



## Thesis Title

Jesse Greenslade

Supervisor:  
Dr. Jenny A. Fisher  
Co-supervisor:  
Dr. Clare Paton Walsh

*This thesis is presented as part of the requirements for the conferral of the degree:*

Doctor of philosophy

The University of Wollongong  
School of School of Chemistry

January 1901

## **Declaration**

*I, Jesse Greenslade, declare that this thesis submitted in partial fulfilment of the requirements for the conferral of the degree Doctor of philosophy, from the University of Wollongong, is wholly my own work unless otherwise referenced or acknowledged. This document has not been submitted for qualifications at any other academic institution.*

---

**Jesse Greenslade**

*December 24, 2016*

# **Abstract**

The Thesis Abstract is written here (and usually kept to just this page). The page is kept centered vertically so can expand into the blank space above the title too...

# Acknowledgments

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 his noodly appendages.

# Contents

<b>Abstract</b>	iii
<b>1 Outline</b>	1
1.1 Overview . . . . .	1
<b>2 Literature review</b>	3
2.1 What are Volatile Organic Compounds (VOCs)? . . . . .	3
2.1.1 VOCs . . . . .	3
2.1.2 Hydroxyl (HO) and other radicals . . . . .	5
2.2 Natural gas and aerosol emissions in Australia . . . . .	6
2.2.1 Australia . . . . .	6
2.2.2 Satellite Measurements . . . . .	6
2.3 Ozone . . . . .	8
2.3.1 Basics . . . . .	8
2.3.2 Sources and sinks . . . . .	8
2.3.3 Measurements . . . . .	8
2.3.4 Estimates . . . . .	9
2.4 Formaldehyde(HCHO) . . . . .	9
2.4.1 Basics . . . . .	9
2.4.2 Sources and sinks . . . . .	9
2.4.3 Measurements . . . . .	10
2.4.4 Relationship with glyoxyl and isoprene . . . . .	11
2.4.5 Satellite measurements . . . . .	12
2.5 Isoprene . . . . .	13
2.5.1 Basics . . . . .	13
2.5.2 Sources and Sinks . . . . .	13
2.5.3 Factors affecting isoprene emissions . . . . .	14
2.5.4 Estimates . . . . .	15
2.5.5 Isoprene to HCHO . . . . .	15
2.5.6 Smearing . . . . .	19
2.5.7 Measurements . . . . .	19

2.5.8	Estimates . . . . .	20
2.5.9	Isoprene products . . . . .	20
2.5.10	Radiative Forcing . . . . .	21
2.6	Dust . . . . .	21
2.7	Models . . . . .	22
2.7.1	Chemical Transport Models . . . . .	22
2.7.2	GEOS-Chem . . . . .	23
2.8	Satellites . . . . .	24
2.8.1	Useful satellites . . . . .	24
2.8.2	Comparisons with Models . . . . .	25
2.8.3	DOAS . . . . .	26
<b>3</b>	<b>Stratosphere to Troposphere Transport of ozone</b>	<b>27</b>
3.1	Background . . . . .	27
3.1.1	Historical estimates . . . . .	27
3.1.2	Tropospheric production . . . . .	28
3.1.3	Stratosphere to Troposphere ozone Transport (STT) . . . . .	30
3.2	Instruments and data sets . . . . .	31
3.2.1	Atmospheric Infrared Sounder (AIRS) . . . . .	31
3.2.2	Sondes . . . . .	31
3.2.3	European Centre for Medium-Range Weather Forecasts (ECMWF) Re-Analysis - Interim data set (ERA-I) . . . . .	31
3.3	STT Detection . . . . .	32
3.3.1	Aim . . . . .	32
3.3.2	Tropopause Heights . . . . .	32
3.3.3	Fourier bandwidth (or bandpass) Filtering . . . . .	33
3.3.4	Bandwidth filter applied to ozonesondes . . . . .	35
3.3.5	Case Studies . . . . .	37
3.3.6	Site summaries . . . . .	37
3.4	Stratosphere to Troposphere flux analysis . . . . .	40
3.4.1	Determining a minimum estimate of stratospheric influence . .	40
3.5	Non-STT influences on ozone at southern latitudes . . . . .	41
3.5.1	Fire Plumes . . . . .	41
3.5.2	Transport Exclusion . . . . .	42
3.6	GEOS-Chem ozonesonde comparison . . . . .	44
3.6.1	Comparing data to GEOS-Chem . . . . .	44
3.6.2	Determining tropospheric ozone from GEOS-Chem . . . . .	45
3.6.3	Ozone profiles compared with GEOS-Chem . . . . .	46
3.6.4	Estimation of southern ocean STT flux . . . . .	58

<b>4 Formaldehyde product over Australia</b>	<b>60</b>
4.1 Australian Biogenic Volatile Organic Compounds (BVOCs) . . . . .	60
4.1.1 Isoprene, Monoterpenes . . . . .	60
4.1.2 Biomass Burning . . . . .	61
4.1.3 MEGAN . . . . .	61
4.2 Satellite HCHO measurements . . . . .	61
4.2.1 Satellite Retrievals . . . . .	61
4.2.2 OMI Algorithm BOAS . . . . .	63
4.2.3 Optical Depth ( $\tau$ ) . . . . .	65
4.2.4 Scattering . . . . .	66
4.2.5 Absorption cross section and number density . . . . .	66
4.2.6 Air Mass Factors . . . . .	66
4.2.7 OMI HCHO data products . . . . .	67
4.2.8 HCHO Vertical Column Calculation . . . . .	67
4.2.9 Uncertainty in OMI total columns . . . . .	70
4.2.10 Reference sector correction for comparison of products to various models . . . . .	71
4.3 Recalculating HCHO from satellite(OMI) data over Australia . . . . .	71
4.3.1 Process Outline . . . . .	71
4.3.2 Quality filtering OMI HCHO slant columns . . . . .	73
4.3.3 Reading OMHCHO daily slant columns . . . . .	75
4.3.4 Regridding to 0.25 by 0.3125 8-day averaged vertical columns	76
4.3.5 Filtering pyrogenic HCHO . . . . .	77
4.3.6 Filtering anthropogenic HCHO . . . . .	78
4.3.7 Recalculating the AMF to create our own vertical HCHO columns . . . . .	79
4.3.8 Determining and applying the pacific ocean reference sector normalisation . . . . .	82
4.3.9 Estimation of error or uncertainty . . . . .	85
4.3.10 Comparison with in-situ measurements . . . . .	86
<b>5 Isoprene Emissions in Australia</b>	<b>87</b>
5.1 GEOS-Chem isoprene mechanisms . . . . .	87
5.1.1 Outline . . . . .	87
5.1.2 Emissions from MEGAN . . . . .	88
5.2 Isoprene emissions estimation . . . . .	89
5.2.1 HCHO Products and yield . . . . .	89
5.2.2 Outline . . . . .	90
5.2.3 Calculation of Emissions . . . . .	92

5.2.4	Calculation of smearing effect . . . . .	93
5.2.5	Extrapolating the circadian cycle . . . . .	93
5.2.6	Comparison with MEGAN . . . . .	93
5.3	New estimates affects on the Australian atmosphere . . . . .	94
<b>A</b>	<b>Appendix A</b>	<b>95</b>
A.1	Grid Resolution . . . . .	95
	<b>Bibliography</b>	<b>95</b>

# Chapter 1

## Outline

### 1.1 Overview

In this thesis I will combine satellite and ground based atmospheric measurements with chemical transport modelling to clarify the impact of Australian natural emissions on atmospheric composition and chemistry. This broadly includes O<sub>3</sub>, NO<sub>X</sub>, VOCs, HO<sub>X</sub>, and the various processes which influence their composition and transport. Satellite measurements of HCHO, calculations of VOC emissions, HCHO and VOC atmospheric distributions, variations, and chemistry will be laid out with isoprene calculations being the main theme. The techniques used to analyse and create the products I use in my thesis will be explained. A background of HCHO modelling and MEGAN and GEOS-Chem will be split between here and the introduction to chapter 2.

A review of the literature on several topics is included with a view to touch on all the possible factors affecting my own work. Ozone and ozone precursors, along with the toxic effects and future projections of tropospheric ozone will be discussed. This leads into my **first chapter: “Stratosphere to Troposphere Transport of ozone”**. Here I analyse a dataset of ozonesondes, and calculate how many STTs occur in the high southern latitudes. An overview of my work categorising Stratospheric transport over the southern high latitudes will be the focus of this chapter as well as a paper I hope to submit next month.

An in depth review of how the GEOS-Chem model compares against ozonesondes will also be part of this chapter. GEOS-Chem is used to estimate how much tropospheric ozone is due to these STT events. These processes are some portion of the tropospheric ozone source - ties into another source: VOCs

My **second chapter: “HCHO total columns in Australia”** is focussed on calibrating a grided HCHO product from the OMI satellite measurements over Australia between January 2005 and April 2013. GOES-Chem and MEGAN model

histories will form some part of the background as they have been used several times to undertake similar inversions of VOC emissions. The process follows that of several other publications, however I focus on Australian emissions (largely unmentioned elsewhere) and parameters including precursor yields and grid resolution. Estimation of uncertainties and sensitivities to various factors including model grid resolution, fire, and anthropogenic masking operations is performed. Finally validation is undertaken using in-situ measurements of HCHO from an FTIR instrument on the roof of the chemistry building at the University of Wollongong.

My **third chapter “Isoprene emissions in Australia”** uses the HCHO product developed in chapter 2 along with various modelled parameters to estimate the emissions of isoprene (and possibly other BVOCs) through a simple linear steady state model, along with an analysis of the assumptions required for this model. Ideally the isoprene emissions estimates can be compared with MUMBA isoprene measurements during summer of 2012 - 2013. Again uncertainties and sensitivities will be examined for various parameters. Notably the effects due to smearing and grid resolution will be examined in detail, as these have never been heavily scrutinised for this inversion technique over Australia. A detailed comparison against other emissions estimates for Australia (ie: MEGAN and any other inventories) will be done with improvements or limitations highlighted. The isoprene emissions product will be analysed further by examination of its simulated VOC products.

# Chapter 2

## Literature review

### 2.1 What are Volatile Organic Compounds (VOCs)?

#### 2.1.1 VOCs

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 ( $\text{CO}_3$ ), and simple oxides of carbon and cyanides. 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, and have a facility to evaporate at low temperatures. Plants contain tens of thousands of organic compounds, it's likely that fewer than 40 are emitted due to the low volatility of most of them ([**Guenther2000**]).

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. VOCs have vapour pressure greater than  $10^{-5}$  atm, and are mostly generated naturally by plants, which emit around 1000 Tg per year ([**Guenther1995**, **Glasius2016**]). Due to their high volatility these compounds are generally seen 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 seen 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**]).

VOC emissions result in radical cycling, acid deposition, and the production of tropospheric ozone, and secondary aerosols ([**Atkinson2000**]). These have impacts on climate (through radiative forcing) and air quality, affecting both human health and crop yields ([**IPCC Chapter2**, **Avnery2011**, **Lelieveld2015**]). Understanding the drivers of trends in biogenic volatile organic compound emissions (BVOCs) is needed in order to estimate future carbon fluxes, changes in the water cycle, air

quality, and other climate responses ([**Yue2015**]).

In the 1990's, the World Meteorological Organisation (WMO) estimated that we are emitting  $360 \text{ Mt yr}^{-1}$  of methane ( $\text{CH}_4$ ), one of the more abundant and potent VOCs, while biogenic emissions were around  $200 \text{ Mt yr}^{-1}$  ([**Atkinson2000**]). At that point in time, emissions of other VOCs (Non Methane VOCs - NMVOCs) were estimated at  $1150 \text{ Mt yr}^{-1}$  (of carbon) from biogenic sources, and  $100 \text{ Mt yr}^{-1}$  from anthropogenic sources ([**Guenther1995**, **Atkinson2000**]). These estimates were based on the Model of Emissions of Gases and Aerosols from Nature (MEGAN, [**Guenther1995**]).

MEGAN initially included a simple canopy radiative transfer model, which parameterised sun-lit and shaded conditions through a canopy. Isoprene emissions were based on temperature, leaf area, and light, but have since been updated to include leaf age activity ([**Guenther1999**]), and a leaf energy balance model ([**Guenther2006**]) in MEGANv2.0. This update included a parameter for soil moisture, to account for drought conditions.

MEGAN has recently been analysed using 30 years of meteorological reanalysis information by [**Sindelarova2014**]. They estimate emissions of Biogenic VOCs (BVOCs) to be  $760 \text{ Tg(C) yr}^{-1}$ , 70% of which is isoprene. This is similar to estimates from other sources (TODO: cites and find those sources)

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  $\sim 566 \text{ TgC yr}^{-1}$ . This estimate is greater than initially thought and leads to decreased ( $\sim 10\%$ ) simulated OH concentrations to  $9.5 \times 10^5 \text{ molec cm}^{-3}$ .

Photolysis and oxidation of many VOCs initially form alkyl radicals ( $\dot{R}$ ), and reactions with ozone (with alkenes or VOCs containing a double bonded carbon) lead to organic peroxy radicals ( $\text{RO}_2$ ). These go on to form many products and lead to (amongst other things) aerosol, formaldehyde, and ozone formation, depending on various other factors such as sunlight and NO pollution ([**Atkinson2000**]).

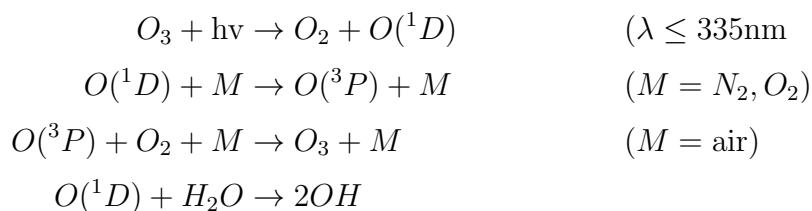
VOCs are removed by wet and dry deposition, or transformed by reaction with OH,  $\text{NO}_3$ , or  $\text{O}_3$ . The process of deposition only accounts for a small fraction of the VOC loss, with the possible exception of the long lived methane compound ([**AtkinsonArey2003**]). Primary reductions occur through photolysis, OH oxidation, ozonolysis, and at night time in polluted areas,  $\text{NO}_3$  ([**AtkinsonArey2003**, **Brown2009**]). In the presence of  $\text{NO}_x = \text{NO} + \text{NO}_2$ , non-methane organic compounds (NMOCs) and NMVOCs end up forming tropospheric ozone. This is achieved through photolysis of  $\text{NO}_2$ , concentrations of which are increased by NMOC and NO reactions ([**AtkinsonArey2003**]).

### 2.1.2 Hydroxyl (HO) and other radicals

TODO: hydroxyl introduction and importance

OH is a key species which reacts with nearly all the organic compounds in the troposphere. The exceptions are chlorofluorocarbons (CFCs), and Halons not containing H atoms ([Atkinson2000]). OH and HO<sub>2</sub> concentrations largely determine the oxidative capacity of the atmosphere. Oxidation and photolysis are the two main processes through which VOCs are broken down into HCHO, O<sub>3</sub>, CO<sub>2</sub> and various other species.

Ozone is an important precursor to HO, as excited oxygen atoms (O(<sup>1</sup>D)) are created through photolysis, which then go on to mix with water and form OH, as shown in this equation taken from [Atkinson2000]:

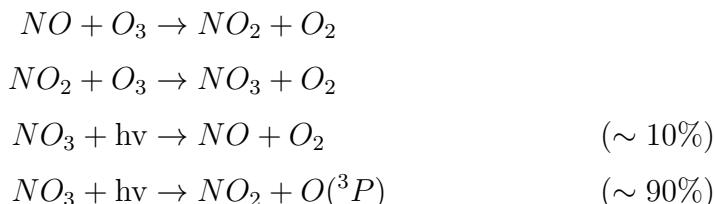


This shows how some of the O(<sup>1</sup>D) recycles back to Ozone, while some forms OH.

NB: The wavelength was updated to 350 nm in [AtkinsonArey2003].

In the late 90's it was understood that OH radicals are formed exclusively from photolysis of O<sub>3</sub>, HONO, HCHO, and other carbonyls [Atkinson2000]. TODO: Since then what do we know of OH formation?

Nitrate radicals NO<sub>3</sub> are also largely formed through ozone reactions. They are photolysed very rapidly during the day, with a lifetime of about 5 s ([Atkinson2000]). If NO and O<sub>3</sub> are both in the atmosphere, the following reactions ([Atkinson2000]) occur:



A build up of NO<sub>3</sub> radicals can be seen at night, when the quick photolysis is not occurring ([Atkinson2000, Brown2009]).

## 2.2 Natural gas and aerosol emissions in Australia

### 2.2.1 Australia

Australia is largely covered by environments which are not heavily influenced by human activity. These regions are natural sources of the trace gases which make up less than 1% of earth's atmosphere. Trace gases in the atmosphere can have a large impact on living conditions. They react in complex ways with other elements (anthropogenic and natural), affecting various ecosystems upon which life depends. Natural emissions affect surface pollution levels and can alter the radiative and particulate matter distribution of the atmosphere with harmful results. For example, ozone in the lower atmosphere is a serious hazard that causes health problems ([**Hsieh2013**]), damages agricultural crops worth billions of dollars ([**Avnery2011**]), and increases the rate of climate warming ([**IPCC 2013 chap8**]). Particulate matter in the atmosphere is also a major problem, causing an estimated 2-3 million deaths annually ([**Hoek2013**, **Krewski2009**, **Silva2013**, **Lelieveld2015**]).

Much of the landscape outside of urban areas is undeveloped and sparsely inhabited. In Australia most long term air quality measurements are performed in or near large cities. However, estimates of atmospheric gas and particulate densities, and their distributions over much of the continent are uncertain and lack in-situ measurements.

One important factor, affecting particulate matter and ozone concentration (both detrimental to human health), is a group called VOCs (Volatile Organic Compounds). The major source of VOCs in the atmosphere is biogenic, with around 90% of all global emissions coming from natural sources ([**Guenther1995**, **Guenther2006**, **Millet2006**]) Atmospheric VOCs can form harmful SOAs, and affect radical levels, which drive much of the chemistry in our atmosphere. Due to the lack of in-situ ground based measurements, estimates of VOC emissions are uncertain, with large scale extrapolation required [**Millet2006**]. VOC emission is based on many factors, including plant type and soil moisture ([**Guenther1995**]), both of which are not well characterised in Australia ([**Sindelarova2014**, **Bauwens2016**]). Changes in parameterisation of soil moisture in MEGAN lead to massive changes in Australian isoprene emission estimates, and soil moisture in Australia is not very well measured ([**Sindelarova2014**]). This has an compounding effect on the large uncertainties of biogenic VOC emissions ([**Guenther2000**, **Millet2006**]).

### 2.2.2 Satellite Measurements

Natural emissions from areas with little anthropogenic influence and no ground based measurements characterise the majority of Australian land mass ([**VanDerA2008**]).

One source of information which covers the entirety of Australia is remote sensing performed by instruments on satellites which overpass daily recording reflected solar (and emitted terrestrial) radiation. These can be used to quantify the abundance of several chemical species as well as estimate their distribution in vertical columns over the land.

The existence of satellite data covering remote areas provides an opportunity to develop more robust models of global climate and chemistry. Understanding of emissions from these areas is necessary to inform national policy on air pollution levels. Satellite data allow us to verify large scale model estimates of natural emissions. These measurements can be used to improve models, which are then able to predict harmful and costly events.

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 from satellite have large uncertainties, which arise due to several factors which arise in the process of transforming spectra to total column measurements, as well as instrument degradation (satellite instruments are hard to tinker with once they are launched). Uncertainty in satellite measurements comes from a range of things, including measurement difficulties introduced by clouds, and instrument sensitivity to particular aerosols ([Millet2006]).

There are two types of error, arguably the worst of these is systematic error (or bias) which 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 the other type (often reported as some function of a dataset's variance or else uncertainty), and this can be reduced through averaging either spatially or temporally. By taking the average of several measurements, any random error can be reduced by a factor of one over the square root of the number of measurements. This is done frequently for the relatively highly uncertain satellite measurements of 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. The main source of error in satellite retrievals of HCHO are due to instrument detection sensitivities, and the vertical multiplication factor (discussed in more detail in Section todo: point to an AMF section) ([Millet2006]).

## 2.3 Ozone

TODO: Start this section!

### 2.3.1 Basics

Ozone ( $O_3$ ) is mostly located in the stratosphere, where it helpfully prevents much of the shorter wave length solar radiation from reaching the earth's surface (ie UV light). However around 10% is in the troposphere (TODO: cite), where it has several deleterious effects. Ozone in the lower atmosphere is a serious hazard that causes health problems ([**Hsieh2013**]), damages agricultural crops worth billions of dollars ([**Avnery2011**]), and increases the rate of climate warming ([**IPCC·2013·chap8**]).

In the late 1990's, ozone transported down from the stratosphere was thought to contribute 10-40 ppb to the tropospheric ozone profile ([**Atkinson2000**]). Thought to match the atmospheric production of ozone (as in equation 2.1), this has since been revised downwards: from 50% to around 10 - 15% (TODO: cites with this number ([**Monks2015**])).

Ozone is a very important substance for formation of radicals in the troposphere ( $NO_3$ ,  $OH$ ), see Section 2.1.2 for more details.

### 2.3.2 Sources and sinks

Ozone is formed in the troposphere through oxidation of VOCs in the presence of  $NO_X$ . Net formation or loss of  $O_3$  is determined by interactions between VOCs,  $NO_X$ , and  $HO_X$ , and is a complicated system of positive and negative feedbacks ([**Atkinson2000**]).

Smoke plumes from biomass burning may carry precursors to ozone production. 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**]).

TODO: more on ozone.

### 2.3.3 Measurements

In the southern hemisphere there are relatively few records of ozone. Since 1986, Lauder, New Zealand ( $45^\circ S$ ,  $170^\circ E$ ) has released ozonesondes which measure ozone up to around 30 km ([**Brinksma2002**]). Kerguelan Island ( $49.2^\circ S$ ,  $70.1^\circ E$ ), 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/>.

TODO: Include ozone hole treaty and things put in place for that

### 2.3.4 Estimates

Recently global chemical transport models (CTMs) have been used to trace how much ozone is being transported to the troposphere from the stratosphere. There are a few methods of doing this, such as [Ojha2016], who use the ECHAM5 CTM with a tracer based on keeping track of ozone formed and transported from the stratosphere. The estimates generally require validation against actual measurements, such as those from ozonesondes or satellites.

## 2.4 Formaldehyde(HCHO)

### 2.4.1 Basics

HCHO, aka methanal, methyl aldehyde, and methylene oxide is of the aldehyde family. HCHO is an OVOC which is toxic, allergenic, and a potential carcinogen. It is dangerous at low levels, with WHO guidelines for prolonged exposure at 80ppb.

HCHO is used as an adhesive in plywood, carpeting, and in the creation of paints and wallpapers. Emissions in enclosed spaces can build up to dangerous levels, especially if new furnishings are installed ([Davenport2015]). One common way to detect and measure HCHO is through the DOAS technique, which takes advantage of the optically thin nature of HCHO in order to linearise the differential determined from the Beer-Lambert intensity law. This method works for both in the home HCHO detection and global measurements from in-situ and remote sensing instruments ([Guenther1995, Abad2015, Davenport2015]).

### 2.4.2 Sources and sinks

In the atmosphere HCHO is primarily produced through the oxidation of methane ( $\text{CH}_4$ ) by the hydroxyl radical ( $\text{OH}$ ). Within the continental boundary layer, the major source of HCHO enhancement is VOC emissions reacting with OH radicals in the presence of  $\text{NO}_X$  ([Wagner2002, Millet2006]). There is a complex relationship between VOCs,  $\text{HO}_X$ , and  $\text{NO}_X$ , and with higher levels of  $\text{NO}_X$  the speed that VOCs are converted into HCHO increases, as does the HCHO concentration ([Wolfe2016]). Isoprene is the main VOC precursor of HCHO in the continental boundary layer, except near fires or anthropogenic sources of HCHO and precursors ([Guenther1995, Wolfe2016]).

Biomass burning can be a source of HCHO, and various other pollutants, precursors, and aerosols. Additionally HCHO is emitted into the atmosphere directly through fossil fuel combustion, natural gas flaring, ethanol refining, and agricultural activity ([**Wolfe2016**]). Background levels of HCHO are provided by methane oxidation, and enhancements to regional and continental HCHO is largely driven by isoprene emissions ([**Guenther1995**, **Palmer2003**, **Shim2005**]). [**Atkinson2000**] summarised the background formation of HCHO with the following reaction:



which shows that methane photolyses and oxidises to form HCHO and ozone in a process that regenerates the OH radicals.

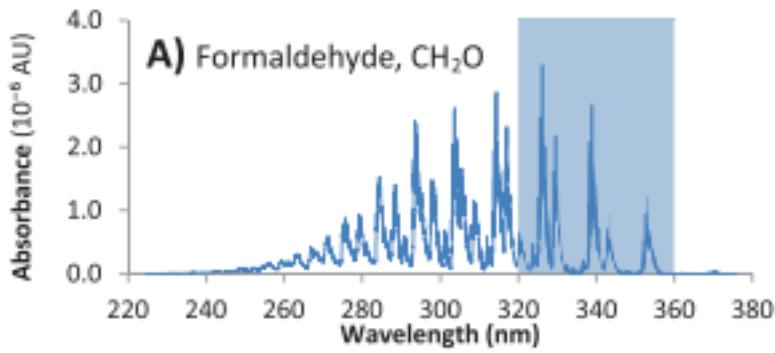
HCHO has two major sinks, one being reactions with OH, the other being photolysis ([**Crutzen1999**, **Wagner2002**, **Levy1972**]). (todo: levy 1972 read abstract and cite). These reactions lead to a daytime lifetime of a few hours ([**Atkinson2000**, **Millet2006**]). Both these loss processes form CO and hydroperoxy radicals ( $HO_2$ ), and have global significance to radiative forcing and oxidative capacity ([**Franco2015**]). The other notable sinks are wet and dry deposition, although these are not as significant ([**Atkinson2000**]) (todo add more cites here).

In the past, HCHO levels were underestimated by models, often with large discrepancies, due to the poor understanding of methyl peroxy radical ( $CH_3OO$ ) chemistry ([**Wagner2002**]).

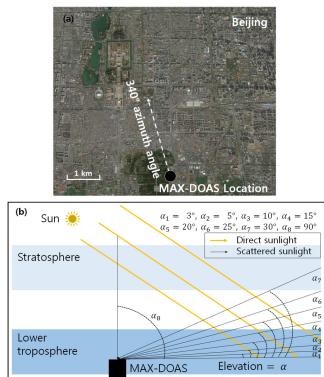
### 2.4.3 Measurements

There are a few ways to measure HCHO, including Fourier Transform Infra-Red Spectrometry (FTIR). As a trace gas HCHO interferes with light over a few wavelength bands, which allows instruments to detect concentrations along a path between a sensor and a known light source like a lamp or the sun. Figure 2.1 shows the interference spectrum of HCHO as well as 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 measurements can have a range of uncertainties, including systematic and random measurement errors and uncertainties in apriori shape factors and water profiles (eg: [**Franco2015**]). Multiple axis differential optical absorption spectroscopy (MAX-DOAS) also examines the infra-red light interference. In [**Franco2015**], an FTIR spectrometer at Jungfraujoch is compared against both MAX-DOAS and satellite data, with two CTMs; GEOS-Chem and IMAGES v2 used to compare total columns and vertical resolution of each instrument.

In MAX-DOAS retrievals, the measurements of light absorption are performed



**Figure 2.1:** HCHO spectrum, with a typical band of wavelengths used for DOAS path measurements. This is a portion of an image from [Davenport2015].



