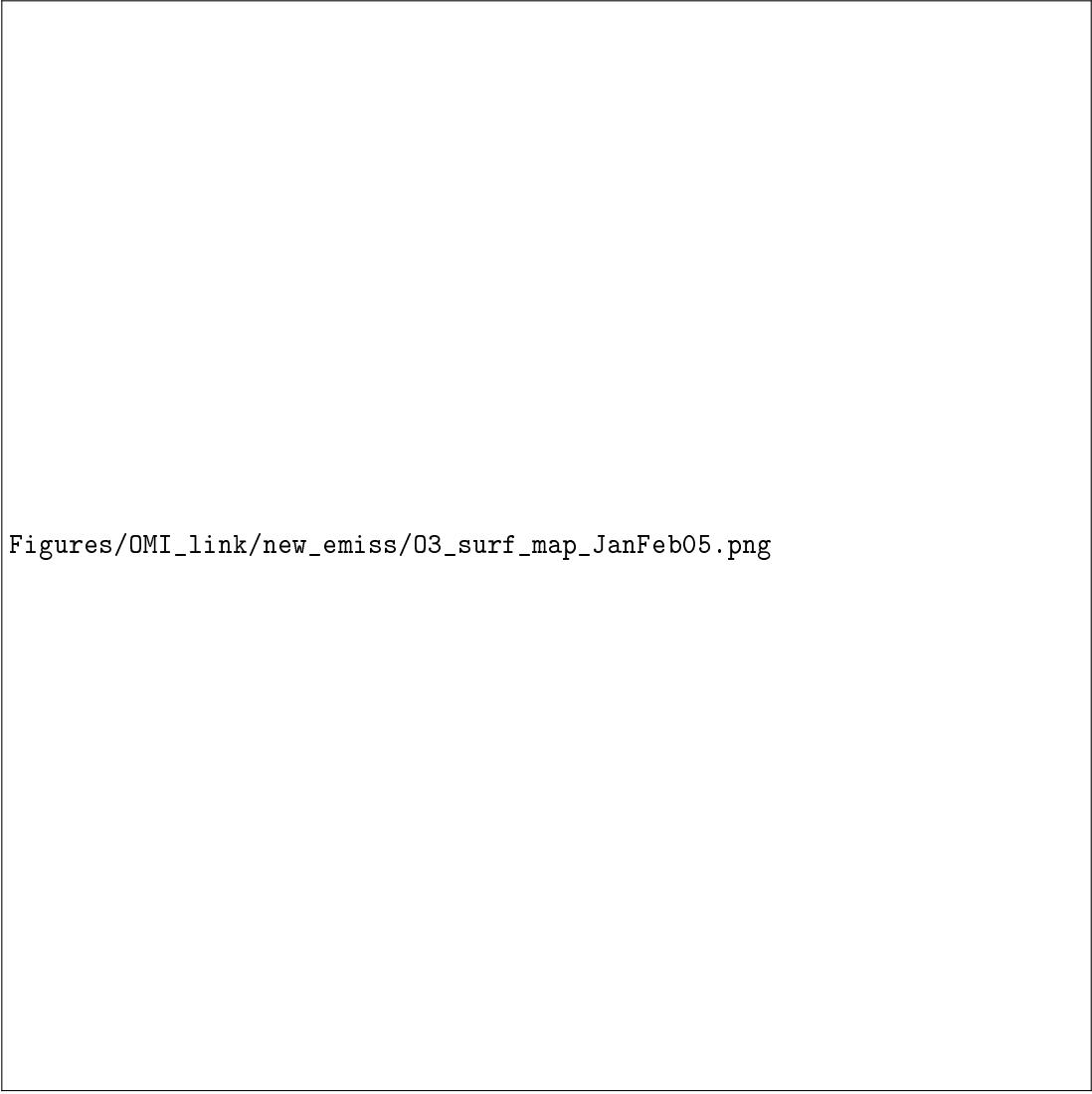
Ozone transported from the stratosphere also may have some affect on surface ozone levels, although this is hard to detect with the methods used in this thesis. We only compare the affect of the two disparate influences on the tropospheric ozone column. STT analysis over Melbourne suggests up to $\sim 10\%$ of the tropospheric column is due to stratospheric influx, with the average increase caused by influx being ~ 1 to 3.5% . We see that a relatively large drop in isoprene emissions causes $\sim 1\%$ reduced ozone concentration over summer. Both of these sources become more important in summer, with more frequent STT and stronger isoprene emissions. However the impacts are separated by altitude with most STT occurring in the free troposphere, while isoprene emission impacts are largely seen at the surface.

TODO: how isoprene is linked to ozone in modelling Figure 5.3 shows surface level (up to ~ 150 m altitude) ozone concentration over 2005 before and after scaling modelled isoprene emissions. Reducing isoprene emissions lowers surface ozone concentrations by TODO: XX to YY % in summer, and XX to YY % in winter. The direct correlation between reduced emissions and surface ozone (Figure TODO) is more or less clear monthly, and using multiple years enhances/reduces the relationship.

5.4 Outputs and future work

Analysis of STT along with an estimate of ozone flux from the stratosphere over a portion of Australia and the southern ocean is published in **Greenslade2017**. Ozone production using GEOS-Chem with updated top-down isoprene emissions is examined in this thesis and may form part of a new publication in the near future with the aim of updating MEGAN isoprene emission factors in Australia.

One of the important parameters in Australia is the soil moisture activity factor (γ_{SM}), which can have large regional effects on the isoprene emissions (**Sindelarova2014; Bauwens2016**). **Sindelarova2014** showed how Australian isoprene emissions could be as much as halved by accounting for lower soil moisture. Generally if soil moisture is too low, isoprene emissions stop (**Pegoraro2004; Niinemets2010**), however in many Australian regions the plants may be more adapted to lower moisture levels. (TODO:



Figures/OMI_link/new emiss/03_surf_map_JanFeb05.png

FIGURE 5.2: Ozone maps before (left) and after (right) scaling isoprene emissions in GEOS-Chem for summer. The bottom panel shows the linear regression between the runs along with a black dashed line representing the 1-1 ratio.



Figures/OMI_link/new emiss/03_surface_20050101_20051231.png

FIGURE 5.3: Surface ozone concentrations (ppb) per region over 2005.

Find cites for this - talk from K Emerson at Stanley indicated this) GEOS-Chem runs MEGANv2.1, which has three possible states for isoprene emissions based on the soil moisture (θ):

$$\begin{array}{ll} \gamma_{SM} = 1 & \theta > \theta_1 \\ \gamma_{SM} = (\theta - \theta_w) / \Delta\theta_1 & \theta_w < \theta < \theta_1 \\ \gamma_{SM} = 0 & \theta < \theta_w \end{array}$$

where θ_w is the wilting point, and θ_1 determines when plants are near the wilting point. The wilting point is set by a land based database from **Chen2001**, while θ_1 is set globally based on **Pegoraro2004**. These moisture states are disabled in GEOS-Chem V10.01. Improved isoprene emissions modelling requires this soil moisture problem to be handled. Simply enabling the parameter in its current form is not quite good enough for Australia, due to both the unknown soil moisture and the poorly understood plant responses in this country. TODO: cite the paper which could update MEGAN soil moisture parameterisation.

Using satellite data to improve isoprene emission estimates such as done in this thesis must be used to analyse model improvements, since fully independent measurements are lacking. Measurements of isoprene emissions in Australia would provide a valuable opportunity to verify improvements in emission estimates. Emissions measurement would be greatly valuable however they remain expensive and difficult, especially over the large sparse environment which makes up the Australia outback.