**Figure 2.2:** Image from [Lee2015].

over several elevations in order to add some vertical resolution to the measurement of trace gas concentrations. An example of this is shown in figure 2.2, which was taken from [Lee2015]. Recently MAX-DOAS has been used to examine HCHO profiles in the clean free troposphere ([Franco2015, Schreier2016]) as well as in polluted city air ([Lee2015]). Depending on orography and atmospheric composition (ie. the influence of interfering chemicals), MAX-DOAS can be used to split the tropospheric column into two partial columns; giving a small amount of vertical resolution to HCHO measurements ([Franco2015, Lee2015]).

Other measurement techniques include chromatographic and fluorimetric methods, both of which differ widely from each other and the spectroscopic method TODO: read ([Hak2005])). This resulted in HCHO not having a consistent network for global measurements like those for GHGs or Ozone ([FortemsCheiney2012]).

#### 2.4.4 Relationship with glyoxyl and isoprene

Glyoxyl (CHOCHO) is important to us as it shares many properties with HCHO, and may provide additional information in determining isoprene emissions. Glyoxyl is another product of VOC oxidation in the atmosphere, with isoprene being the main source globally. Isoprene has been used to estimate isoprene emissions (see

section 2.5.5) but many uncertainties exist. One of these uncertainties is the yield of HCHO from isoprene, especially in low  $\text{NO}_X$  environments. Glyoxyl could prove complementary to HCHO in constraining isoprene emissions (TODO: Read and cite Vrekoussis2009,2010, Chan Miller 2014, Alvarado 2014).

Under high  $\text{NO}_X$  conditions, glyoxyl forms rapidly, similarly to HCHO. However, glyoxyl also forms in low  $\text{NO}_X$  environments both slowly (through isoprene epoxydiols), and rapidly (through di-hydroperoxide dicarbonyl compound photolysisation (todo: read abstract and cite crounse(Autoxidation of Organic Compounds in the Atmosphere) 2013). This process is similar to the proposed mechanisms for hydroperoxyaldehydes by [Peeters2014] and carbonyl nitrates (todo: read and cite M uller 2014).

#### 2.4.5 Satellite measurements

Satellite measurements of HCHO are relatively uncertain, however this can be improved by averaging over larger grid boxes or longer time scales. an example of this can be seen in (TODO: read/cite dufour2009), where monthly averaging is used to decrease the measurements uncertainty. HCHO in Europe is low; near the detection limit of satellite measurements, however taking monthly averages allows enough certainty that useful inversions can be determined to estimate the sources of the HCHO (TODO: read and citepDufour2009).

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 on board the ERS-2) corrects for an instrument artifact using modelled HCHO over the remote pacific ([Shim2015]). OMI HCHO products use a similar technique to account for sensor plate drift and changing bromine sensitivity ([Abad2015])

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 (SEAC<sup>4</sup>RS). [DeSmedt2015] also found a low bias from 20 - 40% when comparing OMI and GOME2 observations against 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 apriori HCHO profiles which are used to calculate the AMFs of the satellite columns.

The OMI measurements used in this research are recalculated using an updated estimate of HCHO profiles and validated against Wollongong total column measurements. Similar to validation performed by other studies TODO: cite and exemplify

a few studies ie: [Zhu2016, Marais2015, Palmer2003]

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 (TODO: use both these methods for HCHO section)([DeSmedt2012, DeSmedt2015]).

GOME suffers from similar uncertainties to OMI, as the same general method of DOAS remote measurements are performed. The uncertainty from slant column fitting has been calculated for GOME to be  $4 \times 10^{15}$  molecules cm $^{-2}$  ([Chance2000, Millet2006]). The conversion factor for slant to vertical columns (AMF) calculation also suffers from errors; primarily from surface albedo, HCHO vertical profile apriori, aerosol, and cloud influence ([Millet2006]). AMF uncertainties for GOME are calculated to be 1 to  $1.3 \times 10^{15}$  molecules cm $^{-2}$  by [Shim2005].

## 2.5 Isoprene

### 2.5.1 Basics

Isoprene, or 2-methylbuta-1,3-diene, is a Volatile Organic Compound (VOC) and has chemical formula of C<sub>5</sub>H<sub>8</sub>. Isoprene effects NO<sub>X</sub> and HO<sub>Y</sub> cycling, and in the presence of NO<sub>X</sub>, forms tropospheric ozone and SOAs ([Wagner2002, Millet2006]). Bottom up inventories of VOCs remain largely uncertain due to extensive extrapolation over plant functional types, changing land cover, and parameterised environmental stressors (todo: read and ([Guenther2000])).

### 2.5.2 Sources and Sinks

Methane and isoprene each comprise around a third of the yearly global total emission of VOCs. However, methane is relatively long lived (years) and is well mixed in the atmosphere while isoprene levels are very volatile and spatially diverse due to a life time of around an hour. Estimates put global isoprene emission at roughly 600 Tg yr $^{-1}$ , emitted mostly during the day. Major emitters are tropical broadleafs (notably eucalypts), and scrubs ([Guenther2006, Arneth2008, Niinemets2010, Monks2015]).

Isoprene has a short lifetime during the day, roughly an hour due to OH oxidation (todo: abstract and cite Atkinson and Arey 2003([AtkinsonArey2003])). At night when OH concentrations have dropped, isoprene can remain in the atmosphere to be transported. Typically less than half of this night time isoprene is removed

through ozonolysis ([**AtkinsonArey2003**]), however, in polluted areas where high levels of  $\text{NO}_X$  exist, isoprene is consumed by a different radical. During the night time, nitrate radicals ( $\text{NO}_3$ ) build up, especially in areas with high  $\text{NO}_X$  levels. In areas with high  $\text{NO}_X$  levels, greater than 20% of the isoprene emitted late in the day ends up being oxidised by the  $\text{NO}_3$  radical over night ([**Brown2009**]). So while night time isoprene is not as highly concentrated, it does have varying biogenic and anthropogenic sinks. At night isoprene has affects on both  $\text{NO}_X$  concentrations and ozone levels, and can form harmful SOAs ([**Brown2009, Mao2013**]).

Isoprene oxidation by OH is less well understood in lower concentrations of NO. Initially isoprene was thought to be a sink for atmospheric oxidants ([[]]eg.)Guenther2000. It was thought that in low NO environments, like those far from anthropogenic pollution and fires, oxidation of isoprene would create hydroxyhydroperoxides (ISOPOOH) and lead to low concentrations of OH and  $\text{HO}_2$  (together known as  $\text{HO}_X$ ) [**Paulot2009b**]. In [**Paulot2009b**], the  $\text{HO}_X$  levels are shown to be largely unaffected by isoprene concentrations. They show that ISOPOOH is formed in yields  $> 70\%$ , and MACR and MVK is formed with yields  $< 30\%$ . The formation of MACR and MVK produces some  $\text{HO}_X$ , although not enough to close the gap. [**Paulot2009b**] goes on to suggest (and provide experimental evidence) that dihydroxyperoxides (IEPOX) are formed from oxidation of the ISOPOOH, which form precursors for SOAs as well as closing the  $\text{HO}_X$  concentration gap. They then use GEOS-Chem, modified to include IEPOX formation, to estimate that one third of isoprene peroxy radicals react with  $\text{HO}_2$ , and two thirds react with NO. Their work showed another pathway for isoprene based SOA creation, and additionally estimated  $95 \pm 45 \text{ TgC yr}^{-1}$  IEPOX being created in the atmosphere without any inclusion in CTMs at that time.

Although understanding of OH production/recycling in these low NO conditions has been improved, many observations of OH are still quite under-predicted in models ([**Mao2012**]). It was shown in [**Mao2012**], for a remote forest in California, that the traditional method of OH measurement may be affected by instrument internally generated OH from VOC oxidation. This lends more credence to the current understanding of VOC oxidation as it closed the gap between measurements and model predictions.

### 2.5.3 Factors affecting isoprene emissions

[**Marais2014**] examine factors affecting isoprene emissions, showing how emissions are sensitive to various environmental factors. Their work used MEGAN ([**Guenther1995**]) and GEOS-Chem to look at how these factors affect surface ozone and particulate matter in Africa. One of the important uncertainties seen in MEGAN within this work is the isoprene emissions due to plant type. Canopy

level isoprene measurements are made using relaxed eddy accumulation (REA) at several sites in Africa. One plant type near a measurement site emits more than other species and it's actual distribution on a larger scale is completely unknown - leading to possible overestimations in MEGAN.

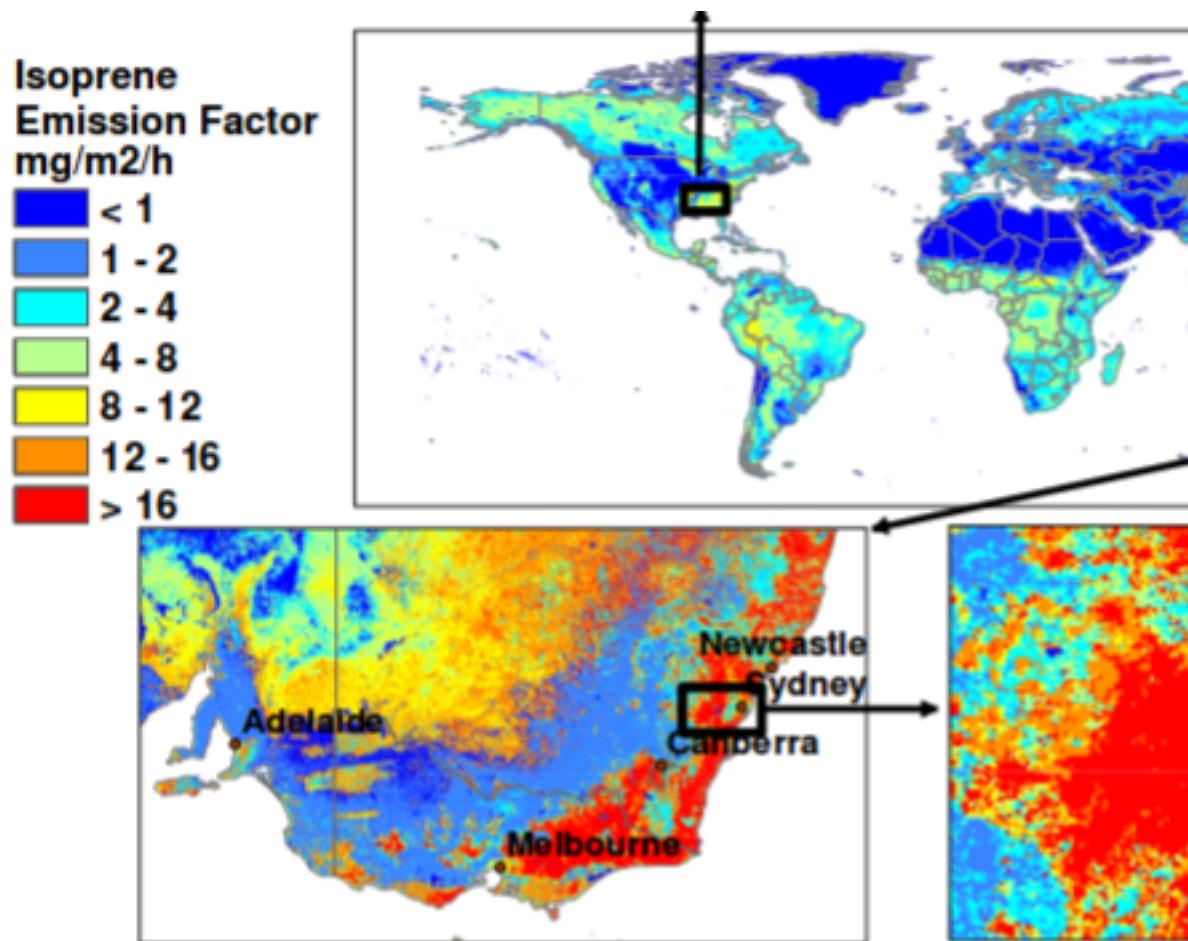
[**Stavrakou2014**] examined modelled Asian emissions and altered model parameters for temperature, plant type emission factors, incoming solar radiation (insolation) intensity, land use changes, and palm tree forest expansion. Changes were constrained by a network of radiation measurements and some experiments with south east Asian forest emissions - and led to reduction in isoprene emissions by a factor of two over the region. The Asian region is also shown to have a strong correlation with the Oceanic Nio Index (ONI), with positive anomalies associated with El Nio.

#### 2.5.4 Estimates

Biogenic isoprene emissions are far greater than anthropogenic emissions of VOCs ([**Guenther2006**, **Kefauver2014**]). The lack of accuracy in BVOC emissions estimates has a large effect on determining with confidence the sources and distribution of pollutants including ozone and organic aerosols. Accuracy in VOC measurements is important: it has been shown that even the diurnal pattern of isoprene emissions has an effect on modelling ground level ozone ([**Hewitt2011**, **Fan2004**]). These uncertainties could explain why models of HCHO over Australia are poor at reproducing satellite measurements ([**Stavrakou2009**]).

[**Guenther2006**] estimates 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 2.3). These emissions factor estimates are not well verified as there is little coverage of isoprene (or other BVOC) emissions measurements over Australia. However, comprehensive coverage of one high yield (generally) product in the atmosphere over Australia exists in the form of satellite measurements.

It is important to note that many estimates of isoprene emission are based on a few algorithms of isoprene emission which can depend greatly on input parameters ([**Niinemets2010**]). [**Yue2015**] has 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 based, while the other (MEGAN) uses the light and canopy temperature (Niinemets et al., 1999; Arneth et al., 2007; Unger et al., 2013).



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

**Figure 2.3:** Part of a figure from [Guenther2006] showing global isoprene emission factors.

### 2.5.5 Isoprene to HCHO

In the remote troposphere HCHO production is dominated by methane oxidation, while in the continental boundary layer (CBL) production is largely due to NMVOCs ([**Abbot2003**, **Kefauver2014**]). NMVOCs are alkanes, alkenes, aromatic hydrocarbons and isoprene. Isoprene is hard to measure directly due to its short lifetime and weak spectral absorption, instead formaldehyde is often used as a proxy ([**Millet2006**, **Dufour2009**, **Marais2012**, **bauwens2013satellite**, **Kefauver2014**, **Bauwens2016**]). Formaldehyde formed in the troposphere is mostly due to VOC (roughly one third each: methane, isoprene, others) oxidation. We can model this oxidation process in order to work out how much VOC is present based on the total HCHO. This requires among other things an idea of which VOCs are present and their yields of HCHO.

The method used to develop top-down isoprene inference using satellites was developed initially by [**Palmer2001**, **Palmer2003**]. Isoprene emissions fluxes were derived using the Global Ozone Monitoring Experiment (GOME) satellite instrument. Palmer's method improved biogenic isoprene emissions estimates (compared with in-situ measurements) over two available inventories: the U.S. EPA Biogenic Emissions Inventory System (BEIS2) and the Global Emissions Inventory Activity (GEIA). This showed an inversion technique which could be used to improve large scale emissions estimates without further expensive measurement campaigns.

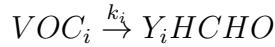
[**Dufour2009**] use HCHO from SCIAMACHY, and examine Europe using CHIMERE as the chemical model.

Satellites recording reflected solar spectra use Differential Optical Absorption Spectroscopy (DOAS) to measure various trace gases in the atmosphere, including formaldehyde. Formaldehyde levels in the continental boundary layer are generally dominated by chemical formation due to VOC (largely isoprene) emissions ([**Kefauver2014**]).

Satellites can be used to measure the seasonal and interannual variability of HCHO over Australia. These records can be compared with modeled estimates of HCHO and used as a proxy to estimate isoprene emissions. This has been done in North America ([**Palmer2003**, **Millet2006**]), South America, Africa, China, Europe ([**Dufour2009**]), and recently globally ([**FortemsCheiney2012**, **Bauwens2016**]). Often these works use two forms of measurement such as satellite and aircraft data combined for validation ([**Marais2014**]).

Initially studies assumed a simple linear steady-state relationship between HCHO and it's precursors ([**Palmer2003**, **Palmer2006**, **Millet2006**]). This allowed a simple calculation of isoprene using the measured HCHO, with estimated reaction rates and yields. The methodology for calculating VOCs from HCHO is laid out in

[Palmer2003], and takes into account the expected lifetime and reaction rates of the precursor VOCs and HCHO. Assuming HCHO is produced quickly from short-lived intermediates, and the column is at steady state:



Where  $Y_i$  is HCHO yield per C atom (a measure of how much HCHO will form per gram of C from a VOC within a system), and  $k_i$  is the reaction rate. Then assuming a steady state of atmospheric HCHO ( $\Omega$  molecules  $cm^{-2}$ ) produced by oxidation of VOCs ( $VOC_i$ ) and no horizontal transport:

$$\Omega = \frac{1}{k_{HCHO}} \sum_i Y_i E_i$$

Where i indexes a chemical species,  $k_{HCHO}$  is the HCHO loss rate due to OH and photolysis,  $Y_i$  is the molar HCHO yield from oxidation of i, and  $E_i$  is emission fluxes (C atoms  $cm^{-2}s^{-1}$ ).

Estimates of  $Y_i$  can be attained from a model as shown in [Millet2006]. This involves a reduced major axis (RMA) correlation calculation between modelled HCHO and isoprene columns, multiplied by their loss rates (to photolysis and oxidation) (as a normalising factor). In high NOx environments where HCHO has a lifetime on the order of 30 minutes, it can be used to map isoprene emissions with spatial resolution from 10-100 kms. Horizontal transport ‘smears’ the HCHO signal so that source location would need to be calculated using windspeeds and loss rates ([Palmer2001, Palmer2003]). For more details on this see section 2.5.6.

Another method of correcting isoprene emissions using observed HCHO total column involves a Bayesian inversion. [Shim2005] work with GOME HCHO observations and GEOS-Chem, looking at areas with high signal to noise ratio (higher HCHO concentrations). They show that the model underestimates isoprene emissions and HCHO concentrations by 14-46%, with the corrected VOC emissions reducing the model biases to 3-25%.

The Bayesian inversion is also used in [Curci2010], where a regional CTM (CHIMERE) simulates HCHO, which is compared against OMI observed HCHO and shown to be regionally biased. This bias is expected to be caused by errors in MEGAN’s natural isoprene emissions. The CHIMERE model is used to derive yields of HCHO from the various local VOCs and these are then used in estimating local emissions. The model is run initially with emissions of BVOCs and reactive anthropogenic VOCs (RAVOCs) turned off in order to work out the background (b) values of these compounds. The Bayesian inversion is used to correct regionally biased biogenic isoprene emissions by optimising these parameters in order to simu-

late HCHO closest to the observed HCHO levels. [**Curci2010**] uses CHIMERE as the forward model to determine the relationship between HCHO ( $y$ ), isoprene and reactive anthropogenic VOCs ( $\mathbf{x}$ ), using

$$y = \mathbf{K}\mathbf{x} + b + \epsilon \quad (2.2)$$

where  $\epsilon$  are the (assumed) independent errors in measurements.  $\mathbf{K}$  is the Jacobian matrix determined from CHIMERE representing the sensitivity of  $y$  to the state variable  $\mathbf{x}$ . This  $\mathbf{K}$  matrix is used in conjunction with error covariance in  $\mathbf{x}$  to determine the Maximum A Posteriori (MAP) solution to calculate the optimal estimate of  $\mathbf{x}$  ( $\hat{\mathbf{x}}$ ).

TODO: Read through this list of sources on the hcho to isop process : taken from Wolfe2015 Such techniques have informed isoprene emission inventories in North America (Abbot et al., 2003; Millet et al., 2008 ([**Palmer2003**, **Millet2006**, **Palmer2006**]), South America ([**Barkley2013**]), 2008), Europe ([**Curci2010**, **Dufour2009**]), Africa ([**Marais2012**]), Asia (Fu et al., 2007; Stavrakou et al., 2014), and globally (Fortems-Cheiney et al., 2012; ([**Shim2005**]); Stavrakou et al., 2009).

More recently, full inversions that better account for transport, source attribution, and chemical schemes have been implemented ([**FortemsCheiney2012**]). TODO: full description of this better inversion technique going through Fortems-Cheiney2012.

### 2.5.6 Smearing

The distance travelled downwind ( $L_{d,i}$ ) by a precursor (i) before becoming HCHO can be estimated using:

$$L_{d,i} = \frac{U}{k_i - k_{HCHO}} \ln \left( \frac{k_i}{k_{HCHO}} \right)$$

where  $U$  is windspeed. [**Palmer2003**] further define a smearing length scale:  $L_{s,i}$  as the distance downwind where a fraction ( $1 - 1/e$ ) of the precursor is completely transformed into HCHO. This equation uses the initial VOC column concentration ( $[VOC]_0$ ) at the point of emission and mass balance equations, and is as follows:

$$\frac{1}{k_{HCHO} - k_i} \left( k_{HCHO} \exp \left[ \frac{-k_i L_{s,i}}{U} \right] - k_i \exp \left[ \frac{-k_{HCHO} L_{s,i}}{U} \right] \right) = \frac{1}{e} \quad (2.3)$$

with limiting values  $L_{s,i} \rightarrow U/k_i$  for  $k_i \ll k_{HCHO}$ , and  $L_{s,i} \rightarrow U/k_{HCHO}$  for  $k_{HCHO} \ll k_i$ .

Accounting for transport of the precursors is important, especially in low NO<sub>X</sub>

conditions in which isoprene has a longer lifetime (days). This allows horizontal transport to occur and complicates the algorithms, as can be seen by the smearing length scale which increases beyond the 100 km. For conditions where VOCs have a lifetime of days determining the major HCHO contributors requires a complex inversion to map HCHO columns to VOC emissions.

### 2.5.7 Measurements

There are relatively few measurements of isoprene in the southern hemisphere, including MUMBA(TODO CITE), other campaigns?, and very recently that girl from Macquarie University with an instrument in the daintree rainforest(TODO CITE, DESCRIBE?). Since 1997, when GOME first measured HCHO over Asia (TODO cite thomas 1998), satellites have been able to provide a total column measurement of one of the primary products of isoprene.

### 2.5.8 Estimates

There are two commonly used ways of estimating isoprene emissions, top-down or bottom-up. Bottom-up emission estimates generally model the flora and events which emit isoprene, like Eucalypts, factories, shrubs, leaf areas under sunlight, etc. Understanding how much isoprene is emitted, when and by what is more complicated than it sounds, and since little data exists with which to verify these bottom-up emission inventories they are uncertain on a large scale. Top-down estimates look at how much of a chemical is in the atmosphere and try to work out how much of it's major precursors were emitted. For isoprene this is done by looking at atmospheric HCHO enhancement, which can be largely attributed to isoprene emissions as long NO<sub>X</sub> and transport effects are accounted for.

### 2.5.9 Isoprene products

Isoprene forms many products with various lifetimes, here I will present an overview of some important mechanisms which affect oxidation capacity, ozone and aerosol production. Isoprene reacts with OH leading to peroxy radical (ISOP<sub>O</sub><sub>2</sub>) formation. In the presence of NO<sub>X</sub> ISOP<sub>O</sub><sub>2</sub> forms organic nitrates after reacting with NO. These affect levels of both HO<sub>X</sub> (H, OH, peroxy radicals) and NO<sub>X</sub>, acting as a sink ([Mao2013] and references therein).

The first generation of organic nitrates produced by isoprene oxidation range from 7% to 12%, shown in laboratory experiments (todo read abstracts and cite papers in the 3rd paragraph of intro to Mao2013), A portion of isoprene nitrates are

recycled back to  $\text{NO}_X$ , so may serve as a reservoir of nitrogen and allow its transport to the boundary layer of remote regions (TODO: as prior todo).

During the night isoprene is oxidised by  $\text{NO}_3$  radicals, which joins to one of the double bonds and produces organic nitrates in high yield (65% to 85%) ([**Mao2013**])(todo: read mao2013 para 3 cites for) These organic nitrates go on to produce further SOAs ([**Rollins2009**]) (todo read Rollins2009).

Todo: More on ([**Mao2013**]) ()chemistry mechanism used in GEOS-Chem v9.02) For specific information on the isoprene oxidation mechanisms used by GEOS-Chem V10-01 (used in this work), see section 5.1.

## 2.5.10 Radiative Forcing

## 2.6 Dust

Australia is the greatest source of dust in the southern hemisphere producing around 120 Tg yr<sup>-1</sup> ([**Li2008**]), however model validation and analysis over Australia is relatively scarce with more focus applied to the northern hemisphere ([**Fairlie2007, Ridley2013**] ). Atmospheric dust has many direct effects including reduced surface insolation, mineral transfer to remote ocean regions, and health degradation in populated areas ([**Shao2007**]). Direct and indirect effects of dust have many implications which are not fully understood, with many models still struggling to explain the atmospheric cycling of dust at larger scales ([**Rotstayn2011**]).

Australian dust emissions involve various weather conditions, convolving the ENSO cycle with flooding, droughts, and winds. Rivers and rain build up the particulate matter in many areas, these are referred to as fluvial deposits. Fluvial deposits in the Eyre basin increase the dust base load, which will only have mobility during suitable dry weather conditions. These deposits are saltated (loosened from the surface) and transported by strong winds([**Zender2003**]).

Synoptic scale measurements of dust concentrations in Australia are made by the Bureau of Meteorology (BOM) and can be used to estimate dust transport caused by large storms. Single storms have been estimated to move up to 2.5 Tg of dust off shore in a single day. Yearly dust emissions in Australia are somewhere between 10 and 110 Tg yr<sup>-1</sup>. These estimates exemplify the large variability in Australian annual dust transport.

Dust plays a large role in the oceanic carbon cycle, as dust is a major source of oceanic iron (Fe) deposition. Some regions in the ocean are high in nutrients, but low in chlorophyll (HNLC), due to a lack of Fe. Oceanic carbon cycling is a complex system in which Fe is a limiting factor, required by plankton in order to fix atmospheric nitrogen into a more bioavailable form such as ammonia. Atmospheric

deposition into the oceans is a very poorly constrained variable in global models ([**Grand2015**]). Model estimates of trace element oceanic deposition are required to quantify the atmospheric impact due to a dearth of in situ measurements in remote open ocean regions.

Measurements of dissolved iron (DFe) at very low concentrations like those found in surface ocean waters are very easily contaminated, which has contributed to the fragmentary and scarce nature of DFe ocean data sets ([**Rijkenberg2014**]). Recent analysis of the US Climate Variability and Predictability (CLIVAR)-CO<sub>2</sub> Repeat Hydrography Program predicted total deposition flux with uncertainty at a factor of 3.5 ([**Grand2015**]). Some headway has been made with the recent GEOTRACES program which has several transects of the major oceans and measures trace elements over multiple depths including Al, Ba, Cu, Cd, Fe, Mn, Ni, Pb, and Zn.

Total iron (TFe) emissions from dust and combustion sources are estimated (by average of several global models) at approximately 35 Tg yr<sup>-1</sup> and 2 Tg yr<sup>-1</sup> respectively. A two fold increase in Fe dissolution may have occurred since 1850 due to increased anthropogenic emissions and atmospheric acidity. This increase may revert by 2100 due to the affects of emission regulations ([**Myriokefalitakis2015**]). Dust, TFe and DFe have strong temporally and spatial variability, with changes having most impact upon HnLC regions.

Another environmental impact of dust is its contribution to fine particulate matter in the atmosphere. Several studies have shown that long term exposure to fine particulate matter (PM2.5) increases mortality. Estimates of yearly premature deaths related to PM2.5 are  $\sim$  2-3 million ([**Hoek2013, Krewski2009, Silva2013, Lelieveld2015**]). These estimates are made using global atmospheric models or model ensembles to quantify population exposure before applying epidemiological models to estimate the increased death rates. The main source of uncertainty in premature death rates arises from the difference and uncertainties between and within the atmospheric models.

Dust affects global climate change through direct radiative forcing. Uncertainties in the atmospheric dust concentrations make accurate determination of radiative forcing from other sources more difficult ([**IPCC 2013 chap8**]).

## 2.7 Models

### 2.7.1 Chemical Transport Models

Chemical Transport Models (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. CTMs normally solve the continuity

equations simultaneously with chemical production and loss for chemicals under inspection. The continuity equations describe transport of a conserved quantity such as mass, which, solved together with production and loss of a chemical forms the basis for a CTM. This basis enables a record of the chemical densities and transport over time as a model runs. 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:

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

where  $\rho$  is density of  $q$  in the field,  $t$  is time,  $\nabla$  is divergence,  $j$  is the flux (the amount of  $q$  per unit area per unit time entering or leaving the field), and  $\sigma$  is the generation of  $q$  per unit volume per unit time. Note that  $\sigma$  can be positive or negative due to sources and sinks.

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 arbitrary sequential steps in time at each location. The mass balance equation must be satisfied in any realistic long term box model and is as follows:

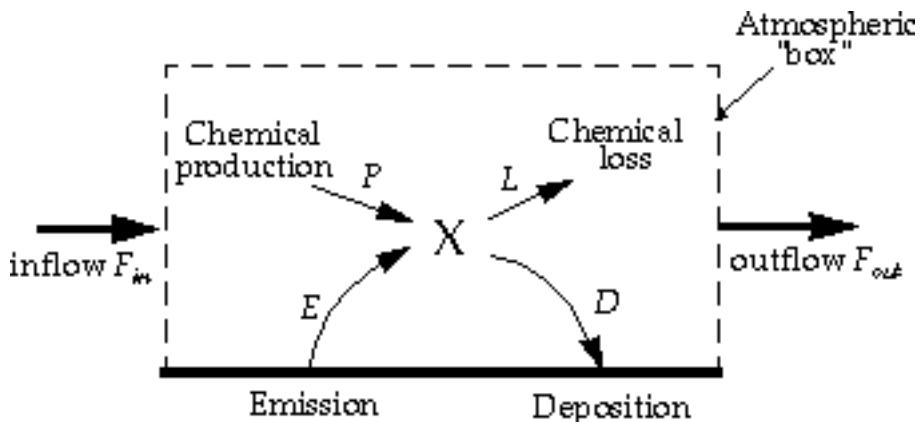
$$\begin{aligned} \frac{dm}{dt} &= \sum \text{sources} - \sum \text{sinks} \\ &= F_{in} + E + P - F_{out} - L - D \end{aligned}$$

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 2.4. 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 can be complex and time consuming.

In many CTMs the isoprene emissions are calculated elsewhere with their own models (EG: [Guenther2006]). These estimates can then be used as boundary conditions. Trace gases with short lifetimes and complex chemistry such as isoprene are often hard to measure which makes verifying model estimates difficult.

### 2.7.2 GEOS-Chem

GEOS-Chem is a well supported global, Eulerian CTM 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). 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



**Figure 2.4:** Standard box model parameters, image taken from [Jacob 1999 book].

atmospheric gases and aerosols. It was developed, and is maintained, by Harvard University staff as well as users and researchers worldwide. Several driving meteorological fields exist with different resolutions, the finest at 0.25 by 0.3125° horizontally at 5 minute time steps with 72 vertical levels.

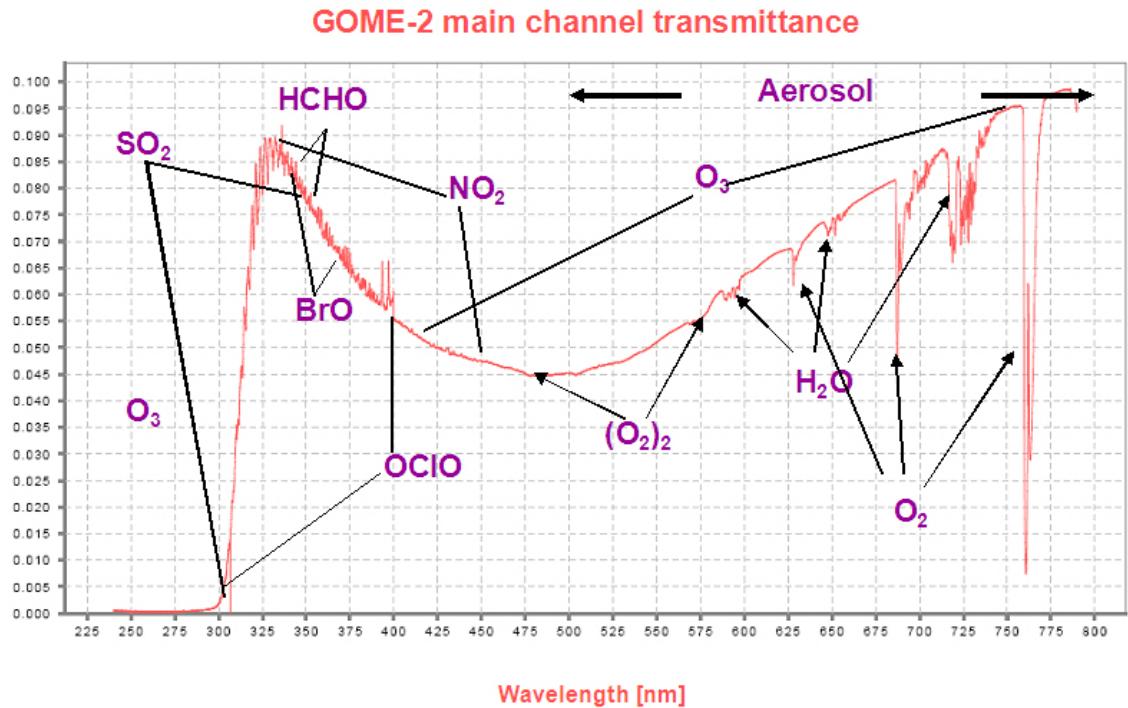
GEOS-Chem uses fire emissions from the GFED4 product ([Giglio 2013]).

Combining satellite data with model outcomes provides a platform for the understanding of natural processes to be tested now and into the future over Australia and anywhere with few in-situ measurements. Due to the low availability of in-situ data covering most of the Australian continent, a combination of the models with satellite data may provide improved understanding of 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.

## 2.8 Satellites

### 2.8.1 Useful satellites

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. They record near nadir (nearly vertical) reflected spectra between around 250-700 nm split into spectral components at around 0.3 nm in order to calculate trace gases including O<sub>3</sub>, NO<sub>2</sub>, and HCHO. An example of a



**Figure 2.5:** An example spectrum showing interferences used for species concentration measurements by GOME-2. Image by EUMETSAT and ESA ([GOME2Image]).

spectrum retrieved from the GOME-2 instrument is given in figure 2.5.

Formaldehyde (HCHO) is often used as a proxy to estimate isoprene emissions ([Marais2012, bauwens2013satellite]). Satellites can use DOAS techniques with radiative transfer calculations on solar radiation absorption spectra to measure column HCHO (eg: [Leue2001]). Several public data servers are available which include products from the satellites just mentioned, including NASA's Mirador (<http://mirador.gsfc.nasa.gov/>) and the Belgian Institute for Space Aeronomy (IASB-BIRA) Aeronomie site (<http://h2co.aeronomie.be/>).

Instruments including MODIS on board the AQUA and TERRA satellites are able to determine aerosol optical depth (AOD), a measure of atmospheric scatter and absorbance. An AOD of under 0.05 indicates a clear sky, while values of 1 or greater indicate increasingly hazy conditions. This is an important atmospheric property allowing us to track dust storms and pollution events as well as determine where measurements from other instruments may be compromised by high interference. Satellite measured AOD requires validation by more accurate ground based instruments like those of AERONET which uses more than 200 sun photometers scattered globally.

### 2.8.2 Comparisons with Models

DOAS methods can be heavily influenced by the initial estimates of a trace gas profile (the *apriori*) 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 unrealistic *a priori* assumptions. One way to remove these *apriori* influences is through the satellite's averaging kernel, which takes into account the vertical profile of the modelled trace gas and instrument sensitivity to the trace gas ([**Eskes2003**, **Palmer2001**]). Measurements done using DOAS often apply a forward radiative transfer model (RTM) such as LIDORT in order to determine a trace gas's radiative properties at various altitudes.

### 2.8.3 DOAS

TODO: some of this is repeated in isoprene chapter satellite section.

The DOAS technique uses solar radiation absorption spectra to measure trace gases through paths of light. The RTM used in DOAS techniques is based on Beer's law relating the attenuation of light to the properties of the medium it travels through. Beer's law states that  $T = I/I_0 = e^{-\tau}$  with T being transmittance,  $\tau$  being optical depth, and I,  $I_0$  being radiant flux received at instrument and emitted at source respectively. Using  $\tau_i = \int \rho_i \beta_i ds$  gives us:

$$I = I_0 \exp \left( \sum_i \int \rho_i \beta_i ds \right)$$

Where i represents a chemical species index,  $\rho$  is a species density (molecules per  $\text{cm}^3$ ),  $\beta$  is the scattering and absorption cross section area ( $\text{cm}^2$ ), and the integral over ds represents integration over the path from light source to instrument. The forward RTM used for satellite data products also involves functions representing extinction from Mie and Rayleigh scattering, and the efficiency of these on intensities from the trace gas under inspection, as well as accounting for various atmospheric parameters which may or may not be estimated (e.g. albedo).

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 should also account for instrument sensitivities to various wavelengths at various altitudes, and is unique for each trace gas under consideration.

# Chapter 3

## Stratosphere to Troposphere Transport of ozone

### 3.1 Background

#### 3.1.1 Historical estimates

Tropospheric ozone is important for both air quality and climate change. The impact of stratospheric ozone on the troposphere is dependent on weather, season, temperature, and many other factors. For example changing ozone in the tropical tropopause layer by 5% causes  $0.5 \text{ K dec}^{-1}$  radiative heating ([Forster2007]). In Australia, records of ozone profiles provided by the Department of the Environment can be used to determine how often stratospheric ozone descends into the troposphere.

Over the industrial period, tropospheric ozone, which is the third most potent greenhouse gas, has been estimated to exert a radiative forcing equivalent to a quarter of the CO<sub>2</sub> forcing. Ozone is present in the troposphere due to a variety of dynamical and photochemical processes, including downward transport from the ozone-rich stratosphere and anthropogenic pollution. The primary sources of tropospheric ozone are chemical creation and stratospheric input, estimated using a model ensemble to be  $5100 \pm 600 \text{ Tg/yr}$  and  $550 \pm 170 \text{ Tg/yr}$ , respectively ([Stevenson2006]). The primary sinks are chemical destruction and dry deposition, estimated to be  $4700 \pm 700 \text{ Tg/yr}$  and  $1000 \pm 200 \text{ Tg/yr}$ , respectively ([Stevenson2006]).

“Further sources of uncertainty in the ozone RF stem from uncertainties in precursor emissions (natural and anthropogenic), as well as changes in climate and stratospheric ozone” ([Stevenson2013]).

There are several ways of estimating stratospheric impact on tropospheric ozone levels, and models can be utilised to this end. Recently, models have been able to trace stratospheric ozone which is transported to the troposphere, which allow

quantification of how much transport should be expected globally or regionally.

### 3.1.2 Tropospheric production

Ozone is a toxic trace gas which increases mortality rates when populations are exposed for extended periods of time. The amount of global premature deaths per year due to atmospheric ozone exposure has recently been estimated at ~150-470 thousand ([**Silva2013**, **Lelieveld2015**]). Long term effects of ozone overexposure increase the risk of respiratory disease and may also increase other cardiopulmonary risks ([**Jerrett2009**]). Further tropospheric ozone enhancements are projected to drive reductions in global crop yields equivalent to losses of up to \$USD<sub>2000</sub> 35 billion per year by 2030 ([**Avnery2011**]), along with detrimental health outcomes equivalent to ~\$USD<sub>2000</sub> 11.8 billion per year by 2050 ([**Selin2009**]).

The Ambient Air Quality (AAQ) National Environment Protection Measure (NEPM), which is the Australian framework for air quality measurement and reporting aiming for “adequate protection of human health and well-being” has set national standards and benchmarks for reporting. The NEPM covers six chemical groups including Ozone (O<sub>3</sub>), and the benchmarks are shown in figure 3.1.

**NEPM standards and goals specified in Schedule 2 of the AAQ NEPM**

Pollutant	Averaging period	AAQ NEPM standard (maximum concentration)	AAQ NEPM goal (maximum number of allowable exceedences)
Carbon monoxide	8-hour rolling average	9.0 ppm	1 day a year
Nitrogen dioxide	1-hour average	0.120 ppm	1 day a year
	1-year average	0.030 ppm	None
Photochemical oxidants – as ozone	1-hour average	0.100 ppm	1 day a year
	4-hour rolling average	0.080 ppm	1 day a year
Sulfur dioxide	1-hour average	0.200 ppm	1 day a year
	1-day average	0.080 ppm	1 day a year
	1-year average	0.020 ppm	None
Particles as PM <sub>10</sub>	1-day average	50.0 µg/m <sup>3</sup>	5 days a year
Lead	1-day average	0.50 µg/m <sup>3</sup>	None
Particles as PM <sub>2.5</sub> <sup>1</sup>	1-day average	25.0 µg/m <sup>3</sup>	Gather sufficient data nationally to facilitate a review of the Advisory Reporting Standard
	1-year average	8.0 µg/m <sup>3</sup>	

<sup>1</sup> Reporting standard only

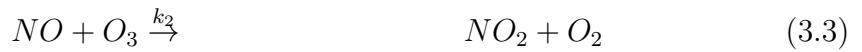
**Figure 3.1:** NEPM standards taken from National Environment Protection Council annual report 2012-2013 ([**nepc annuals**]).

The primary source of ozone in the lower troposphere is chemical formation following emissions of precursor gases, including VOCs, and NO<sub>X</sub>. Globally the greatest sources of NO<sub>X</sub> include fossil fuel combustion (~50%), biomass burning

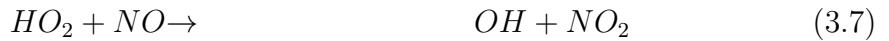
(~20%), lightning, and microbial activity in soils ([**Delmas1997**]). Estimates using CHASER (a global Chemical Transport Model (CTM)) constrained by measurements from two satellites as well as the in-situ measurements taken through LIDAR and aircraft (INTEX-B) put global tropospheric  $\text{NO}_X$  emissions at  $45.4 \text{ TgN yr}^{-1}$  in 2005 ([**Miyazaki2011**]).

The majority of this chemical formation is due to photochemical oxidation of carbon monoxide (CO), methane ( $\text{CH}_4$ ), and other Volatile Organic Chemicals (VOCs) in the presence of nitrogen oxides ( $\text{NO}_X \equiv \text{NO} + \text{NO}_2$ ) ([**Stevenson2006**]).

Photolysis of  $\text{NO}_2$  forms  $\text{NO} + \text{O}(3\text{P})$ , which combines with  $\text{O}_2$  to form  $\text{O}_3$ , leading to reaction with NO to form  $\text{NO}_2 + \text{O}_2$ . These reactions reach a steady state where  $\text{O}_3$  is proportional to the ratio between  $\text{NO}_2$  and NO ([**Sillman2002**]). The following formulae show an example of this with CO, however, similar reactions occur for many VOCs:

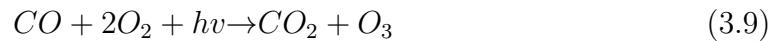


$$[\text{O}_3] \sim \frac{k_1 [\text{NO}_2]}{k_2 [\text{NO}]} \quad (3.4)$$



(3.8)

where  $k_1$  and  $k_2$  are reaction rates, and  $h\nu$  represents photons. The balance of these reactions is:



Note that this reaction pathway only occurs during the day.

Isoprene ( $\text{C}_5\text{H}_8$ ) is a precursor to ozone through radical oxidative chemistry. Isoprene in the atmosphere reacts rapidly with hydroxyl radicals ( $\text{OH}$ ) and then  $\text{O}_2$  to form peroxy radicals ( $\text{RO}_2$ ). These react with nitrogen oxides and can lead to ground-level ozone formation similarly to the CO reaction listed prior.

Formaldehyde (HCHO) together with  $\text{NO}_2$  regulate tropospheric oxidation capacity through  $\text{O}_3$  production, as well as being health hazards. The HCHO/ $\text{NO}_2$  ratio can be used to determine whether surface  $\text{O}_3$  is  $\text{NO}_2$  or VOC limited ([**Mahajan2015**]). If  $\text{O}_3$  is  $\text{NO}_2$  limited then an increase in  $\text{NO}_2$  will increase  $\text{O}_3$  levels while an increase in HCHO will not, and vice versa when  $\text{O}_3$  is HCHO limited.  $\text{NO}_2$  is a common pollutant in populated areas, released primarily by combustion in power generation

and transport. Outside of cities in Australia, VOCs and NOx are emitted from biogenic sources, although lightning, and biomass burning (most clearly in the Northern Territory) also play a role ([**Guenther2006**, **VanDerA2008**]).

### 3.1.3 Stratosphere to Troposphere ozone Transport (STT)

While the amount of tropospheric ozone is small compared with that found in the stratosphere, it is an important constituent. Ozone-rich air mixes irreversibly down from the stratosphere during meteorologically conducive conditions ([**Sprenger2003**, **Mihalikova2012**])); these are referred to as Stratosphere - Troposphere Transport events (STTs). In the extra-tropics, STTs most commonly occur during synoptic-scale tropopause folds ([**Sprenger2003**]) and are characterised by tongues of high Potential Vorticity (PV) air descending to low altitudes. PV is a metric which can be used to determine whether a parcel of air is stratospheric, based on its local rotation and stratification. Within the troposphere PV values are typically low, increasing rapidly into the stratosphere due to the increased static stability.

These ozone rich tongues become elongated and filaments separate from the tongue which mix into tropospheric air. Stratospheric ozone brought deeper (lower) into the troposphere is more likely to affect the surface ozone budget and tropospheric chemistry ([**Zanis2003**, **Langford'2009**]). A high correlation is found between lower stratospheric and tropospheric ozone ([**Terao2008**]) with the highest STT associated with the jet-streams over the oceans in winter. Irreversible STT of ozone is important for explaining tropospheric ozone variability ([**Tang2010**]). Stratosphere to Troposphere ozone transport can potentially increase regional surface ozone levels above safe levels ([**Zhang2014**]).

While photochemical production is the dominant source, stratosphere to troposphere transport of ozone is also important and climate change may drastically increase this quantity ([**Hegglin'2009**]). In a future climate, a warmer, wetter troposphere will change the chemical processing of ozone. Dynamical processes such as STT, boundary layer ventilation and convection changes will alter tropospheric ozone distributions. [**Hegglin'2009**] estimate that climate change will lead to increased STT of the order of 30 (121)  $\text{Tg yr}^{-1}$  relative to 1965 in the southern (northern) hemisphere due to an acceleration in the Brewer Dobson circulation.

STT events are characterised in the ozonesondes' vertical profiles of ozone as altitudes in the troposphere where the ozone mixing ratio exceeds a specified threshold. Usually stratospheric ozone mixes irreversibly down into the troposphere in a synoptic-scale tongue of air: the vertical ozone profile observed by the ozonesonde depends upon the time in this cycle that it is observed ([**Sprenger2003**]). As such, the altitude of the tropospheric ozone peak due to an STT event, and the amplitude

of the event above the background tropospheric ozone profile, vary in space and time.

## 3.2 Instruments and data sets

### 3.2.1 Atmospheric Infrared Sounder (AIRS)

AIRS is an instrument on board NASA's AQUA satellite, which overpasses the globe roughly daily on a sun-synchronous orbit gathering measurements at 1:30pm local time. AIRS is a high spectral resolution spectrometer with 2378 bands in the thermal infrared ( $3.7 - 15.4 \mu\text{m}$ ) and 4 bands in the visible ( $0.4 - 1.0 \mu\text{m}$ ). One of the products shared by NASA is total column carbon monoxide (AIRS3STD ([AIRS3STD])). Using this product to exclude possible fire plume transport from STT analysis is done in section 3.5.2.

### 3.2.2 Sondes

Ozonesondes are weather balloons with an attached instrument which measures ozone concentrations roughly every 100m up to around 30km. These ozonesondes provide a high-vertical resolution profile of ozone.

In this work sonde data from three sites are utilised: Davis (lat, lon, UTC +7), Macquarie Island (lat, lon, UTC +11), and Melbourne (lat, lon, UTC +11). Ozonesondes are launched approximately weekly from these three sites, we examine data from up to 2013, starting in 2004 at Melbourne and Macquarie Island, and 2006 at Davis. More frequent ozonesonde launches occur at Davis during the spring ozone hole season than at other times of the year ([Alexander2013]).

### 3.2.3 European Centre for Medium-Range Weather Forecasts (ECMWF) Re-Analysis - Interim data set (ERA-I)

Ozonesondes provide much higher vertical resolution profiles of ozone than that available from reanalyses products. The downside is that one data point per week from an ozonesonde release is too low to be in itself useful to diagnose the evolution of STT exchange over time-scales associated with normal synoptic scale weather patterns present in the extra-tropics. The ECMWF provides a useful meteorological model based on assimilated data. Here, ozonesonde data are supplemented with the ERA-I dataset ([Dee2011]) to enable construction of an STT exchange climatology. The dataset is called a re-analysis due to the fact that the whole timeline gets re-run

when the climatological model is updated. This provides uniform data output over a long time period.

The ERA-I data we used for synoptic weather was of one degree horizontal resolution with pressure levels at 200, 300, 400, and 500 hPa. For individual cases ERA-I data was downloaded at .25 degree horizontal resolution with the full 34 pressure levels from 1000 to 1 hPa. Note that 34 levels up to the top of the atmosphere is not enough to determine vertical transport when considering only a single vertical profile, however a precise view of coincident weather can determine probable cause of ozone flux.

## 3.3 STT Detection

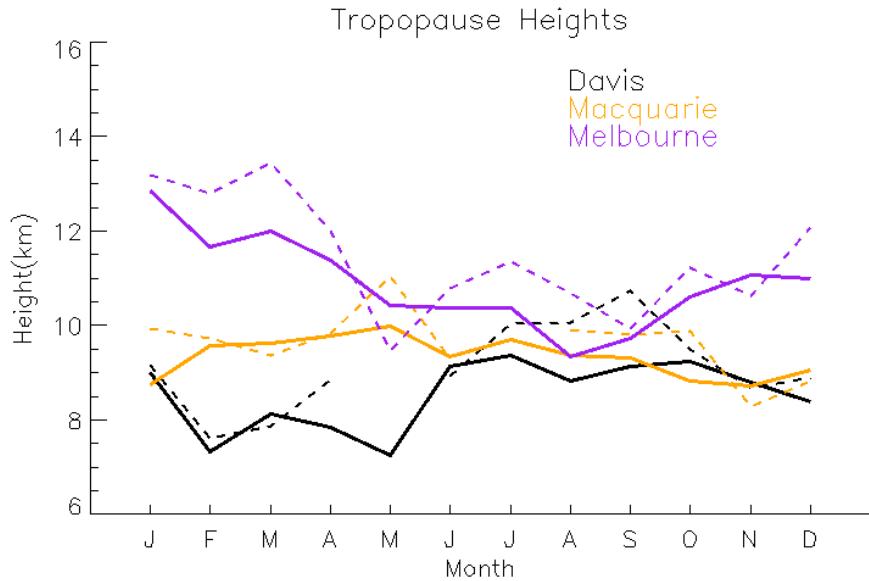
### 3.3.1 Aim

Using several years of ozonesonde flights from three locations spanning the latitudes of the Southern Ocean, we create and examine a quantitative method for detecting ozone STT events from ozone vertical profiles. With a quantitative method of STT detection characterisation of the seasonal cycle of STT events and determination of their contribution to the total amount of tropospheric ozone is possible. Examination of the ozone intrusions and case studies allows us to relate these STT to meteorological events. Finally we use the same filtered sonde data in order to extract a lower bound estimate of how much of the tropospheric column ozone is due to STT events.

### 3.3.2 Tropopause Heights

Two definitions of the tropopause height are calculated: the standard lapse rate tropopause ([**WMO1957**]), and the ozone tropopause ([**Bethan1996**]). At Davis, the ozone tropopause defintion is modified for polar sites, following [**Tomikawa2009, Alexander2013**]. While the ozone tropopause can be less robust during stratosphere-troposphere exchange, it performs better than the lapse rate tropopause at polar latitudes in winter and near jet streams in the lower stratosphere ([**Bethan1996**]). The lower of these two tropopause altitude-definitions is referred to as the tropopause for this study. This choice of the lowest altitude of the tropopause avoids occasional unrealistically high tropopause heights due to perturbed ozone or temperature measurements.

The monthly mean tropopause altitudes at each location are shown in Figure 3.2, along with the subset of altitudes from profiles for which an STT event was determined. The seasonal cycle in tropopause altitude at Melbourne is clearly apparent, as is the decreasing tropopause altitude poleward. Seasonally averaged ozone as



**Figure 3.2:** Monthly mean tropopause altitudes (minimum of lapse-rate and ozone defined tropopauses). Dashed lines show 'event only' seasonal tropopause altitudes.

recorded over the three stations (Figure 3.3) shows increased ozone extending down through the stratosphere during the peak STT months over Melbourne. It is worth noting that tropopause altitudes at Davis may exceed 11 km altitude under certain synoptic conditions ([Alexander2013]): the relation of tropopause altitude with individual STT events will be investigated in detail below.

### 3.3.3 Fourier bandwidth (or bandpass) Filtering

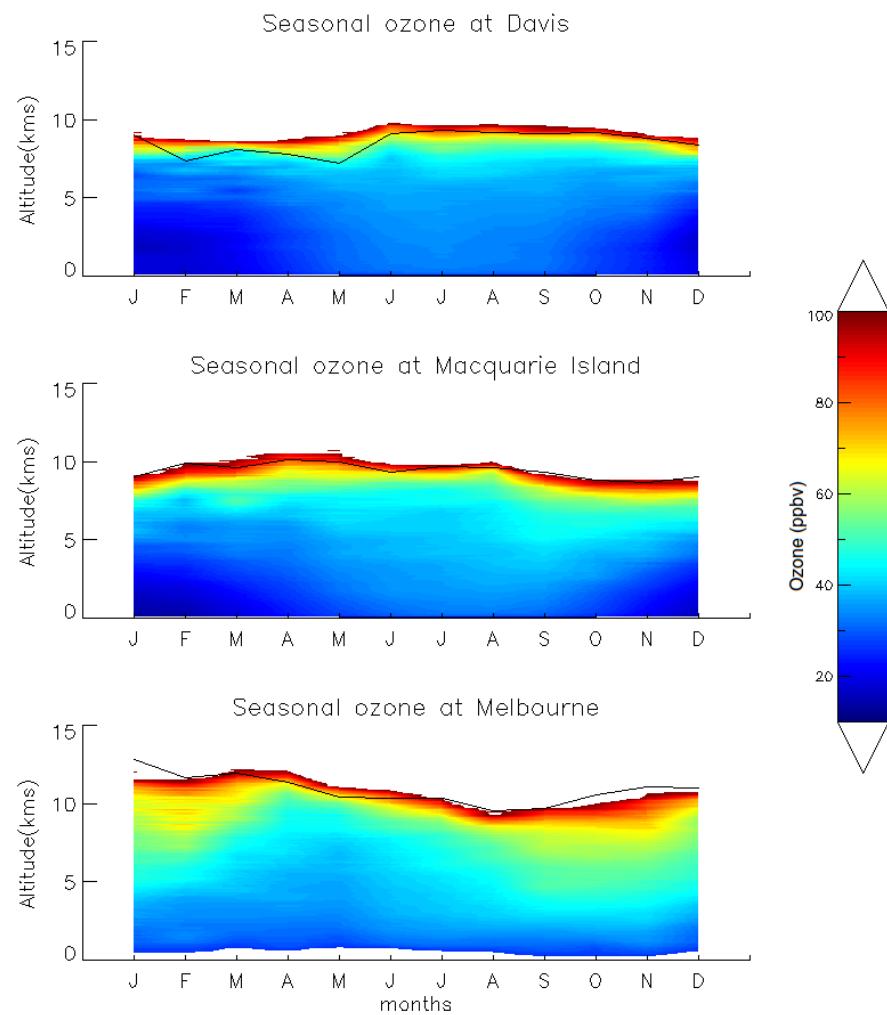
A Fourier bandwidth filter can remove components of a line based on the components frequency. For example: a noisy ozone profile can be cleaned by removing the high frequency components, while growth of ozone with altitude in the profile can be removed as a low frequency component.

TODO: Add description, equation(s) and maybe a simple example plot here as well as limits of filter. The basic idea of a Fourier bandwidth filter is that any finite length function can be written as a series of trigonometric functions:

$$f(t) = \sum_i C_{w_i} \cos(w_i t - \theta_i) \quad (3.10)$$

Once we split our function into specific frequencies ( $w_i$ ) we can simply remove the terms which fall outside our desired wavelength range. One limitation is that the reconstructed function may be quite different at either end of the input range.

In the continuous spatial domain, this is done is through taking the Fourier



**Figure 3.3:** Seasonally averaged ozone over Davis, Macquarie, and Melbourne measured by ozonesondes. Black solid lines show seasonal tropopause heights.

transform which can convert our function to a complex frequency domain:

$$F(w) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(t)e^{-iwt} dt$$

where t describes the spatial input range (in our case altitude). Then it's possible to remove the portion of the function outside the desired frequency before inverting the Fourier transform:

$$f(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} F(w)e^{iwt} dw$$

This whole process is slightly changed when we consider discrete dimensions as we must whenever solving numerically or handling real world resolved datasets. There is a “shortcut” method which

### 3.3.4 Bandwidth filter applied to ozonesondes

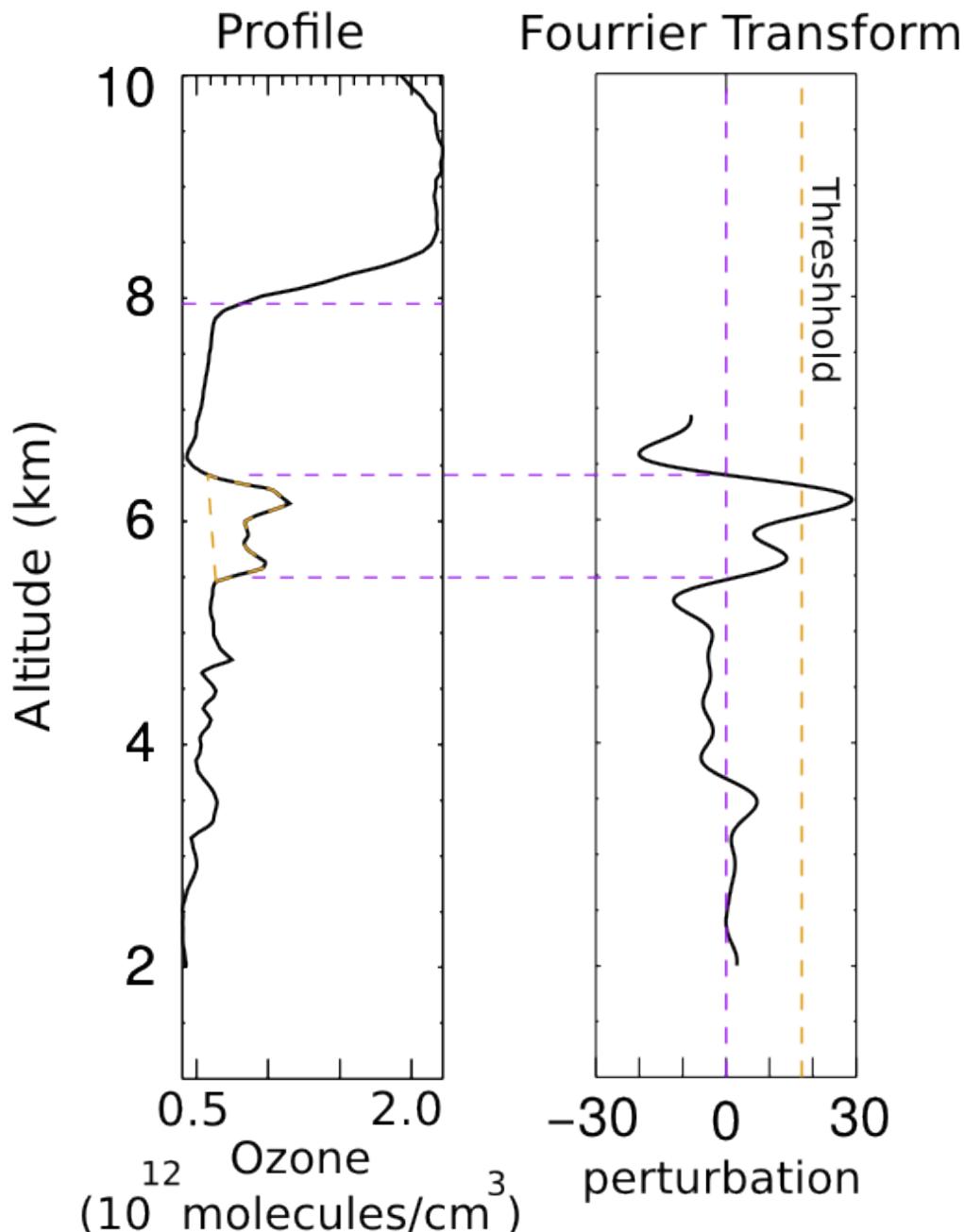
With a Fourier bandwidth filter used on ozone profiles over Davis, Macquarie, and Melbourne we quantitatively determine instances of Stratosphere to Troposphere Transport events through the following method. The vertical profiles of ozone volume mixing ratio are linearly interpolated to a regular grid with 20m resolution up to 14km altitude and are then bandpass filtered so as to retain perturbations which have vertical scales between 0.5km - 5km. The choice of band limits is set empirically, but we note that to define an STT event, a clear increase above the background ozone level is needed, and a vertical limit of  $\sim 5$  km removes seasonal-scale effects. The ozone perturbation profile is analysed at altitudes from 4 km above the surface (to avoid surface pollution events) and 1 km below the tropopause (to avoid the sharp transition to stratospheric air producing spurious false positives). Perturbations above the 99 th percentile (locally) of all ozone levels are initially classified as STT events.

In order to remove unclear ‘near tropopause’ anomalies we remove events where the gradient between the maximum ozone peak and the ozone at 1 km below the tropopause is greater than  $-20$  ppbv  $\text{km}^{-1}$  and simultaneously require that the perturbation profile does not drop below zero between the event peak and the tropopause. The addition of these filters removes several events, each with an ozone peak which could not be definitively said to be seperated from the stratosphere.

To provide a conservative estimate of ozone flux into the troposphere for each event, the ozone concentration is integrated vertically over the interval for which an STT event is identified. An example of an ozone profile is illustrated in Figure 3.4 and indicates how the algorithm detects an STT event, defines the event boundaries, and calculates the ozone flux.

# Ozone at Melbourne

## 2004/01/08



**Figure 3.4:** (a) An ozone profile between 2km altitude and the tropopause (indicated by the dashed vertical line). The 'flux' area shows the estimate of stratospheric impact on tropospheric ozone. (b) The 99th percentile of filtered ozone perturbations (green dashed line) and the technique for determining the vertical extent of the 'event' (red dashed and solid lines).

### 3.3.5 Case Studies

We examine two STT case studies in detail to illustrate the synoptic scale conditions in which they can occur above Melbourne.

A cut-off low pressure system passed over Melbourne on 3 February 2005 (Figure 3.5b). The ozonesonde profile indicated low lapse-rate and ozonesonde tropopauses (both  $> 450$  hPa, see Figure 3.5a). An ozone intrusion into the troposphere is identified by our detection algorithm at  $\sim 520$  hPa.

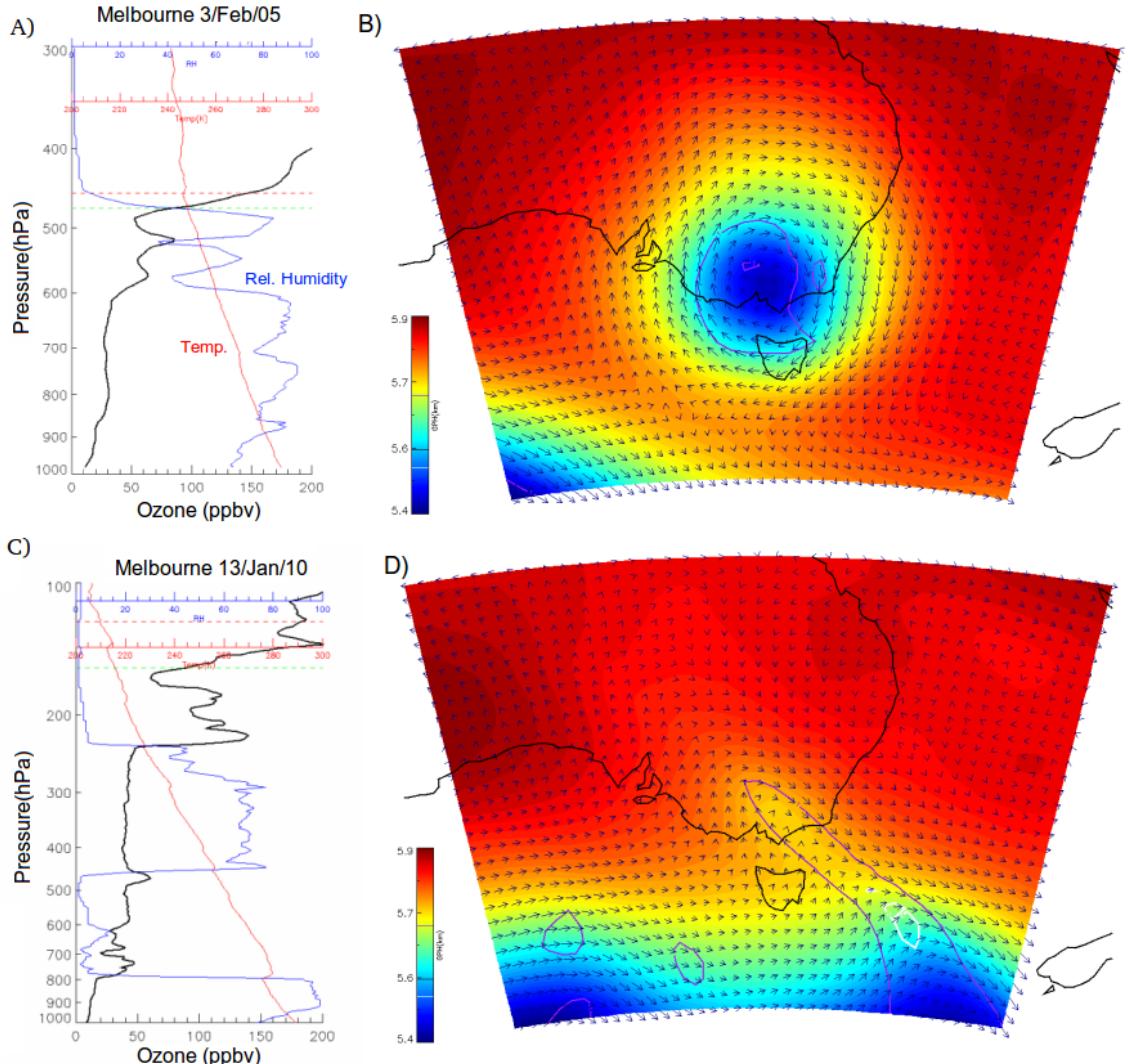
STT events also occur during frontal passages, an example of which is illustrated in Figure 3.5d over south-eastern Australia. The tropopauses are much higher at this time and an ozone intrusion is identified centred around 200 hPa. Note the separation between this intrusion and the ozone tropopause (marked by the green dashed line), indicating the start of the stratosphere above Melbourne. During the frontal passage, stratospheric air descends and streamers of ozone-rich air likely break off and mix into the troposphere ([Sprenger2003]). TODO: talk about pvu lines here.

The relative humidity profiles are anticorrelated with ozone in the upper troposphere for these events, indicating again the stratospheric origin of the ozone-rich air mass. TODO: show latlon plots for the sites and describe lowered tropopause in more detail here. Some of this stratospheric air gets mixed into the troposphere, with one ozonesonde column showing an intrusion at around 200 hPa, with dry ozone rich air peaking below the tropopause (Figure 3.5c).

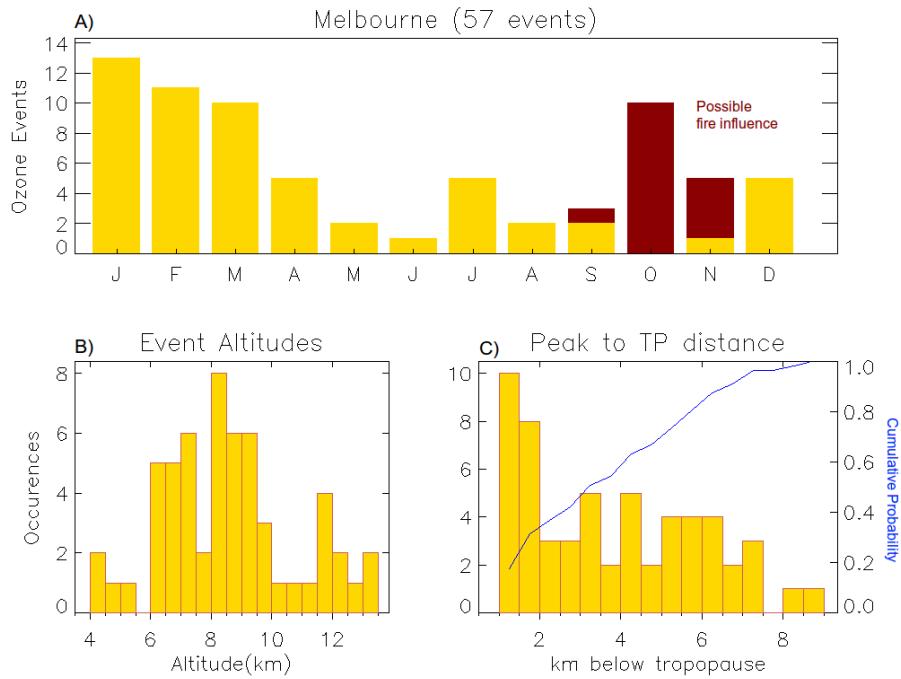
### 3.3.6 Site summaries

Running the ozonesonde dataset through our bandpass filter and analysing the results allows an overview of the yearly cycle and event characteristics. These seasonal cycles and event characteristics for each of the three locations are presented in Figure 3.6 to Figure 3.8. There is an annual cycle in the occurrence frequency of STT events (with a summertime peak) above Melbourne and Macquarie Island. However, the occurrence frequency of STT events above Davis is relatively constant throughout the year.

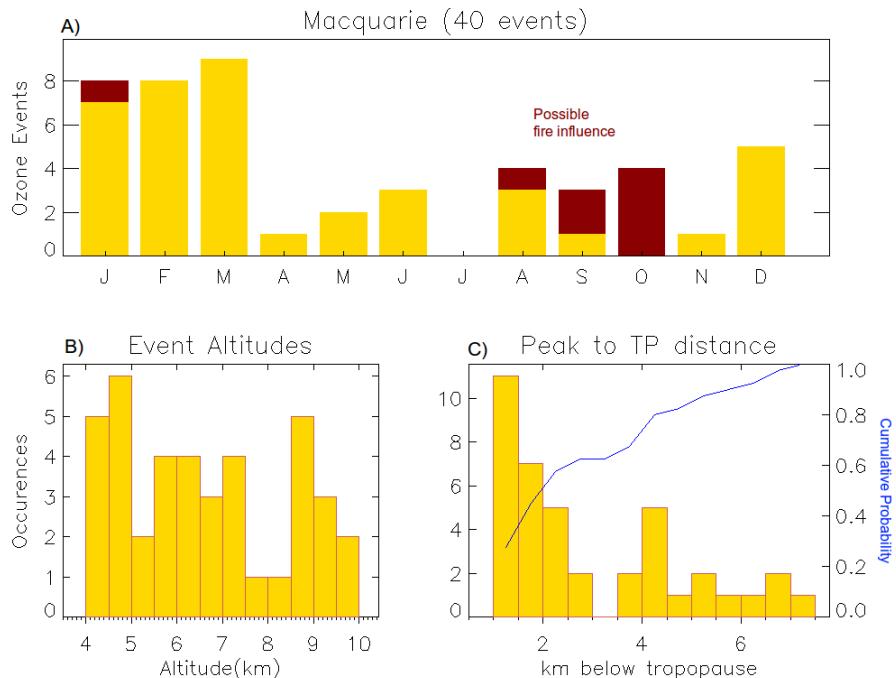
The majority of events occur within 3 km of the tropopause at both Melbourne and Macquarie Island, and within 2 km of the tropopause at Davis. STT event altitudes most commonly occur at 6 – 10 km above Melbourne and below 8 km at Davis but are distributed more evenly in altitude at Macquarie Island.



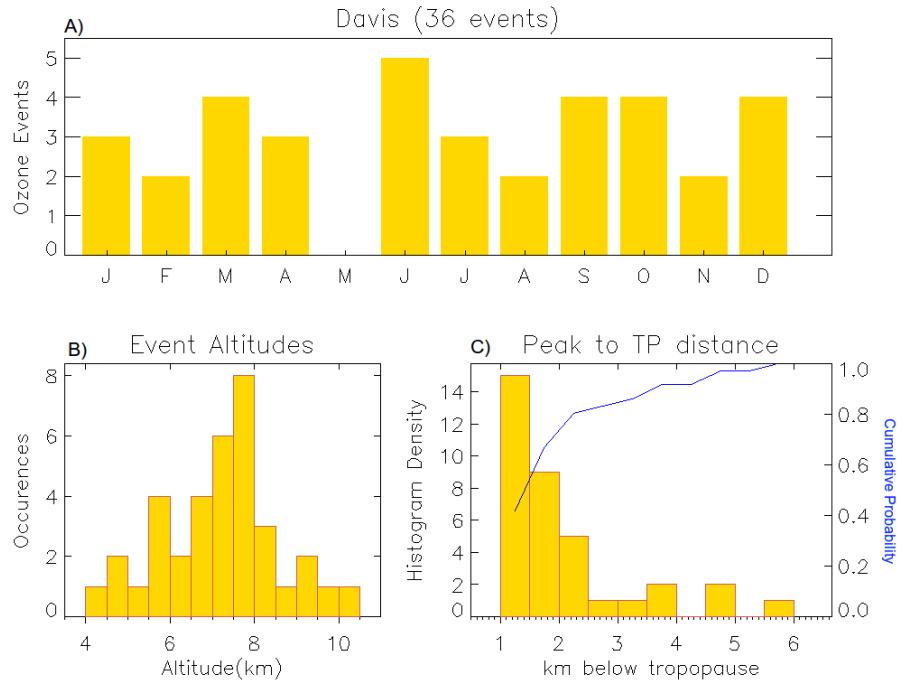
**Figure 3.5:** Vertical profiles show ozone ppbv (black line), relative humidity (blue line), and temperature (red line) for (a) 3 February 2005 and (c) 13 January 2010. Synoptic weather maps show the 500 hPa pressure level taken from the ERA-Interim reanalysis on (b) 3 February 2005 and (d) 13 January 2010. Vectors show wind direction and speed while the colour indicates the geopotential height. Also visible are the line contours of potential vorticity units, 1 PVU in purple and 2 PVU (often used to determine dynamical tropopause height) in white.



**Figure 3.6:** The climatology of STT events at Melbourne: (A) Events sorted by month from the entire Melbourne ozonesonde dataset. The events filtered out as possibly smoke plume influenced are indicated in red. (B) The occurrence distribution of the ozone peak altitude. (C) The distance between the ozone peak and the tropopause (bars) and the cumulative probability function of these distances (blue line).



**Figure 3.7:** As for Figure 3.6 except showing the Macquarie Island STT events.



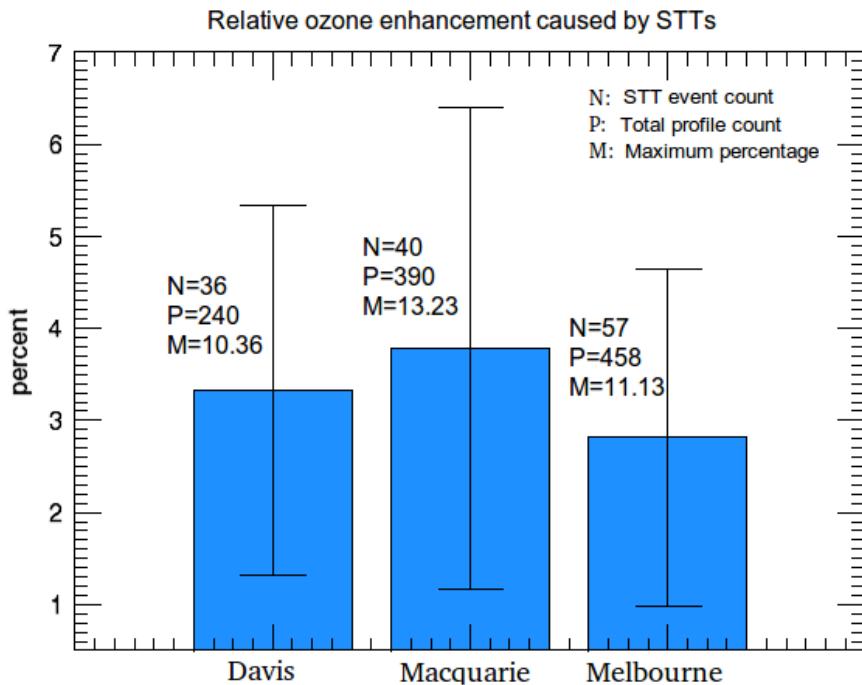
**Figure 3.8:** As for Figure 3.6 except showing the Davis STT events.

## 3.4 Stratosphere to Troposphere flux analysis

### 3.4.1 Determining a minimum estimate of stratospheric influence

Determining how much ozone is transported from the stratosphere with only a two dimensional line vector of ozone concentrations is questionable. It is assumed that the two dimensional profile is a vertical straight line, which accurately represents the ozone profile in the horizontal dimensions. A conservative estimate on stratospheric ozone influence is determined through analysis of the ozone spike above a baseline amount for each column. This is a lower bound as it ignores dispersed ozone, baseline enhancements, and any secondary peaks which may also be due to stratospheric transport.

Using our estimate of STT ozone flux (see section 3.3.3) we find a lower bound for the STT ozone flux over each of our three sites, excluding possible fire influence. Figure 3.9 shows the climatological mean fraction of total tropospheric column ozone attributed to stratospheric ozone intrusions at each site, on days when an STT event occurs. These flux amounts are calculated after removal of the biomass burning events, although leaving the burning events in changes the means by less than 5%. The mean fractions of stratospheric ozone are 2–4%, although the largest fractional ozone in the tropospheric column attributed to stratospheric air exceeds 10% at all locations.



**Figure 3.9:** Fraction of total tropospheric column ozone attributed to stratospheric air intrusions during STT events. Error bars indicate one standard deviation.

## 3.5 Non-STT influences on ozone at southern latitudes

### 3.5.1 Fire Plumes

Ozone production due to fire smoke plumes is complex and affected by photochemistry, fuel nitrogen load, and atmospheric plume interactions both during transport and at the plume's destination. Ozone precursors include nitrogen oxides ( $NO_x = NO + NO_2$ ) and non methane volatile organic compounds (NMVOCs). Large biomass burning events emit substantial ozone precursors, some of which are capable of being transported far from their origins.  $NO_x$  can be transported far downwind due to peroxyacetyl nitrate (PAN) thermal decomposition and this can lead to enhanced ozone far from the source of a fire ([Jaffe 2012]).

[Pak2003] show a tropospheric ozone spike over SE Australia in September and using back trajectories suggest that the air masses over Melbourne on this day also passed over South Africa and the higher altitude air also passed through South America. The Australian fire season spans most of the year, sweeping southwards as shown in figure and may occasionally affect the ozone columns above either of Macquarie and Melbourne. However, the influence from regional burning is unlikely to affect the ozone concentration above 4 km altitude and most of the smoke from

the northern half of Australia is not transported over these sites.

South of 10° S, [Jaffe 2012] estimates that 158 and 54 Tg per year of CO and ozone are emitted emitted and produced respectively by wildfires, with 50% uncertainty in the ozone production. Due to this possible source of ozone column perturbation, care must be taken to exclude biomass burning influence when analysing STE impacts.

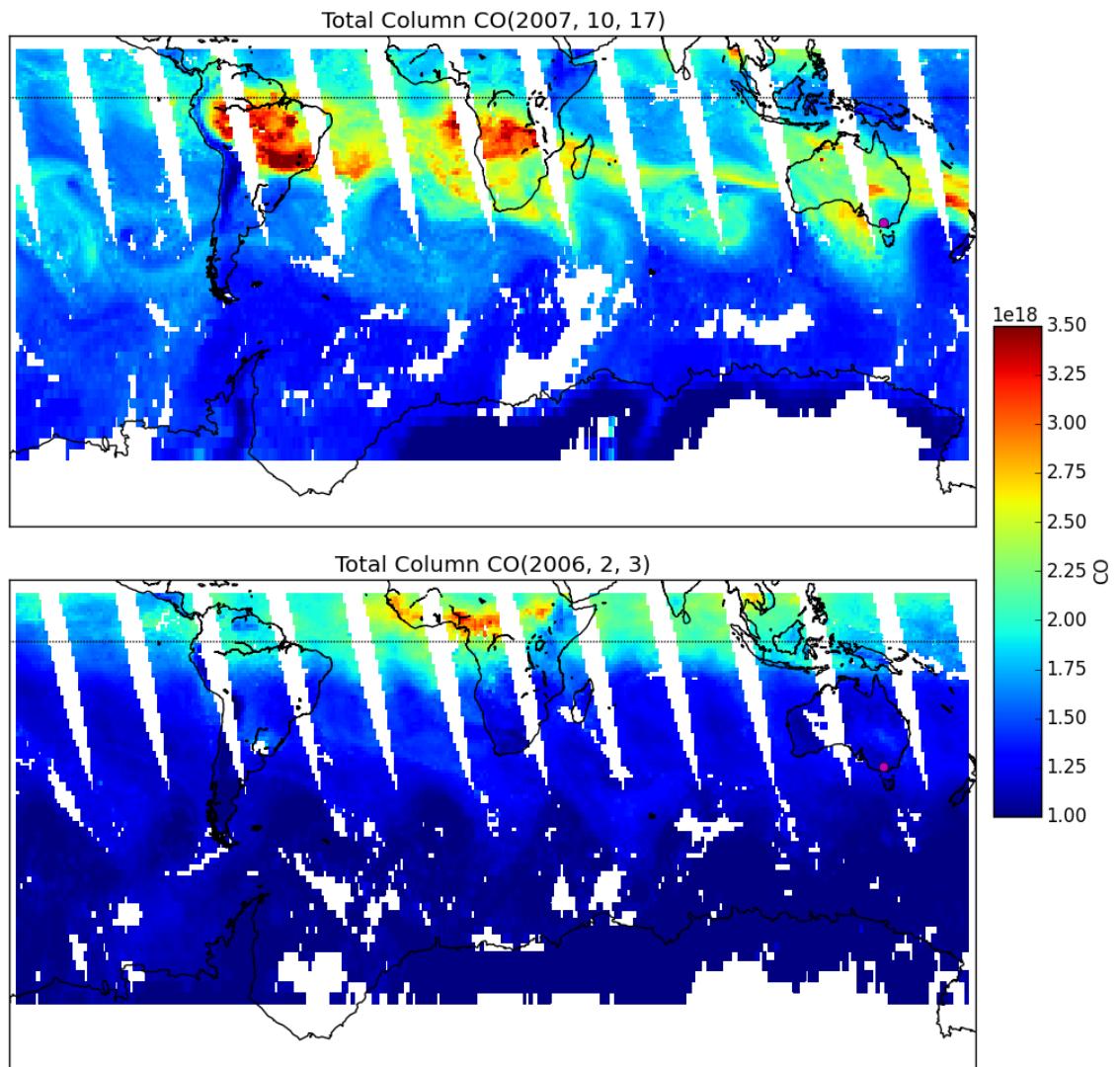
### 3.5.2 Transport Exclusion

Biomass burning from the southern tropics (South Africa and South America) is a likely source of interference to and perturbation of ozone measurements. Transported BB plumes influence the southern mid latitudes generally between July and December, as indicated by the methane and CO enhancements ratio ( $\Delta CH_4/\Delta CO$ ) which averages 0.31 in this time ([Pak2003]).

One method of detecting transported emissions influences is through satellite column analysis as done in [Sinha2004]. Smoke plumes from biomass burning create a heavy haze as well as elevated CO concentrations which can be seen from satellites.

Ozone production due to fire smoke plumes is complicated and dependent on many chemical and meteorological factors. Using high CO levels as a proxy in order to determine where fire smoke plumes exist has been done in several studies (eg: ([Edwards2003, Sinha2004, Edwards2006, Mari2008])). The AIRS (Atmospheric Infrared Sounder) instrument on board the AQUA satellite records, among other trace gases, column CO. The AQUA satellite overpasses most of the globe twice per day, using the data from the ascending mode swathes at local time 1330 allows some idea of the atmospheric CO on a given day.

In this work a visual inspection of vertical columns of CO (provided by NASA ([AIRS3STD])) over the southern hemisphere is used to exclude possible foreign smoke plume influence on the ozone profile at our three sites. Whenever high CO concentrations coincide with sonde detected ozone events it's possible that the tropospheric ozone spike could be due to transported chemicals. All occasions where these coincidences occur are removed. Using a scale of 1e18 up to 3.5e18 molecules/cm<sup>2</sup> can show burning influence as exemplified in figure 3.10. Using 458 ozonesonde profiles over Melbourne, 72 ozone events are detected, of which 15 are discarded as possibly caused by transported fire smoke plumes. Over Macquarie island 48 events are detected from 380 ozonesondes of which 8 are discarded due to possible smoke influence. We also include on Figure 3.6 to Figure 3.8 the events which have possible fire influence. These events are concentrated in Spring at melbourne and Macquarie Island.



**Figure 3.10:** AIRS total column CO image showing two days separate days of swathes. The top panel shows an example of an excluded ozone event which could have been caused by a transported biomass burning plume on October 17th, 2007.

## 3.6 GEOS-Chem ozonesonde comparison

### 3.6.1 Comparing data to GEOS-Chem

Ozone sondes are useful for looking at specific locations with high resolution, and in this work they provide an estimate of both STT occurrence rates and STT ozone flux. This information can be used in conjunction with global scale information in order to extrapolate ozone transport over a large area. GEOS-Chem is used to simulate the global ozone concentrations. In order to check that the model is reasonable, some simple validation is performed comparing sonde data to colocated grid boxes in GEOS-Chem, using the tropospheric columns of ozone as the metrics. After the model is validated this way, an extrapolation is performed and the stratospherically sourced ozone is estimated over the latitude range from 30°S to 75°S (TODO: Update these numbers with final decided ones, and give a reason). Examination of the GEOS-Chem output also gives us an insight as to whether the simulation can be used on its own in order to estimate STT event distribution and magnitude.

A secondary reason to compare ozonesondes to GEOS-Chem is to check that GEOS-Chem chemistry is reasonable and robust. Model validation at any scale increases confidence in the model as well as providing background information to scientists who seek to expand or improve the model.

GEOS-Chem is a global chemical transport model ([[Bey2001](#)]), which includes transport, emission, deposition, chemical production and destruction of ozone and 103 other trace gases throughout the troposphere along with stratospheric chemistry, including photolysis. Stratosphere-troposphere coupling is calculated using the stratospheric unified chemistry extension (UCX) ([[Eastham2014](#)]), which includes a further 28 trace gases. 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.

GEOS-Chem values used within this chapter are based on a simulation using UCX from 2004-2013, with 2004 used as a spinup year. The simulation uses 2° latitude by 2.5° longitude horizontal resolution, with 72 vertical levels from the surface to 0.1 hPa. Biogenic emissions of organic chemicals are determined by the Model of Emissions of Gases and Aerosols from Nature (MEGAN) version 2.1 extended by Guenther et al ([[Guenther2012](#)]). Anthropogenic emissions are given by the Emissions Database for Global Atmospheric Research (EDGAR) version 4.2. Our simulation was modified from the standard v10-01 to a fix a bug in the treatment of the Total Ozone Mapping Spectrometer (TOMS) satellite data used to calculate photolysis (see TODO:wiki entry for fix or supplementary). GEOS-Chem uses measurements from the Total Ozone Mapping Spectrometer (TOMS) satellite instrument with the Fast-JX photolysis rate solver customised for use in the stratosphere

in order to simulate stratospheric ozone photolysis.

### 3.6.2 Determining tropospheric ozone from GEOS-Chem

GEOS-Chem allows certain diagnostics, along with any tracer, to be output at higher temporal resolution for a list of latitude and longitude based boxes. Storing output every six hours allows an examination of vertical profiles at a list of specific latitudes and longitudes during both day and night. Using the ozone mixing ratio ( $C_{O_3}$  in molecules  $O_3$  per molecule of air) at 72 vertical layers and the air density ( $N_{Air}$  in molecules per  $cm^3$ ), provides us with the ozone density profile ( $N_{O_3}$ ):

$$N_{O_3}[z] = C_{O_3}[z] \times N_{Air}[z]$$

where  $z$  is the vertical level index.

In order to determine the tropospheric ozone column ( $TC_{O_3}$  in molecules  $cm^{-2}$ ), we first use the modelled tropopause pressure (TPP in hPa) to determine where the troposphere ends. The TPP is used to determine how many vertical levels exist in the troposphere. This is done through comparison with the pressure edges of each level. The linear fraction (frac) of the level containing the TPP is then obtained from the pressure edges below and above the TPP (pb and pa respectively):

$$frac = \frac{p_b - TPP}{p_b - p_a}$$

At every time step, using the above calculations along with the layer heights ( $H$  in cm), the tropospheric ozone ( $TVC_{O_3}$  in molecules  $cm^{-2}$ ) is determined as follows:  $N_{O_3}[z] \times H[z]$  gives the vertical profile of molecules  $cm^{-2}$ , which is then summed up to the TPP:

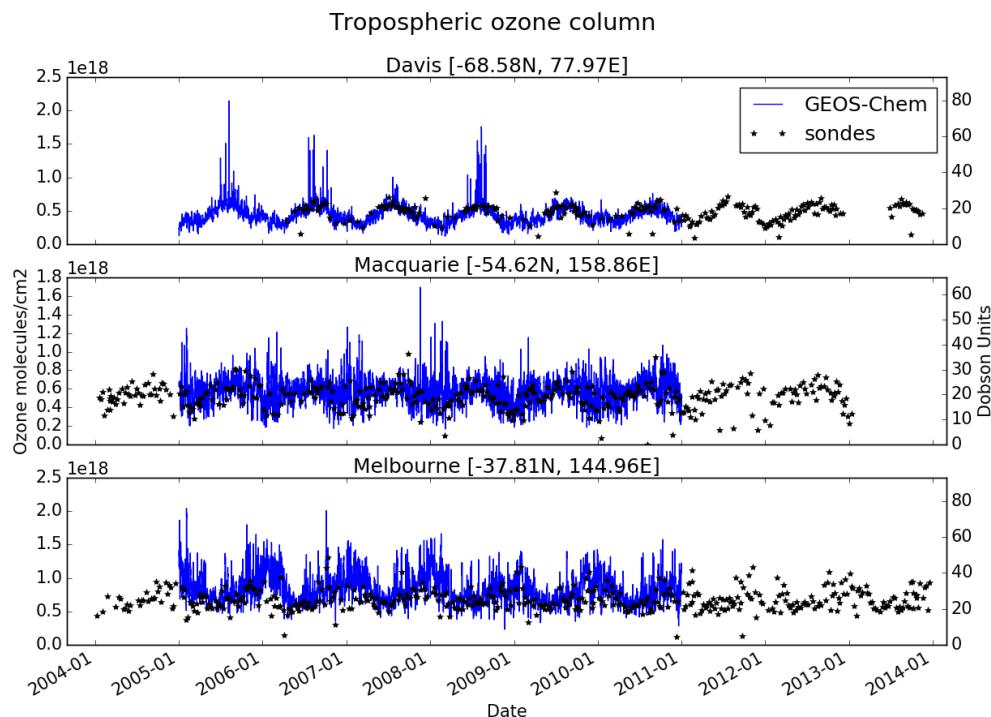
$$TVC_{O_3} = \sum_{z=0}^{z_{TPP}-1} (N_{O_3}[z] \times H[z]) + frac \times N_{O_3}[z_{TPP}] \times H[z_{TPP}]$$

where  $z_{TPP}$  is the index of the level containing the tropopause.

Figure 3.11 shows the time series of tropospheric ozone ( $TVC_{O_3}$ ) simulated over our three stations from January 1 2005 to January 1 2011. Coplotted as stars are the measured tropospheric ozone columns calculated using the GPH and ozone partial pressure recorded by the ozone sondes. The GEOS-Chem simulation uses horizontal resolution of  $2^\circ$  latitude by  $2.5^\circ$  longitude, and the profiles and totals above each site are actually the average over the horizontal area containing each station. There is good agreement between the modelled tropospheric ozone columns and the measurements, with matching seasonal cycles and magnitude. All three sites have a discernable yearly cycle, with Davis having the least spread as well as the greatest

outliers. These outlying tropospheric ozone columns all occur during the July to September months (TODO: reasons for why? simulated summer turbulence?). Both Macquarie Island and Melbourne have more variance in their tropospheric columns. For both sites the data is more spread out over the winter months. This variability is shown in more detail in figure 3.12.

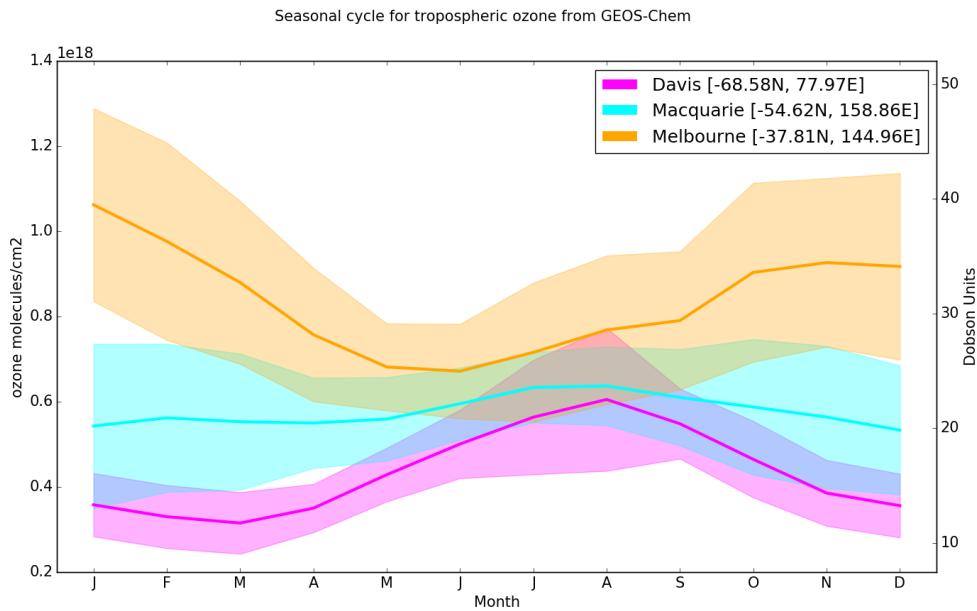
Figure 3.12 shows the yearly averaged tropospheric ozone column over our three sites. This figure is created using the monthly averages along with their standard deviations. Melbourne has the largest tropical ozone column throughout the year, with a summer peak and winter minimum. Davis has the lowest ozone levels, with an opposite seasonal cycle to that of Melbourne. At Macquarie Island, there is a more subtle seasonal cycle, with slightly more ozone occurring in winter than in summer.



**Figure 3.11:** Tropospheric ozone in molecules  $\text{cm}^{-2}$  every six hours simulated by GEOS-Chem (blue line) from January 1 2004 until December 31 2010. The ozonesonde calculated tropospheric ozone columns are shown as stars, each representing one measurement.

### 3.6.3 Ozone profiles compared with GEOS-Chem

GEOS-Chem provides 72 vertical levels of data, of which about 30 will be in the troposphere. The ozone sondes generally have more than 100 vertical levels below

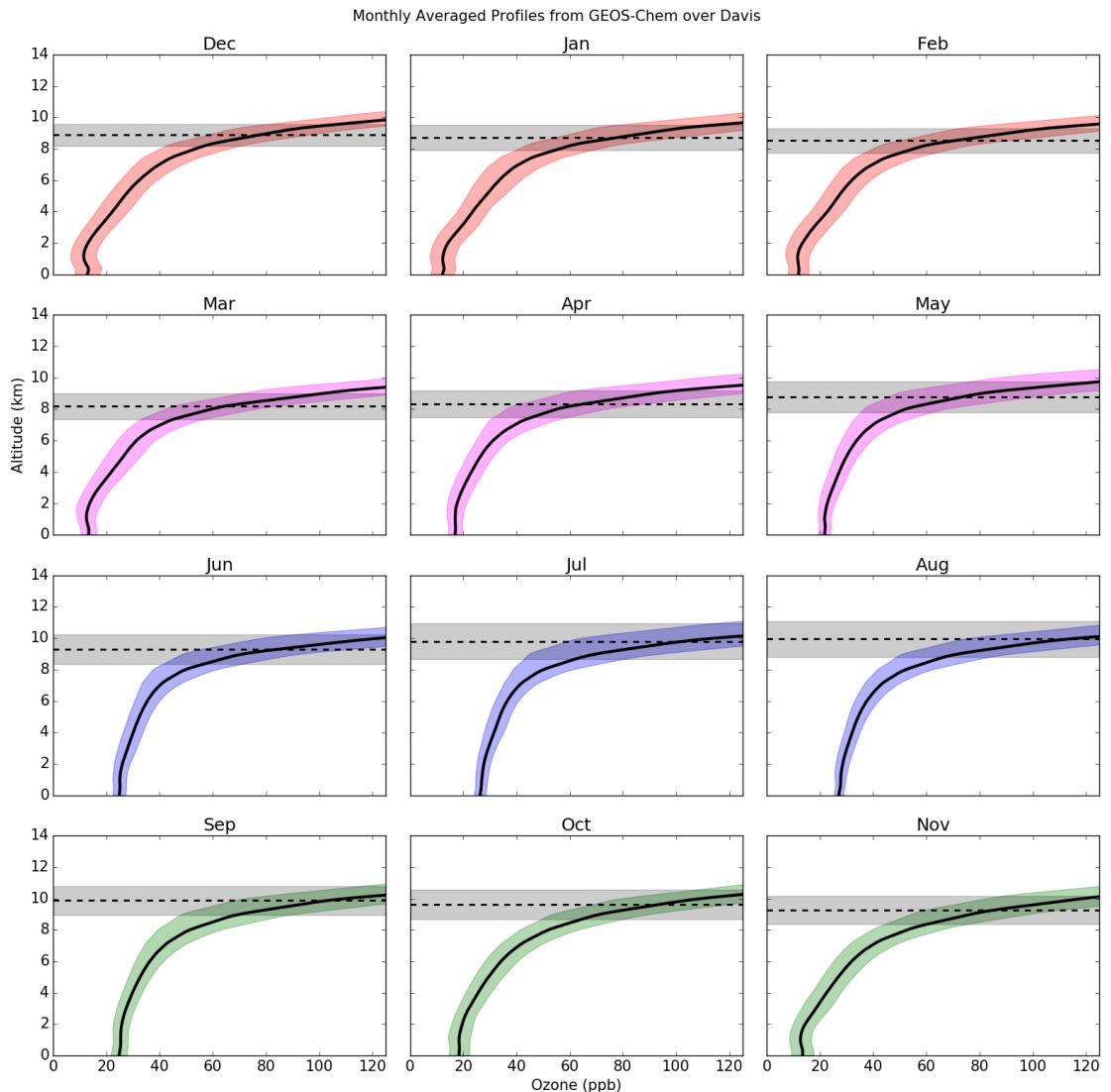


**Figure 3.12:** Tropospheric ozone in  $\text{molecules cm}^{-2}$  seasonal cycle simulated by GEOS-Chem from January 1 2004 until December 31 2010. The monthly averages are taken for each year and the mean plotted with one standard deviation shaded.

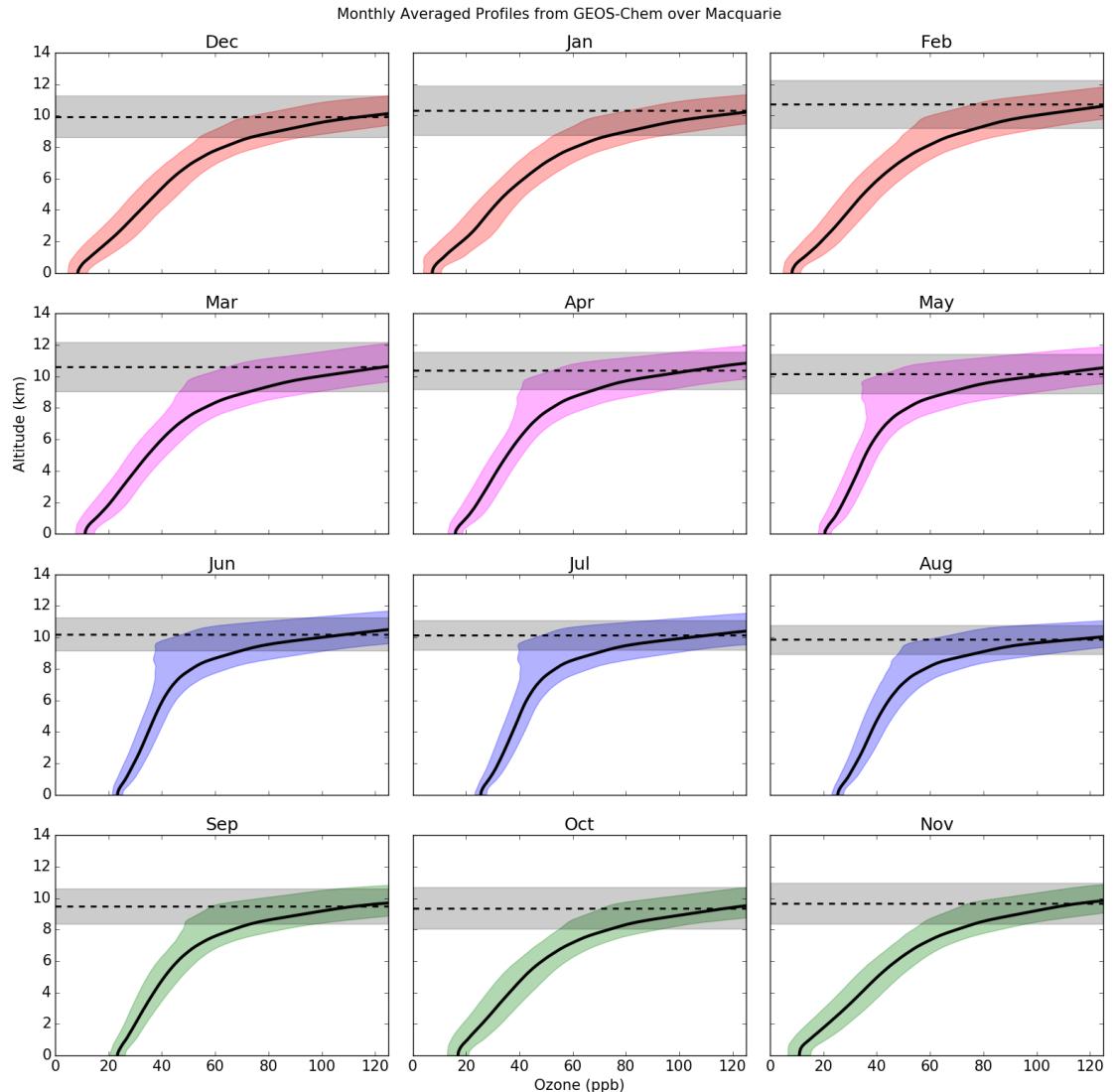
the tropopause. Interpolating the vertical levels to a standard set of altitudes allows an examination of both the mean ozone profiles and the standard deviation at each altitude. For each of the GEOS-Chem and sonde vertical profiles, the data is interpolated to 100 points between 0 and 14 km, at which the mean and standard deviation is determined for each month.

Figures 3.16-3.18 show the simulated monthly averaged ozone profile at Davis, Macquarie Island, and Melbourne respectively. The shaded areas show  $\pm 1$  standard deviation, while the horizontal dotted line shows the mean tropopause height. The effect of pollution and mainland influence can be seen over Melbourne, mostly during the summer months (DJF), as the lower altitudes have increased ozone mean as well as more variance. Also the yearly cycle of tropopause height for each site is noticeable, and it matches the ozonesonde recorded tropopause seasonal cycle. Examining the mean profiles at particular hours only shows a noticeable difference (not shown unless someone tells me to) over Melbourne, the other two stations are very similar regardless of which hour we examine.

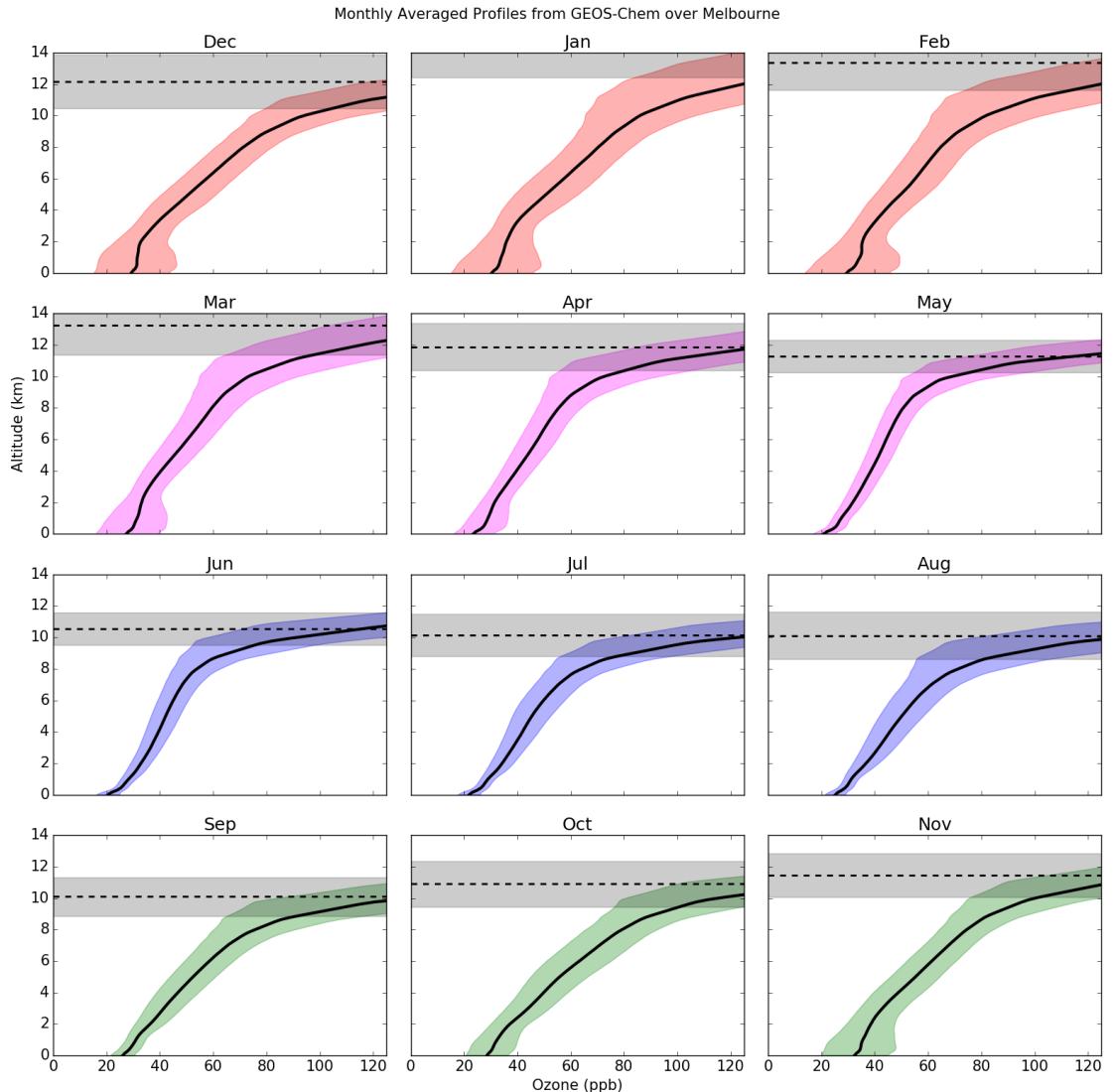
Figures ??-?? show the measured monthly averaged ozone profile at Davis, Macquarie Island, and Melbourne respectively. At Davis, the tropopause level slightly lower in summer and autumn than in winter and spring. This yearly variation matches the tropopause heights simulated by GEOS-Chem, and is opposite to the yearly tropopause height cycle seen in Melbourne. Macquarie does not have a seasonal cycle of tropopause heights, which agrees with the GEOS-Chem profiles, although the GEOS-Chem tropopause heights are slightly higher at all months. For



**Figure 3.13:** Tropospheric ozone (ppb) simulated by GEOS-Chem over Davis from January 1 2004 until December 31 2010, averaged monthly. Horizontal dotted line shows the mean tropopause height, shaded areas show one standard deviation.

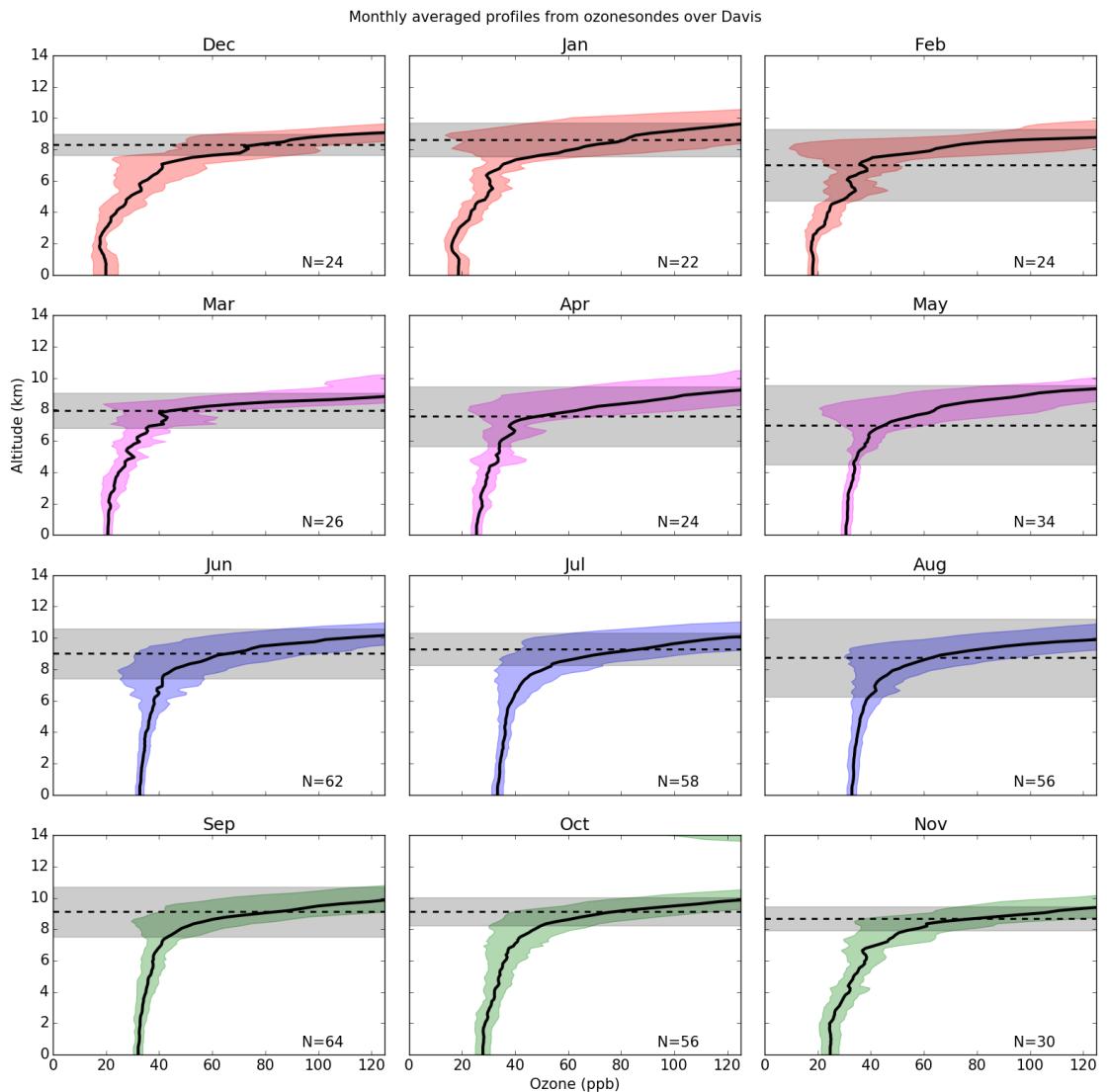


**Figure 3.14:** As figure 3.16 over Macquarie Island.



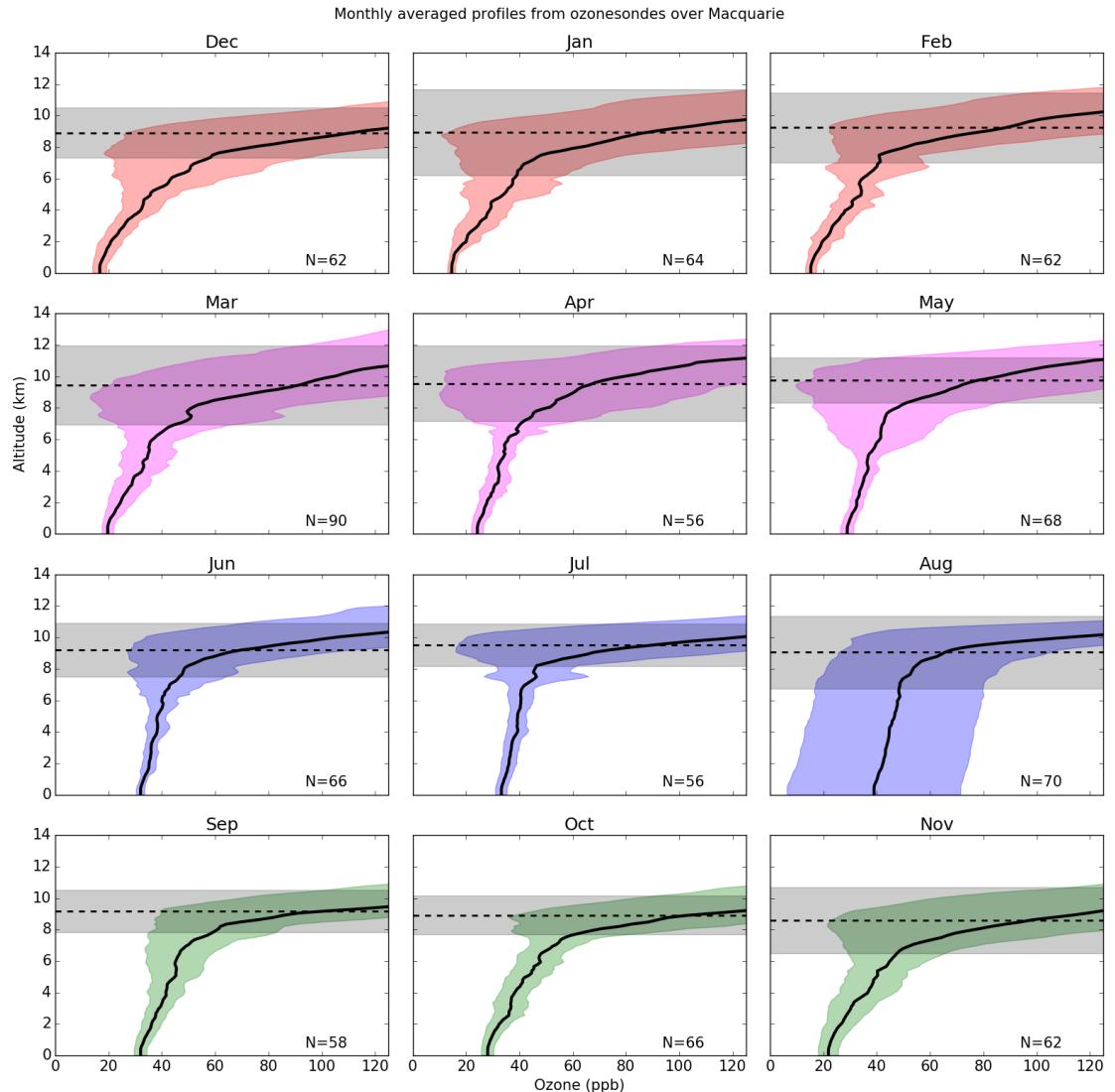
**Figure 3.15:** As figure 3.16 over Melbourne.

Melbourne the gradient of increasing ozone with increasing altitude is stronger in the GEOS-Chem output than the sonde measurements.

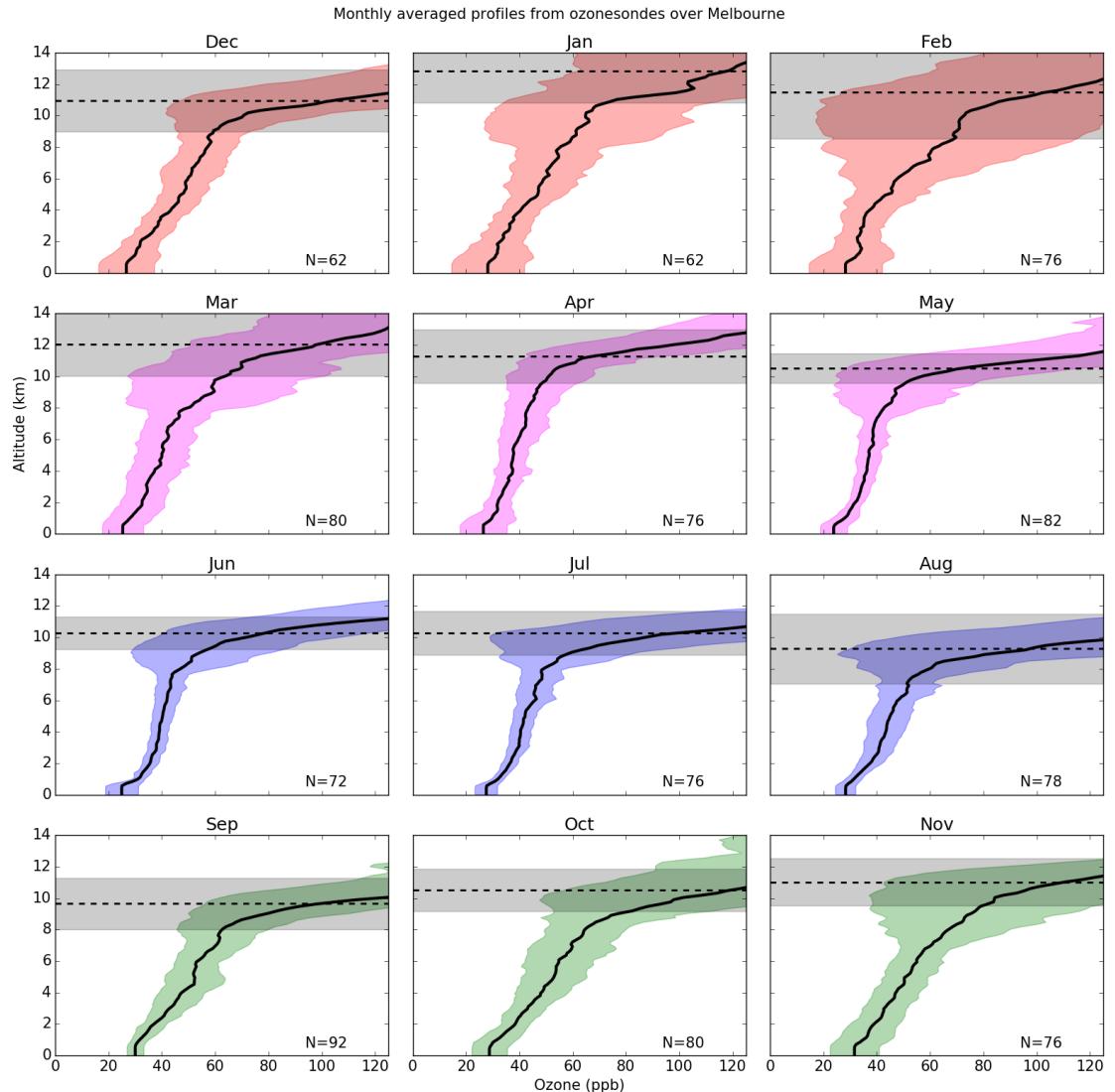


**Figure 3.16:** Tropospheric ozone (ppb) measured by ozone sonde over Davis, averaged monthly. Horizontal dotted line shows the mean tropopause height, shaded areas show one standard deviation.

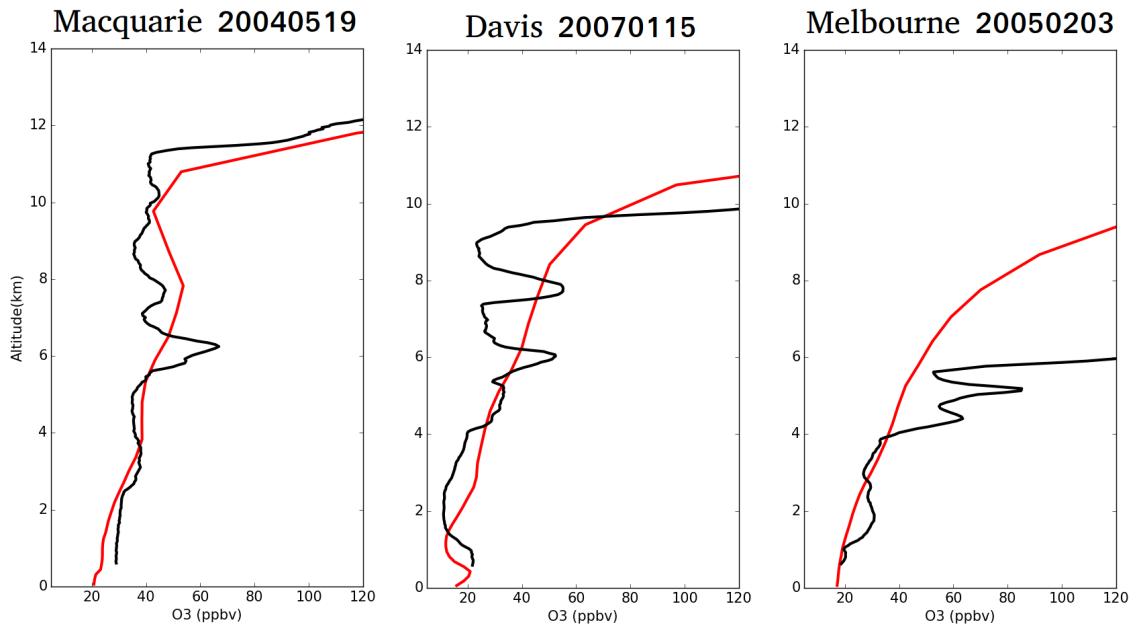
Although GEOS-Chem reasonably matches the ozonesonde tropospheric ozone column, it does not have the resolution required to capture STTs. Figure 3.19 shows the best and worst comparisons of ozone profiles up to 14 km between the ozonesondes and GEOS-Chem. The model output is shown in red, and is the average over  $2^{\circ}$  latitude by  $2.5^{\circ}$  longitude which contain the respective sonde release site. The vertical resolution from GEOS-Chem is too low to allow detection of STTs, there are roughly 30 vertical levels up to the tropopause, while sondes have upwards of 100. These best and worst examples are determined qualitatively, and are shown as examples of the model-measurement differences.



**Figure 3.17:** As figure 3.16 over Macquarie Island.



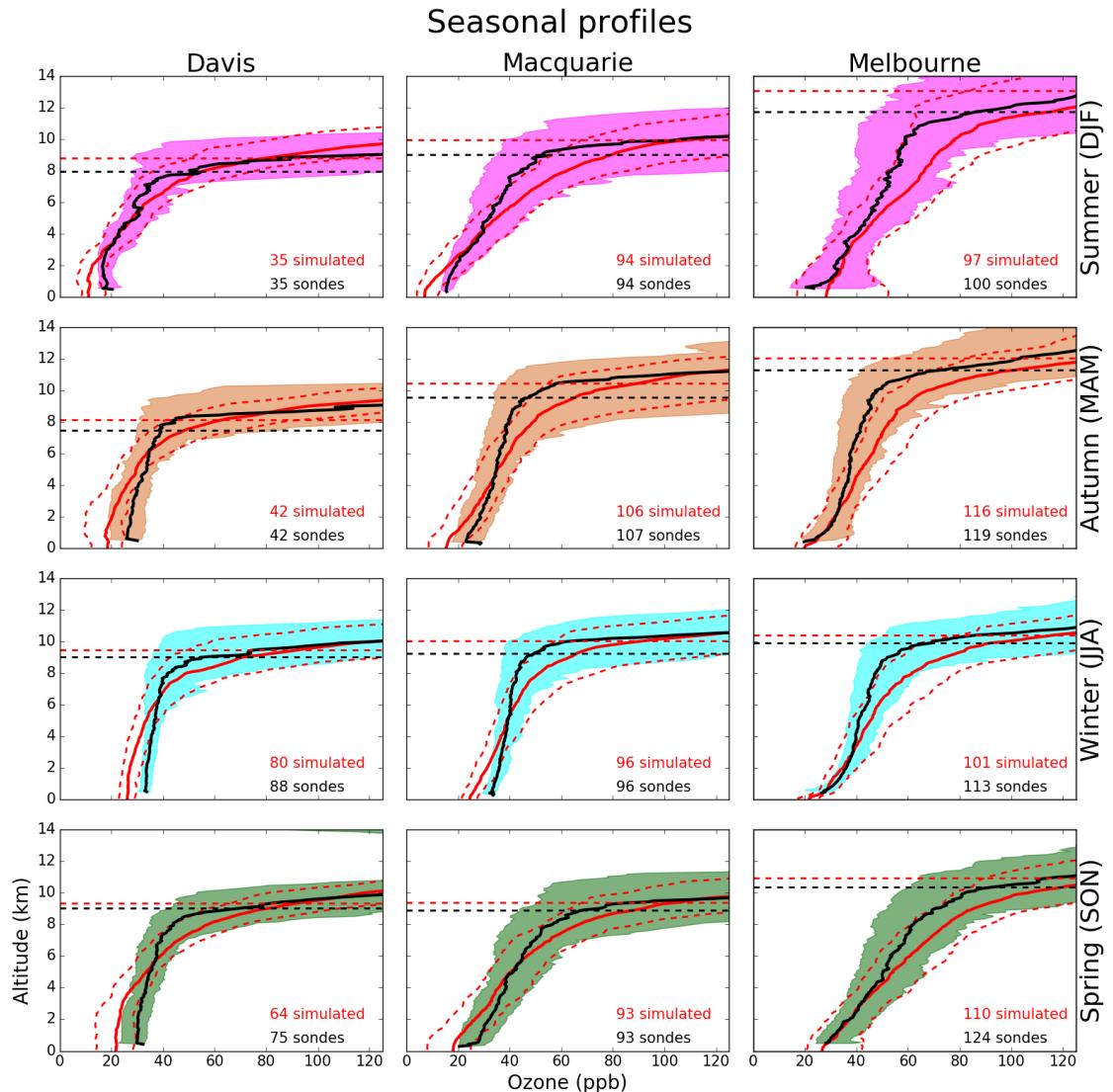
**Figure 3.18:** As figure 3.16 over Melbourne.



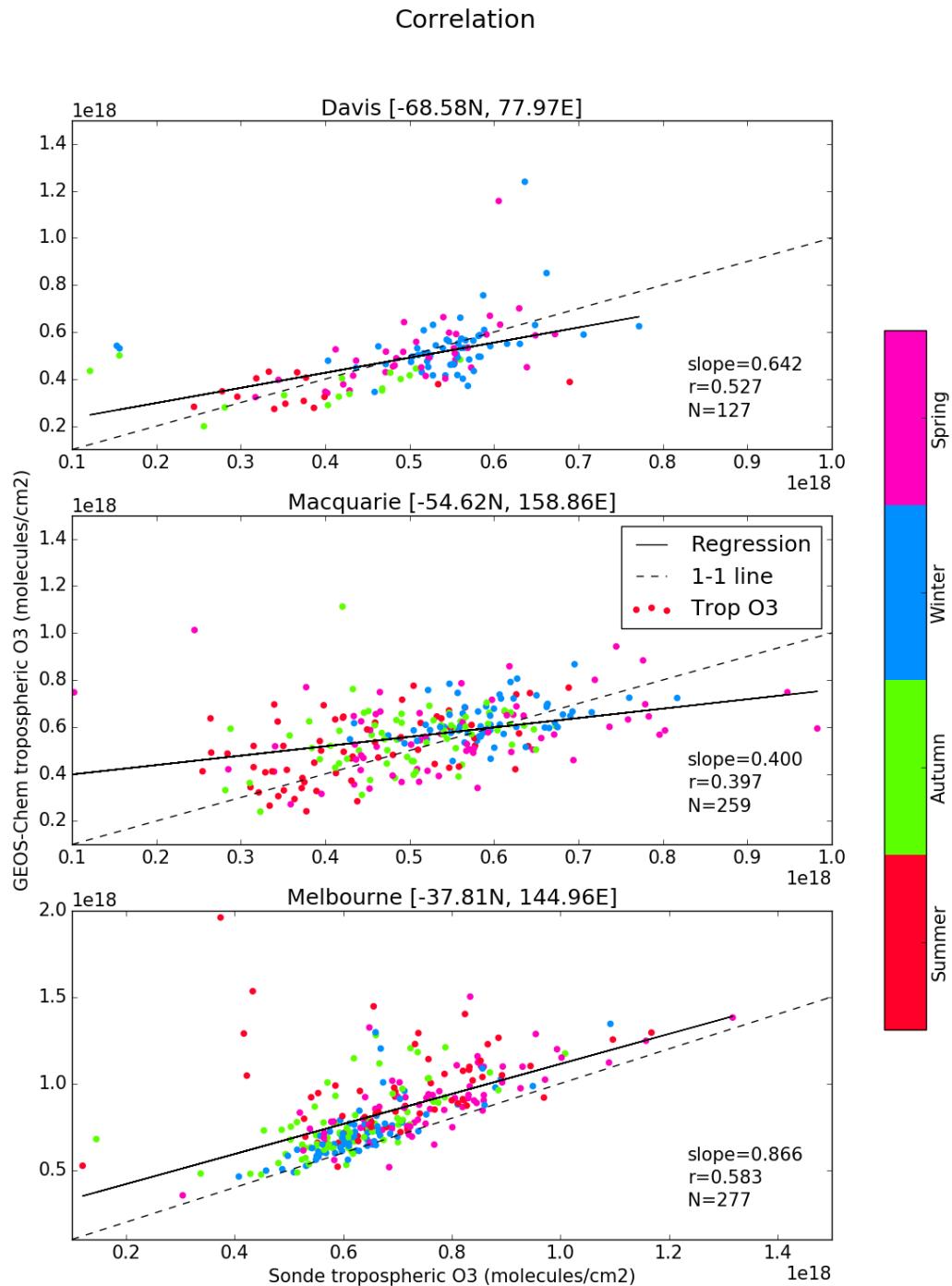
**Figure 3.19:** Ozonesonde profiles (black) against GEOS-Chem profiles (red) above Davis, Macquarie Island, and Melbourne respectively from left to right. These columns from show the best matching profiles at each site (determined qualitatively).

Considering the tropospheric ozone column, I determine the correlation between the GEOS-Chem simulated profiles and the ozone sondes using least squares linear regression. This assumes that the ozonesondes are the true values, while the model output is linearly dependant on them with the addition of some random error. This correlation is determined between simulated and measured tropospheric ozone in molecules  $\text{cm}^{-2}$ , for each site, for each day where there exists both a measurement and model output. Figure 3.21 shows the correlation between mean modelled tropospheric ozone column. The colours show which season the datapoints lie within, and suggest that seasonality plays a large role in this correlation; since summer and winter points are clumped at opposite ends.

Removing the monthly mean from each dataset, and dividing by that mean gives the relative monthly difference of each datapoint from the respective dataset determined annual cycle. Figure 3.22 shows the least squares correlation between ozonesonde relative monthly anomalies and their GEOS-Chem counterparts. Again the season is coloured, and removal of the monthly mean removes the summer and winter clumping which could be seen in figure 3.21. The correlation coefficients are lower, since the model and the measurement both show the same seasonal trends this is expected. What this plot shows is the correlation between simulated tropospheric columns being above (or below) average as well as the ozonesonde measured columns being above (or below) average. The slopes are not one to one, which suggests the variance in sondes is not well matched by variance in simulated output. This effect

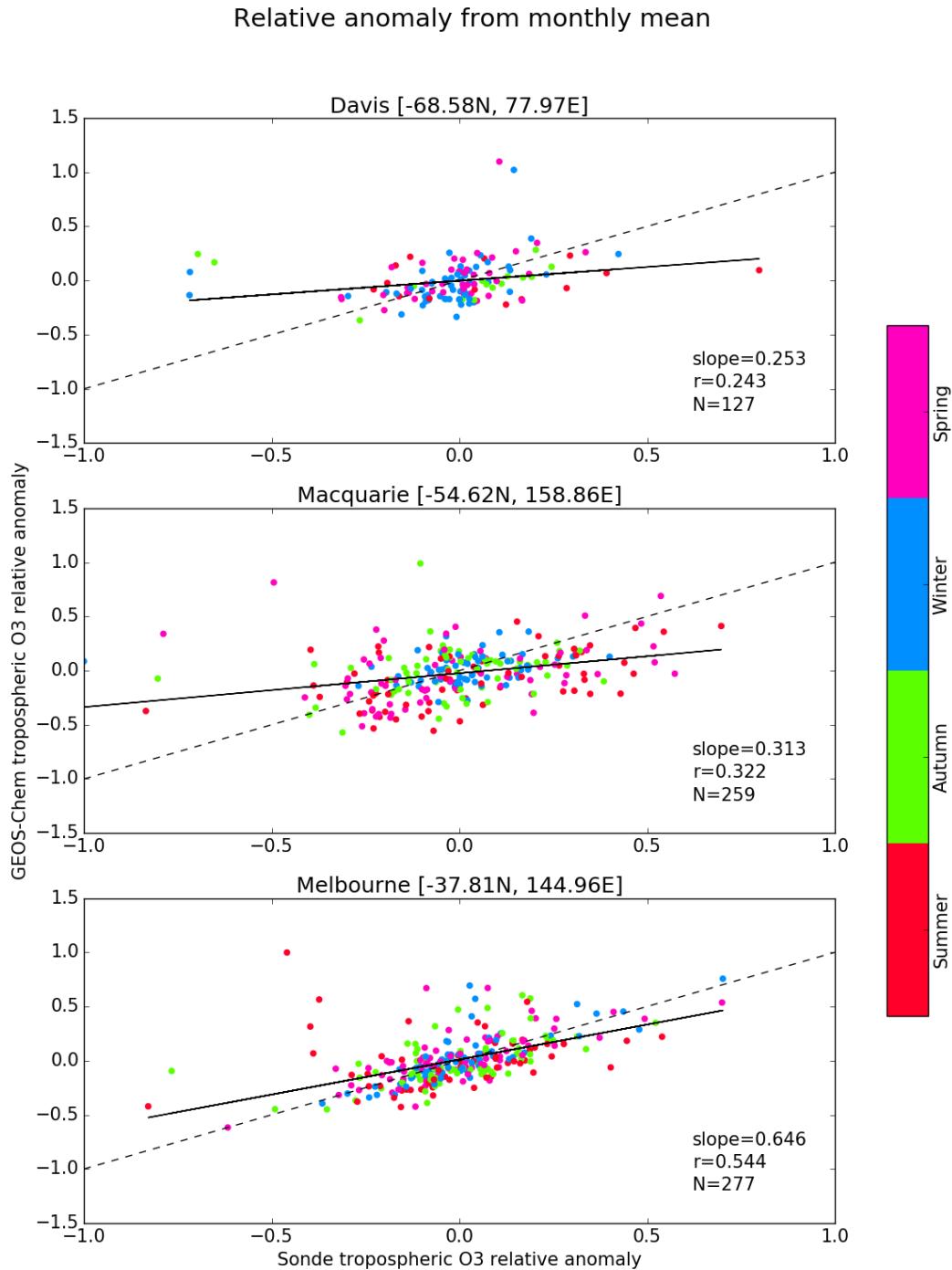


**Figure 3.20:** Observed and simulated tropospheric ozone profiles over Davis, Macquarie, and Melbourne, averaged seasonally. Model means (2005–2013 average) is shown as red solid lines, with red dashed lines showing the 10th and 90th percentile. Ozonesonde means (over each season, for all years) are shown as black solid lines, with coloured shaded areas showing the 10th and 90th percentile. The horizontal dotted line shows the mean tropopause heights from the model (red) and the observations (black).



**Figure 3.21:** Correlation between tropospheric ozone column modelled by GEOS-Chem (vertical axes) and calculated from ozone sondes (horizontal axes). Points are coloured based on their season, shown by the colour bar on the right. The line of best fit (black) and one to one line (dashed) is also shown, note the axes are different for Melbourne.

is strongest at the Davis and Macquarie Island stations, and is exacerbated when the seasonal cycle is removed.



**Figure 3.22:** Correlation between relative difference of tropospheric ozone column from the monthly mean between profiles modelled by GEOS-Chem (vertical axes) and profiles calculated from ozone sondes (horizontal axes). The line of best fit (black) and one to one line (dashed) is also shown.

(TODO: Update after model is rerun) Quantitatively, the modelled tropospheric

Site	direct		anomaly <sup>a</sup>		compared (N)
	r	r <sup>2</sup>	r	r <sup>2</sup>	
Davis	0.527	0.278	0.243		127
Macquarie Island	0.397	0.158	0.322		259
Melbourne	0.583	0.340	0.544		277

**Table 3.1:** Correlations between ozone and tropospheric ozone for monthly mean from GEOS- column.

<sup>a</sup>Relative and normalised ozone to sonde for respective data.

profile correlates reasonably well with the ozone sonde profiles, with correlation coefficients listed in table 3.1. Davis has a strong correlation ( $r = 0.527$ ) between sonde and simulated tropospheric column, which weakens greatly ( $r = 0.243$ ) when the season is removed. This suggests the model has the season and amplitude of tropospheric ozone fairly well simulated over larger time scales, but may lack accuracy when examining time periods of less than a few months. Macquarie and Melbourne are less affected by the removal of the seasonal cycle. Melbourne has a correlation coefficient greater than 0.5 with or without the seasonal cycle, showing that the model not only represents the larger time scales fairly well, but also has some accuracy catching days which are higher or lower than the monthly average.

TODO: show modelled daily cycle? Recall that the profiles are output every 6 hours, so as well as getting the overall profile average it is easy to determine the daytime and night time average by only looking at particular hours. GEOS-Chem uses GMT/UTC time, outputting 4 profiles per day at 0 hrs, 6 hrs, 12 hrs, and 18 hrs. The local offset in time at Davis, Macquarie Island, and Melbourne is 7 hrs, 11 hrs, 10 hrs respectively.

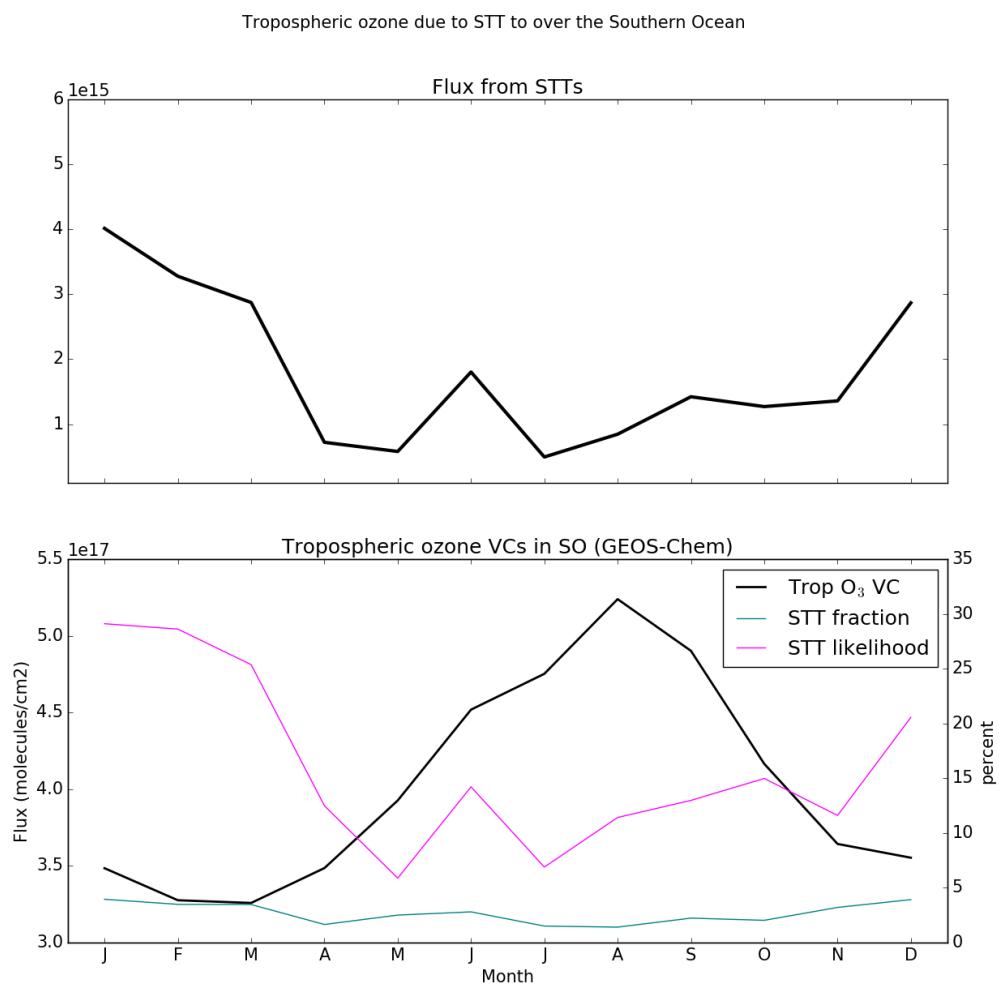
### 3.6.4 Estimation of southern ocean STT flux

A simplistic calculation of how much ozone in the southern high latitudes is due to STT is made through the following method. STT event likelihood, multiplied by STT event tropospheric column contribution, multiplied by southern high latitude tropospheric ozone columns, averaged seasonally. The monthly likelihood of STT events is determined from the ratio of detected STT events over sonde launches. The monthly fraction of tropospheric ozone column due to STT is calculated from the averaging the STT contributions binned by month. The southern high latitude tropospheric column is computed from the monthly averaged tropospheric ozone columns output by GEOS-Chem, binned into months between 2005 and 2013 in-

clusive. This modelled ozone column is the average between 55°S and 85°S for all longitudes. The equation is straightforward once the required components are all computed:

$$\text{Flux}_i = l_i \times f_i \times \text{tropO}3_i$$

where i subscripts index the month (Jan - Dec), Flux is our estimate of STT contribution to the tropospheric ozone column, l and f represent the likelihood and contributory fraction of STT events respectively. Figure 3.23 shows the seasonal STT contribution calculated this way.



**Figure 3.23:** Top panel shows the estimated STT contribution to tropospheric ozone VC. Bottom panel shows the three factors multiplied together in order to produce the estimation. Units for l and f are on the right, while units for ozone VC amounts are on the left.

# Chapter 4

## Formaldehyde product over Australia

### 4.1 Australian Biogenic Volatile Organic Compounds (BVOCs)

#### 4.1.1 Isoprene, Monoterpenes

Isoprene has a large impact on the oxidative properties of the atmosphere, as it reacts quickly with the OH radical to form RO<sub>2</sub> and then in the presence of NO<sub>X</sub> various OVOCs(largely HCHO) and other SOAs(notably ozone) are formed.

Estimates of isoprene emissions require more work in order to generate confidence at a global scale. Due to isoprene being optically thin (see section 4.2.3) and having a short life time (around an hour) there are relatively few accurate measurements against which a comparison and verification can be made.

The emissions of isoprene have been modelled at around 500 Tg C yr<sup>-1</sup> in [Guenther1995]. The global emission factors model used to derive this estimate is based on modelling emissions from different plant species (phenotypes), and very few are used to set the emission factors of Australian forests.

Globally around 710 - 1150 Tg C yr<sup>-1</sup> of BVOCs are emitted ([Guenther1995, Lathiere2006, Guenther2012]). 90% of these emissions come from plants and trees, with the most dominant species being isoprene (C<sub>5</sub>H<sub>8</sub>) (~ 50%), monoterpenes (C<sub>10</sub>H<sub>16</sub>), methanol (CH<sub>3</sub>OH), ethanol (C<sub>2</sub>H<sub>6</sub>O), acetaldehyde (CH<sub>3</sub>CHO), acetone ((CH<sub>3</sub>)<sub>2</sub>CO), ethene (C<sub>2</sub>H<sub>4</sub>) and propene (C<sub>3</sub>H<sub>6</sub>) (together making up ~ 30%) ([Guenther2012]). The larger 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. Another model (ORCHIDEE, with inputs similar to MEGAN) estimates 752 ± 16 Tg C yr<sup>-1</sup>, sensitive to terrestrial vegetation

variations ([**Lathiere2006**]).

One problem with current estimates of biogenic VOC emissions in Australia is that the emission rates from various species of eucalypt 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. Isoprenoid emissions remain to be verified in Australia and the few monoterpenene emission rates we have may be underestimated by a factor of 2-4 ([**Winters2009**]).

#### 4.1.2 Biomass Burning

One complication when computing HCHO yield from VOC emissions is biomass burning interference, as smoke plumes can contribute to column HCHO. Influence from biomass burning can be removed through measurements of acetonitrile and CO (eg: ([**Wolfe2016**]), TODO: other bb papers). [**Wolfe2016**] disregard HCHO measurements when acetonitrile  $\geq$  210 pptv and CO  $\geq$  300 ppbv.

#### 4.1.3 MEGAN

One method used to estimate global isoprene (among other species) emissions is the Model of Emissions of Gases and Aerosols from Nature (MEGAN). MEGAN is a global model with resolution of around 1 km, and is used to generate the BVOC emissions used in various global chemistry models such as GEOS-Chem. MEGAN uses leaf area index, global meteorological data, and plant functional types (PFTs) to simulate terrestrial isoprene emissions. The various PFTs are used to generate emission factors which represent quantities of a compound released to the atmosphere through an associated activity. For example, an emission factor for isoprene within a forest would include the requirement of sunshine and suitable temperature. The schematic for MEGAN, taken from [**Megan Website**], is shown in figure 4.1

MEGAN output in Australia is adversely affected by poor emission factor estimation, which is based on plant type classifications (PFTs) and local conditions like soil moisture and weather. An example can be seen in [**Muller2008**] where MEGAN overestimates isoprene in northern Australia. Underestimates of monoterpenes emissions are also seen from MEGAN (TODO: ask Jenny what other paper showed this?), which may be due simply to underestimated emission rates for many Eucalypt species ([**Winters2009**]).

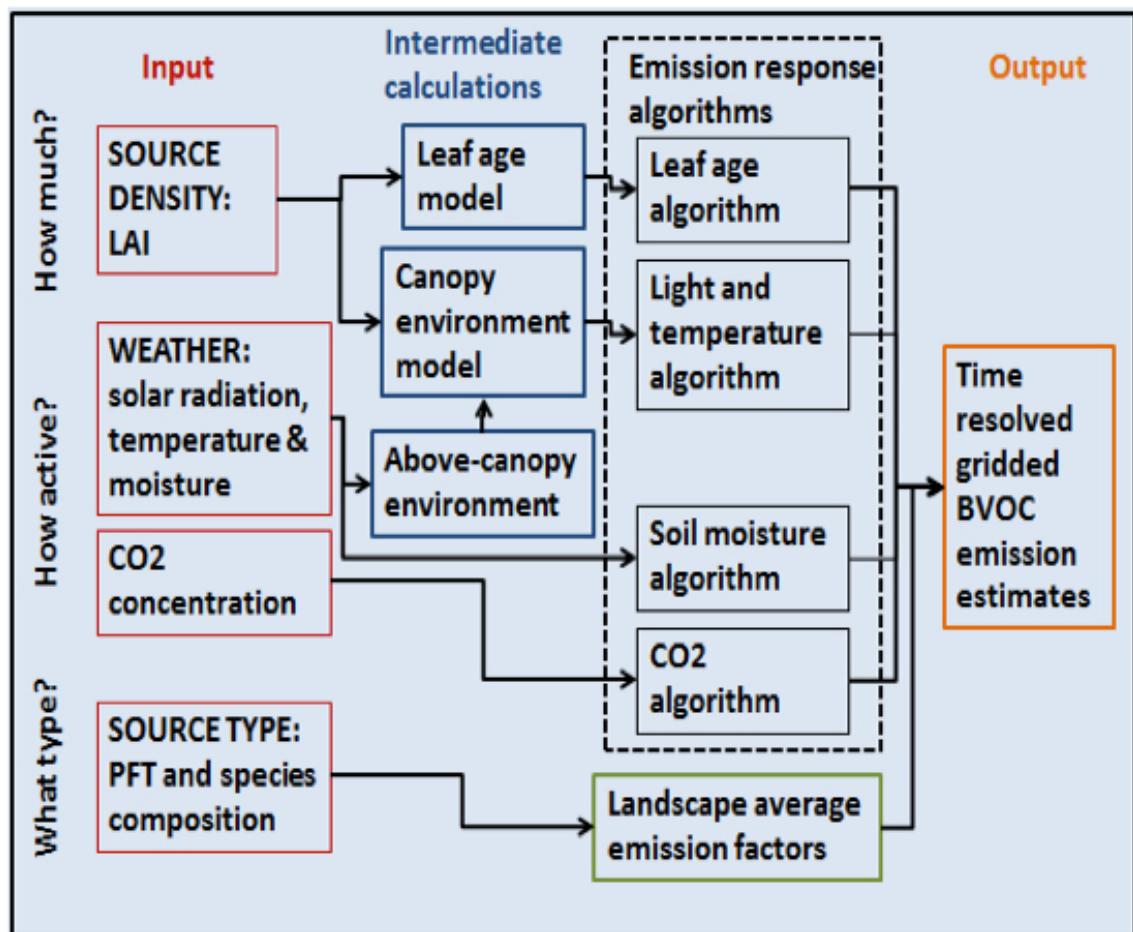


Figure 4.1: MEGAN schematic, copied from [Megan'Website]

## 4.2 Satellite HCHO measurements

### 4.2.1 Satellite Retrievals

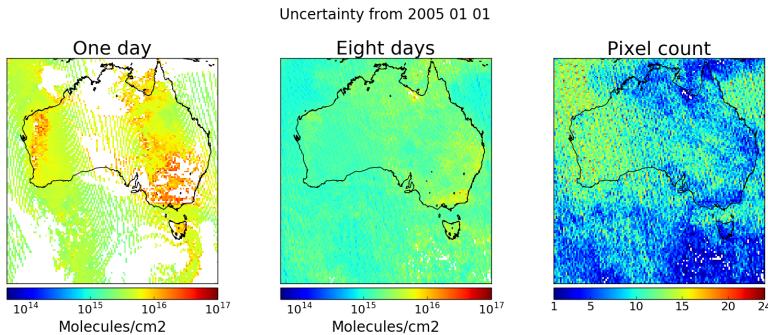
Satellites remotely sense atmospheric trace gases through irradiance measurements of solar light which has reflected off the earth's surface. These irradiances are affected by gases which exist along the reflected path of light between the detector, earth, and sun. The irradiance is then used to estimate 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).

One satellite is NASA's Earth Observing System's (EOS) Aura, which houses the Ozone Monitoring Instrument (OMI), a near-UV/Visible Charged Coupled Device (CCD) spectrometer. Aura orbits the earth in a polar sun-synchronous pattern, circling the earth on a plane coincident with the sun and the poles. OMI measurements are used to map several atmospheric trace gases, including NO<sub>2</sub>, SO<sub>2</sub>, BrO, HCHO, O<sub>3</sub>, 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 13x24 km<sup>2</sup> at nadir to 26x135 km<sup>2</sup> at the swath edges ([Abad2015]). The swaths cover Earth daily, although half of these are at night time and contain no useful near-UV/Visible information. From here on the word pixel is used to describe one data point retrieved by OMI, which will be located by latitude and longitude within OMI's data product.

Atmospheric HCHO can be measured using Differential Optical Absorption Spectroscopy (DOAS), as long as trace gases with similar features near the same wavelength are accounted for. TODO: Go through Lee2005 or Volkamer2005 and detail the DOAS Retrieval of HCHO. 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. The full method details for slant column retrieval by OMI are outlined in section 4.2.2.

Uncertainty in a single pixel for OMI is quite high (todo: link to picture with one day average plus omi uncertainty). Figure 4.2 shows uncertainty over Australia after one and eight days of averaging at 0.25° longitude by 0.3125° latitude. There are several methods of calculating this, one of which is used here and compared against the provided uncertainty (TODO) as shown in Section 4.2.9. 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. Random error can be reduced by either temporal or spatial averaging, decreasing uncertainty by a factor

of  $1/\sqrt{N}$  where N is the number of observations being averaged.



**Figure 4.2:** OMI uncertainty before and after gridding and averaging 8 days from Jan 1 2005 to Jan 8 2005. The third panel shows the number of pixels in each grid box after 8 days of averaging, before accounting for fire.

#### 4.2.2 OMI Algorithm BOAS

The following information comes from the OMHCHO dataset documentation at [Kurosu2014] and [Chance2002]. The method of HCHO total column retrieval depends heavily on measured solar radiation. Radiance is directional radiant flux, expressed in Watts per square metre per steradian (a unit of angle used in three dimensional geometry). Irradiance is radiant flux received by a surface, expressed in watts per square metre. An OMI granule is the sunlit portion of an orbit (one day).

The BOAS algorithm used by OMI is as follows. Slant column abundance can be determined by fitting measured radiance ( $I$ ) at particular wavelengths ( $\lambda$ ), using modelled absorption cross sections ( $\sigma$ ), effective albedo (A) including Rayleigh scattering, a correction for the Ring effect ( $c_R\sigma_R$ ), and a closing polynomial (coefficients  $c_0-c_3$ ).

$$I(\lambda) = AI_0 \exp(-\sum_i S_i \sigma_i) + c_R \sigma_R + c_0 + c_1(\sigma - \bar{\sigma}) + c_2(\sigma - \bar{\sigma})^2 + c_3(\sigma - \bar{\sigma})^3 \quad (4.1)$$

For HCHO, absorption cross sections and number densities for interfering gases are determined beforehand. This is due to HCHO being so optically thin and interferences must be accounted for precisely ([Chance2002]).

In version 3.0 of the OMI satellite data retrievals, HCHO is determined using the spectral window 328.5 nm–356.5 nm. The algorithm used is based on direct fitting of radiances and irradiances. An OMI radiance measurement over the remote Pacific ocean is used instead of an irradiance measurement. This means that the slant columns ( $\Omega_S$ ) are actually the difference with respect to the radiance reference column ( $\Omega_{S_0}$ ).

The model that is fitted to the measurements is made up of the radiance reference attenuated by HCHO contributions, inelastic (rotational Raman) scattering, and interferences from ozone, NO<sub>2</sub>, BrO, and the O<sub>2</sub>-O<sub>2</sub> collision complex. It includes additive and multiplicative closure polynomials and parameters for spectral shift and squeeze, and an undersampling correction and “common mode” spectrum. The spectral fitting results in HCHO slant columns, which are converted to vertical columns through a look-up table of AMFs (see section 4.2.6) Undersampling is a problem caused by the wavelength resolution of the instrument. Nyquist theorem requires that the sampling rate be at least twice the highest frequency of the signal in order to uniquely reconstruct it, otherwise the signal is undersampled (contains errors).

There are three main stages in the algorithm:

1. Radiance wavelength calibration, finding the optimum wavelength registration for a representative swath of radiance measurements, and determination of a common wavelength grid for auxiliary data (molecular reference cross sections, etc.).
2. On-line common mode spectrum calculation from residual fits of the central portion of the orbit. This accounts for systematic features not considered in the semi-empirical model.
3. Nonlinear least-squares fitting of all swath lines in the OMI granule. Fitting is performed individually for the 60 cross-track pixels in each swath line.

Cross-track striping is systematically higher or lower column values along a whole track. Several methods are used to reduce cross-track striping of the HCHO columns. These include soft calibration, which is the use of a daily radiance reference, and outlier screening in the fitting residuals.

### 4.2.3 Optical Depth ( $\tau$ )

Optical Depth, also called optical thickness, is the natural logarithm of the ratio of incident radiant power to transmitted radiant power through a material. 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  $\phi_e^i$  is radiant flux seen at the earth surface,  $\phi_e^t$  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.

#### 4.2.4 Scattering

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 light. Mie scattering is more dominant from larger particles, and has less wavelength dependence. These scattering functions are described in detail at (TODO:section? reference?).

#### 4.2.5 Absorption cross section and number density

TODO: Fill in this section, describe cross sections.

$\tau$  can be described using the attenuation cross section (the attenuation coefficient divided by its number density), with the following relation:

$$\tau = \int_0^l \alpha(z) \eta(z) dz$$

where  $\alpha(z)$  and  $\eta(z)$  represent absorption cross section in  $\text{m}^2 \text{ molecule}^{-1}$ , and number density in molecules  $\text{m}^{-3}$  respectively, and  $l$  represents the length of the path the light is travelling through.

#### 4.2.6 Air Mass Factors

DOAS retrieval columns are an integration of a trace gas over the instrument's 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 all altitudes using an RTM. These properties are combined into a single array called the AMF which can be combined with a SC to produce a VC. AMFs are unique to each trace gas and due to their complexity and the influence of cloud cover they remain one of the larger error sources in remote sensing of BVOCs ([Palmer2001, Millet2006]).

The latest OMI algorithm uses a shape factor determined from GEOS-Chem using 47 vertical levels at monthly temporal resolution and  $2^\circ$  latitude by  $2.5^\circ$  longitude horizontal resolution ([Abad2015]). The GEOS-Chem model has been substantially updated since then, and using the more recent version V10.01 to recalculate

the AMF is performed within this thesis, details are shown in section 4.2.8.

The vertical distribution of a trace gas determined by CTM is independent of the vertically dependent observation sensitivity provided by RTM, which prevents model contamination of remotely sensed data. Two examples of this are GOME-2 products on the MetOp-A satellite and OMI products which use IMAGESv2 combined with LIDORT and GEOS-Chem with LIDORT for product processing respectively ( TODO: ([Abad2015]). The actual formulation of an AMF is explained in section 4.2.8).

#### 4.2.7 OMI HCHO data products

OMI 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 using an AMF calculated through radiative transfer models. Level 3 data is a temporally aggregated version of the this, for instance monthly or yearly averages.

In order to filter fire or anthropogenic influence on HCHO data one generally requires the level 2 data along with extra data such as MODIS and AATSR fire counts, and national or regional industrial emissions estimates.

#### 4.2.8 HCHO Vertical Column Calculation

The method used here largely follows that of [Palmer2001]. The Beer-Lambert law of extinction allows spectroscopic measurement of absorbing chemical species (absorbers) in the atmosphere:

$$I_B = I_{B_0} e^{-\tau_s} \quad (4.2)$$

where  $I_B$ ,  $I_{B_0}$  is backscattered intensity with and without the absorber respectively, and  $\tau_s$  is the optical thickness of the slant column absorber (see section 4.2.3).

The AMF is needed to transform the slant column, as viewed by the satellite, into a vertical column:

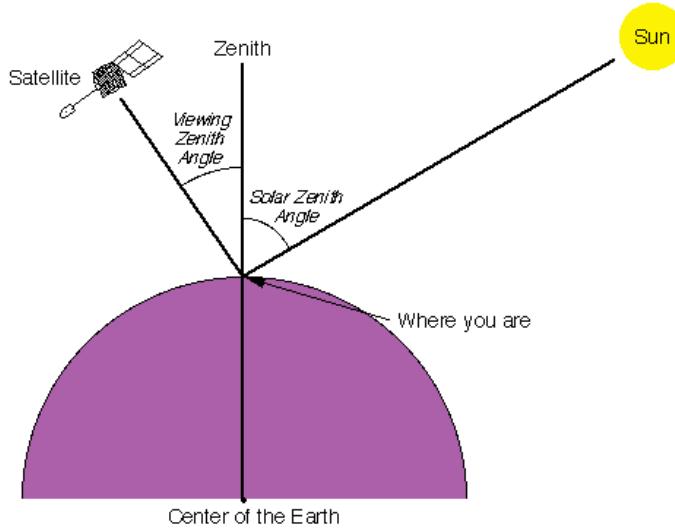
$$AMF = \frac{\Omega_s}{\Omega_v} = \frac{\tau_s}{\tau_v} \quad (4.3)$$

where s and v subscripts refer to slant and vertical values, while  $\Omega$  represents a column of absorber in molecules  $\text{cm}^{-2}$ .

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 ( $\theta_s$ ) and the satellite

viewing angle ( $\theta_v$ ) are shown in image 4.3. However, in the UV-VIS region of the spectrum, Rayleigh and Mie scattering (see section 4.2.4) must be accounted for. Using equations 4.2 and 4.3 gives us an expression for the AMF which includes scattering:

$$AMF = \frac{\ln I_{B_0} - \ln I_B}{\tau_v} \quad (4.4)$$



**Figure 4.3:** Solar and viewing zenith angles, image copied from [SZA\*Image], originally from a NASA website.

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

$$AMF = \frac{\ln \left( 1 - \frac{\nabla I}{I_B} \right)}{\tau_v} \approx \frac{-\frac{\nabla I}{I_B}}{\tau_v} \quad (4.5)$$

$\nabla I$  can also be expressed as the integral of the absorption slices over optical depth increments:

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

which can be placed into equation 4.5:

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

and rewritten as:

$$AMF = \frac{-1}{\tau_v} \int_0^{\infty} \frac{\partial \ln I_B}{\partial \tau} \alpha(z) \eta(z) dz \quad (4.6)$$

where  $\alpha(z)$  and  $\eta(z)$  represent absorption cross section in  $\text{m}^2 \text{ molecule}^{-1}$ , and number density in molecules  $\text{m}^{-3}$  respectively. This uses the attenuation cross section relationship to optical depth (see section 4.2.5).

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

$$\alpha_e = \frac{1}{\Omega_v} \int_0^\infty \alpha(z) \eta(z) dz$$

This is the same as  $\tau_v = \Omega_v \alpha_e$ , which we can place into equation 4.6 to obtain:

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

Defining  $w(z)$  as the scattering weights describing the sensitivity of the backscattered spectrum to the abundance of an absorber at altitude  $z$ :

$$\omega(z) = - \frac{1}{AMF_G} \frac{\alpha(z)}{\alpha_e} \frac{\partial \ln I_B}{\partial \tau} \quad (4.7)$$

and vertical shape factor  $S_z(z)$  representing a normalized vertical number density profile:

$$S_z(z) = \frac{\eta(z)}{\Omega_v} \quad (4.8)$$

Now the AMF can be expressed as

$$AMF = AMF_G \int_0^\infty w(z) S_z(z) dz \quad (4.9)$$

It's worth noting that in the OMI satellite product, the provided  $\omega(z)$  term does not include the  $\frac{1}{AMF_G}$  term and a the calculation in equation 4.9 does not use the  $AMF_G$  term. This is not noted in any of the papers which recalculate the AMF from the OMI product, due to them recalculating the  $\omega$  term themselves with a radiative transfer model such as LIDORT.

For this equation  $\omega$  is a function of atmospheric scattering which can be calculated using an RTM to determine the absorption cross section and optical thickness along the path.  $S_z(z)$  can be calculated using an apriori vertical profile, which may be sourced from any atmospheric chemistry model. Note that for level two non-gridded OMI satellite HCHO total column data, the  $w(z)$ ,  $S_z(z)$ , and  $AMF_G$  are all provided.

Mie scattering and clouds can complicate the calculation of  $\omega(z)$ , however tables of values for this function at various parameter inputs can be used with modeled vertical shape factors for local AMF calculations.

Using the vertical coordinate sigma ( $\sigma$ ), which is related to pressure ( $P$ ) by  $P = \sigma (P_S - P_T) + P_T$ , where subscripts S and T represent earth surface and top of the atmosphere respectively. The hydrostatic relation  $P = -\rho_a g z$ , with  $\rho_a$ ,  $g$ , being

density of air, gravity, respectively lets us switch to the sigma coordinate using:

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

Substitution into 4.9 gives AMF using the sigma coordinates:

$$AMF = AMF_G \int_0^1 w(\sigma) S_\sigma(\sigma) d\sigma \quad (4.10)$$

Where  $S_\sigma$  is defined as a vertical shape factor representing a normalised mixing ratio:

$$S_\sigma(\sigma) = \frac{\Omega_a}{\Omega_v} C(\sigma) \quad (4.11)$$

where  $\Omega_a$  is the vertical column of air and  $C(\sigma)$  is the mixing ratio of the absorber. This normalised shape factor is dimensionless. 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.

When comparing satellite observations to a chemical model, one needs to re-calculate the satellite AMF using their own modelled vertical gas profiles as the a-prior shape factor in order to remove any total column bias which may be due to the satellite's apriori. Another way of removing this bias is through deconvolution (TODO: EQNS) of the averaging kernal (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  $w(z)$  function as follows:

$$AK(z) = \frac{\omega(z)}{AMF} \quad (4.12)$$

Note that this is an approximation for the OMI product, which does not include the AK but does include the  $\omega$  and AMF, as explained in [Abad2015].

#### 4.2.9 Uncertainty in OMI total columns

Provided with the OMI product is the measurement of uncertainty in each pixel, calculated by SAO from the backscattered solar radiation fit ([Abad2015, Abad2016]). BIRA use another method, and calculate the standard deviation of HCHO over the remote pacific ocean as the uncertainty ([DeSmedt2012, DeSmedt2015]). In the remote pacific, it can be assumed that HCHO variations are weak. The standard deviation over this region is therefor a proxy for determination of the instrument error.

Taking the eight day grided average with horizontal resolution of 0.25 by 0.3125

degrees (latitude by longitude) typically reduces uncertainty by a factor of 1.5 to 4. An analysis of the uncertainty within our recalculation of the OMI HCHO columns is performed in section 4.3.9.

#### 4.2.10 Reference sector correction for comparison of products to various models

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**]). Since this oceanic background is used instead of a solar irradiance spectrum, in order to compare the output against different models, the vertical columns need to be corrected by an absolute amount. The corrected vertical column ( $\Omega_V$ ) is calculated as the slant column ( $\Omega_S$ ) minus the reference slant column ( $\Omega_{S_0}$ ) multiplied by the AMF, plus the modelled reference sector column ( $\Omega_{V_B}$ ):

$$\Omega_V = \frac{(\Omega_S - \Omega_{S_0})}{AMF} + \Omega_{V_B}$$

This method is used in various papers, including [**DeSmedt2008, DeSmedt2012, DeSmedt2015, Barkley2013, Bauwens2016**].

Recently this correction was expanded (for OMI data) to include latitudinal and instrument track influence by [**Abad2015**]. The updated correction is explained in detail in section 4.3.8.

### 4.3 Recalculating HCHO from satellite(OMI) data over Australia

#### 4.3.1 Process Outline

First satellite slant columns of formaldehyde for the years 2004 - 2013 will be retrieved from NASA, BIRA, etc (TODO: list satellites + sources or reference section). This data needs to be quality assured and will undergo a basic analysis and filtering criteria as is done in several other studies, for example [**Marais2012, Barkley2013, Bauwens2016, Zhu2016**] TODO: more cites. This filtering generally removes cloudy and uncertain data points, along with instrument anomalies such as the row anomaly experienced by OMI ( todo: cite row anomaly analysis/details ). A full account of the filters used when reading OMI satellite data is given in section 4.3.2.

In order to reduce uncertainty and increase the utility of the satellite data we

regrid it from pointwise single time data points to 8-day averages on a latitude longitude grid which matches our model input and output grid spacing. Using the 8-day average reduces the uncertainty in each datapoint significantly, TODO: citet marais and barkley uncertainty improvements, details of the uncertainty estimations is shown in section 4.3.9.

Once the slant columns are quality filtered and gridded, additional data sources need to be used to account for anthropogenic and pyrogenic sources of HCHO. MODIS fire counts are used in conjunction with NO<sub>2</sub> enhancements (also measured by satellite) to remove data points which may be affected by fires. TODO:If it is easier to use OMI smoke smoke aaod I'll do it that way instead of using the NO<sub>2</sub>, write here if that is the case. One possible solution to anthropogenic filtering is the national pollution index (TODO:cite:<http://www.npi.gov.au>) which contains industrial HCHO and NO<sub>X</sub> emissions from 2003 to 2014. This has a negative affect on uncertainty, as fewer measurements are averaged over the 8-days. The affect of the fire filtering on uncertainty, and how many points are removed is shown in section 4.3.5.

Each satellite slant column measurement is corrected by some amount, based on the divergence from a modeled reference sector. The reference sector correction method corrects for several problems, however it introduces some apriori model influence. One of the problems removed through this correction method is instrument degradation, which can introduce bias over time. Another is the possible influence of varying dead/hot pixel masks across 2-D detector arrays such as OMI ([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]).

The reference sector we use is defined over the pacific ocean at 140 to 160°W and 90°N to 90°S, as in [Abad2015]. HCHO concentrations are assumed to be at background levels over the pacific ocean, with their only source being CH<sub>4</sub> oxidation. A correction for each instrument pixel is created based on the difference between the background HCHO measurements from OMI and the GEOS-Chem modelled HCHO columns within the reference sector. This correction is calculated daily and applied to all good pixels based on their latitude. The full process for this is shown in section 4.3.8.

In order to visualise and analyse satellite column data it is generally transformed into vertical columns. This is done using AMF calculations as shown in section 4.2.8. Taking the biogenic slant columns, scattering weights, and apriori estimates of HCHO vertical profiles we determine vertical HCHO column amounts. This is an in depth process involving radiative transfer modelling in order to work out satellite sensitivities at various altitudes, as well as the effect from the local HCHO profile

on those sensitivities. Several of these required data are available from the satellite data products, including the scattering weights and the zenith angles required to determine an AMF at any particular measured point. In this work the shape factor is recalculated from GOES-Chem, with the associated OMI per-pixel scattering weights unchanged. The satellite shape factor is replaced by GEOS-Chem's overpass time simulated HCHO profile, normalised and saved daily along with air density.

When comparing satellite measurements against models it is important to recognise the impact of the apriori shape factor on the total column values. This is due to the sensitivity of instruments varying vertically through the atmosphere, and how the simulated distribution of HCHO is accounted for. In order to remove a possible bias caused by systematic differences between the old model and the current model, the shape factor used by the satellite is replaced using the profile from the current model before satellite total columns are recalculated (generally using equation 4.9). Without performing this step a bias between modeled and measured total column values may be due to an apriori rather than actual chemistry or measurements.

### 4.3.2 Quality filtering OMI HCHO slant columns

TODO: Quality flags and cloud cover metric uses, and discussion, along with statistics like how many datapoints are removed.

The OMI 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 4.1 which is taken from [Kurosu2014]. Filtering bad or missing measurement pixels is preformed prior to any other filtering. Any pixel with a cloud fraction of greater than 40% is also removed, after the pixel is used in determining the reference sector correction, as is done in [Abad2015, DeSmedt2015]. In [Martin2003] (TODO: find this paper), pixels with more than half of the backscattered intesity from the cloudy sky fraction are removed, in effect removing pixels with more than 40% cloud fraction before calculating the AMF for NO<sub>2</sub> slant columns.

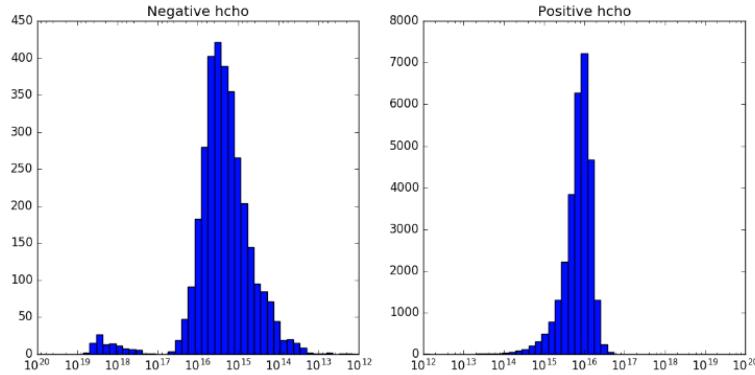
The cloud fraction with each pixel is provided with the OMHCHO dataset, however its source is the OMI cloud product, OMCLDO2. To give an idea of how much data is filtered out, around 30% of the pixels which remain after filtering out the bad or missing data are subsequently removed due to cloudiness.

Further filtering is performed to remove the measurements which are most likely to be unrealistic: pixels wit a solar zenith angle greater than 60°, and those with column density outside the range  $-0.5 \times 10^{16}$  to  $10^{17}$  molecules cm<sup>-2</sup>. These are similar filters to those applied in (TODO: read [Zhu2016], add similar justification if succinct). This final filter is required due to currently unexplained large negative values which occur in the OMI HCHO product increasingly over time. Figure

**Table 4.1:** OMI quality flag values table from [Kurosu2014]

<b>Value</b>	<b>Classification</b>	<b>Rational</b>
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"> <li>• Fit convergence flag is <math>&lt; 300</math> but <math>&gt; 0</math>: Convergence at noise level</li> <li>• Column <math>+2\sigma</math> uncertainty <math>&lt; 0 &lt;</math> Column <math>+3\sigma</math> uncertainty</li> <li>• Absolute column value <math>&gt;</math> Maximum column amount (<math>1\text{e}19</math> molec <math>\text{cm}^{-2}</math>)</li> </ul>
2	Bad	Avoid using as one of the following conditions are present: <ul style="list-style-type: none"> <li>• Fit convergence flag is <math>&lt; 0</math> : No convergence, abnormal termination</li> <li>• Column <math>+3\sigma</math> uncertainty <math>&lt; 0</math></li> </ul>
$< 0$	Missing	No column values have been computed; entries are missing

4.4 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 histograms here show the negative (left) and positive (right) total column HCHO measurements from a subset of swaths over Australia, on the 18th of March 2013. The highly negative values can be seen around the  $-10^{19}$  molecules  $\text{cm}^{-2}$  region.



**Figure 4.4:** 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.

One final selection of the data is removed prior to calculation and analysis due to instrument sensitivity issues. This removed data is based on the latitude of the satellite measurement, high northern and southern latitudes are removed as the effect of both the high solar zenith angle and (over arctic regions) the high albedo cause anomalous readings which are hard to correct. The latitudes removed from analysis for OMI satellite data are those outside of  $0^\circ \pm 60^\circ$ . This is where the satellite measurements are less robust, and often publications will remove this same region (TODO: cites where they did this).

### 4.3.3 Reading OMHCHO daily slant columns

Each  $\sim 90$  minutes the AURA satellite sweeps over the sunny side of the planet, with OMI recording roughly 90 k pixels, of which around 50 k – 80 k 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 (molecules  $\text{cm}^{-2}$ ), 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 apriori shape factor rather than NASA's included apriori shape factor. Each pixel and it's relevant data are saved in a

long list, around 1.1 million pixels per day. As well as the data directly read from the OMI swath files, further information is added to each pixel. This is the new AMF calculated through replacing the apriori vertical profile with the newer GEOS-Chem simulated profile, which is described in section 4.3.7. The shape factors and scattering weights for each pixel lie along a z-axis which is vertically resolved to 47 layers and is shown in figure TODO: make figure showing this stuff and an example profile.

TODO: Show an example of OMI swaths.

#### 4.3.4 Regridding to 0.25 by 0.3125 8-day averaged vertical columns

Regridded OMI HCHO columns from the are based on 14-15 daily swaths of measurements provided by NASA. Each swath contains roughly  $9 \times 10^4$  pixels, each of which contains various data including latitude, longitude vertical column HCHO, etc. In order to regrid these columns each pixel is mapped into a global grid of  $0.25^\circ \times 0.3125^\circ$  latitude by longitude (matching GEOS-5 native resolution) which may contain up to 15 entries from a particular day's orbits. Total vertical column amounts (both the satellite original and the columns reprocessed as follows) are averaged over each 8 days starting on January 1st 2005.

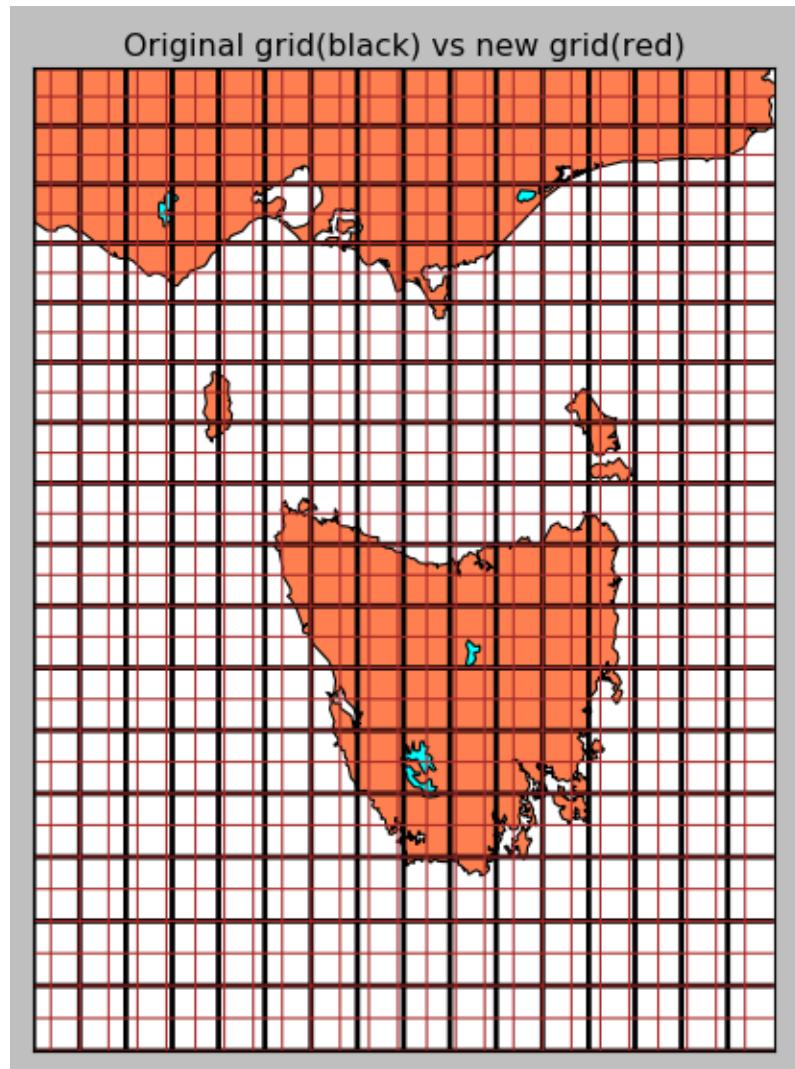
The process of regridding is performed by first reading all relevant information at each pixel for a single day into a list in python, which is then processed, with the shape factor read from GEOS-Chem output and AMF recalculated, before being saved as a gridded array of total columns and pixel counts. The total columns are averaged into each grid box for each day, and eventually averaged over eight day time periods.

TODO: time per regridding and reprocessing: This mapping requires some processing time as well as RAM and computer storage space, and has been performed on the National Computing Infrastructure (NCI) supercomputer cluster. In order to reprocess one year of swath files, X GB of daily data was downloaded and then transformed into Y GB of 8-day averaged gridded data. This takes around 90 minutes per day, and is very parallelisable as each day is completely independent. Using  $N \times 8$  concurrent processors on NCI's computer cluster running Python allows for very fast reprocessing of our entire timeline. As much as possible, processing is done using the HDF-EOS5 format, with GEOS-Chem output being read and processed from bitpunch to HDF-EOS5 prior to reprocessing. The scripts to regrid and reprocess the swath data set are available in the supplementary (TODO).

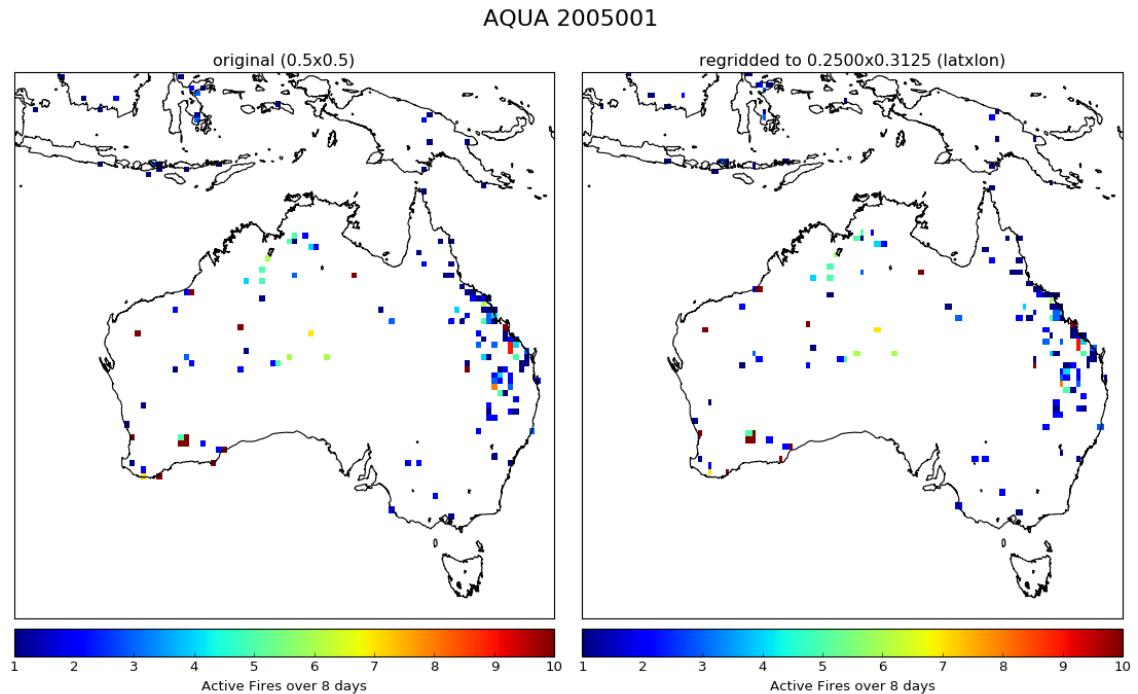
### 4.3.5 Filtering pyrogenic HCHO

TODO: How modis fire counts are used as well as statistics on removed data points.

On board NASA's AQUA satellite, the MODIS instrument is used to detect fire activity. The product used here is called MYD14C8H ([Giglio2006]), which looks at fire activity over eight days on a  $0.5^\circ$  square grid globally. Regridding the product to the native meteorological grid of GEOS5 at  $0.25^\circ$  latitude by  $0.3125^\circ$  longitude is done in python with an interpolator which maps the values of the new grid rectangles to the value of the nearest grid square. An example of the change in resolution is provided in figure 4.5, where the grids are shown over a basic map of Tasmania. The direct affect of this interpolation is shown as an example in figure 4.6, which is showing the regridded MODIS fire count over Australia from January 2005 (avg of first 8 days) in two subplots.



**Figure 4.5:** Example of grid space change using  $0.5 \times 0.5$  and  $0.25 \times 0.3125$  latitude by longitude resolution.



**Figure 4.6:** Example of MODIS 8 day grid interpolation from  $0.5 \times 0.5$  to  $0.25 \times 0.3125$  latitude by longitude resolution. This example uses MODIS fire counts for 1-8 January 2005.

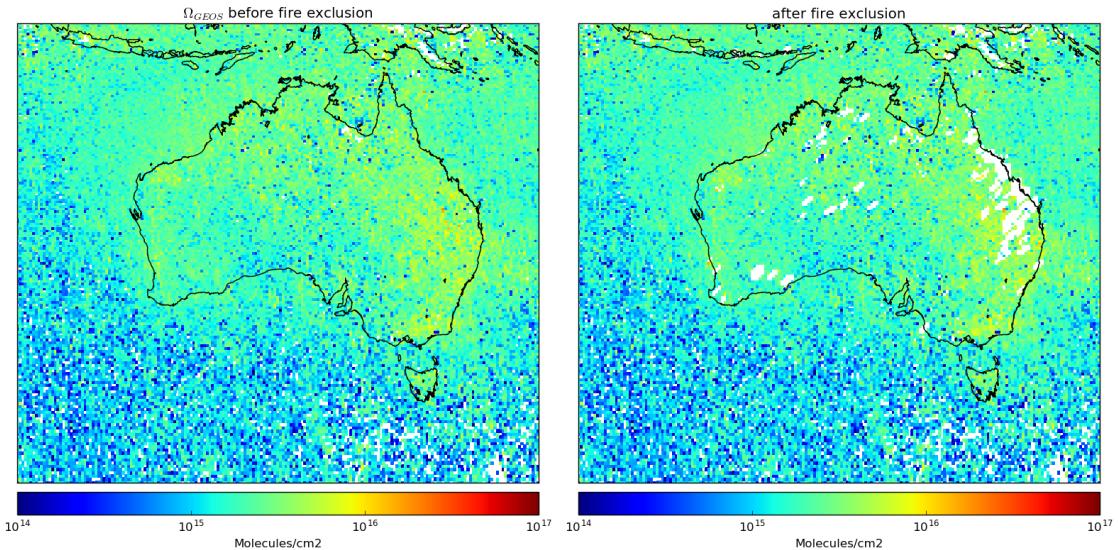
Figure 4.7 shows an example of the total column HCHO calculated using GEOS-Chem aprioris ( $\Omega_{GEOS}$ ) before and after using the MYD14C8H product to exclude fire influenced pixels. (TODO: show time series of how many pixels are removed and discuss if this causes any issues down the line)

This filtering ends up removing too much information, and the recalculation of HCHO is too negatively influenced. To deal with this a separate product from the same instrument has been downloaded: MYD14A1, which keeps daily fire counts. Less disruptive filtering can be achieved by removing pixels which coincide with fires on the same day, as shown in figure TODO: which compares the 8 and 1 day filtering. TODO: The script to read and regrid these one day fire counts was adapted from X. Figure (TODO: effect on uncertainty and time series of fire pixels removed) shows the daily filtering effect on uncertainty and time series of fire pixels removed.

### 4.3.6 Filtering anthropogenic HCHO

TODO

### Eight day average $\Omega_{GEOS}$ for Jan 1, 2005



**Figure 4.7:** Vertical column HCHO calculated using OMI satellite swaths with GEOS-Chem aprioris, averaged over 1-8 January 2005 with and without fire affected squares removed.

#### 4.3.7 Recalculating the AMF to create our own vertical HCHO columns

OMI's apriori shape factor is based on the GEOS-Chem (v9) model, which uses 47 layers between the earth's surface and the top of the atmosphere using a pressure-eta hybrid (the actual values are shown in table A.1). Taking a more recent GEOS-Chem apriori shape factor and integrating along the vertical axis using equation 4.9 gives us a new AMF ( $AMF_n$ ). Since we are using the  $\omega$  provided by OMI, we remove the  $AMF_G$  term from this calculation. The integration is done in Python using a simple rectangular method, which multiplies the integrand midpoints by the change in height, and then takes the sum. This is identical to calculating the integral if we assume the integrand is linear between each measured point, and introduces no new uncertainty. All that remains for recalculating the total vertical column using our new apriori shape factor is to apply the new AMF and remove the old:

$$\Omega_{new} = \Omega \frac{AMF}{AMF_n}$$

The vertical column scattering weights and apriori shape factors provided in the OMHCHO dataset are defined on 47 levels. In order to reformulate the vertical column using updated GEOS-Chem hcho apriori shape factors I have run GEOS-Chem

version 10.01 on the full 72 level vertical grid at 2 by 2.5 (lat by lon) degree monthly resolution. The simulated vertical profiles of HCHO are averaged from 1300-1400 local time in order to match the satellite overpass time of roughly 1330. These vertical profiles then provide the apriori shape factor for the higher horizontally resolved satellite columns, which pick the nearest apriori from the model. TODO: determine which of these is correct! a)The new apriori profiles are monthly averages, which is the same temporal resolution used by the OMI apriori shape factors. b)The new apriori profiles are simulated daily and averaged over 8 days along with the recalculated total vertical columns.

A new AMF is determined using equation 4.9) with the apriori shape factor set by our GEOS-Chem model run. In order to reformulate the AMF, GEOS-Chem's 72 level vertical profile is transformed from ppb to a normalized number density profile in order to match equation 4.8. This conversion uses the following equation:

$$\eta_{HCHO} = ppb_{HCHO} \times \eta_a \times 10^{-9} \quad (4.13)$$

where  $\eta_{HCHO}$  is the number density of a HCHO, and  $ppb_{HCHO}$  is the molecules of that species per billion molecules of air. In order to normalize these vertical density profiles over the globe, we divide by the modelled total vertical column  $\Omega_{HCHO}$  which is determined by:

$$\Omega_{HCHO} = 2.12 \times 10^{13} \sum_z (ppb_{HCHO}(z)(P(z) - P(z + 1)))$$

where  $P(z)$  is the pressure (hPa) at the bottom of altitude level  $z$ , the constant 2.12e13 is determined from equation (TODO: run through this number in another section?). In effect this equation sums over the molecules per  $\text{cm}^2$  in each altitude level.

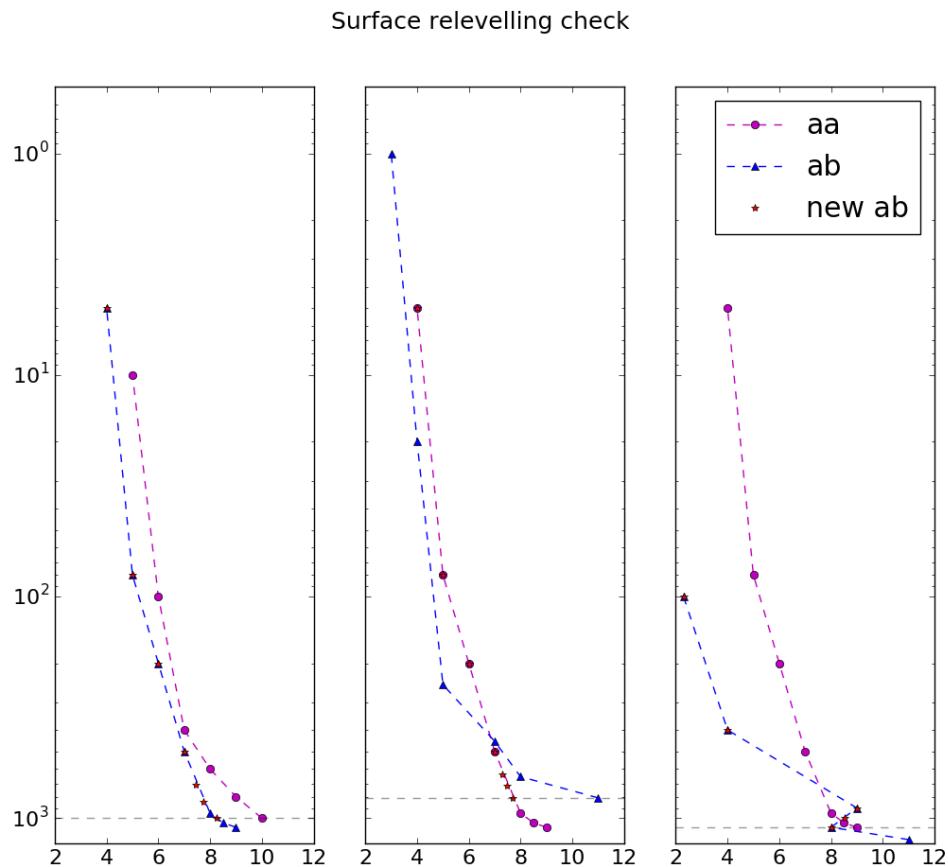
We have  $S_z(z)$  and  $\omega(z)$  over the vertical pressure coordinate  $z$  at all latitude and longitude points on whatever grid we wish. A conversion to the sigma ( $\sigma$ ) vertical coordinate is performed using  $P = \sigma(P_S - P_T) + P_T$ , where  $P_T$  is pressure at the top of the atmosphere and  $P_S$  is surface pressure. In the sigma coordinate system we calculated the shape factor as follows:

$$S_\sigma(\sigma) = \frac{\Omega_a}{\Omega_v} C_{HCHO}(\sigma) \quad (4.14)$$

where  $\Omega_a$  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. This equation comes from [Palmer2001], and is unitless since  $\Omega_a/\Omega_v$  is molecules of air per molecule of HCHO; the opposite of  $C_{HCHO}$ .

Pressure dimension from OMI are the surface pressures from each gridbox (of-

fine conversation with Dr Christopher Miller). Determining the geometric pressure midpoints (here onwards pressure levels) and interpolating to our increased vertical resolution involves a few steps. The lowest level (with highest pressure) in whichever pressure dimension (ours or OMI's) extends to the lowest altitude (or highest pressure) is interpolated upwards to match the lowest level in the other dimension. Secondly, if the OMI dimension has been changed, the scattering weights are interpolated onto this updated dimension. Figure 4.8 shows how these first two steps are applied using three fake array comparisons and updating the array with the lower surface level. Finally, once our dimensions match at the surface (we are not so worried about the very top of the atmosphere) we interpolate the scattering weights onto our updated GEOS-Chem pressure dimension.



**Figure 4.8:** Constructed example of the initial interpolation of OMI's  $\omega$  onto a pressure dimension with mismatched surface pressure.

$S_\sigma(\sigma)$  is determined after running GEOS-Chem, which outputs vertical profiles of air density and HCHO mixing ratio, at 72 vertical levels with associated metadata such as vertical layer height and pressure, grid box location, height, and surface pressure. Using these outputs the vertical columns ( $\Omega_a, \Omega_v$ ) are calculated for each

horizontal grid point (i, j) as follows:

$$\begin{aligned}\Omega_a(i, j) &= \Sigma_z (N_a(i, j, z) \times H(i, j, z)) \\ \Omega_z(i, j) &= \Sigma_z (N_{HCHO}(i, j, z) \times H(i, j, z))\end{aligned}$$

where  $N_a$ , and  $N_{HCHO}$  are the densities of air and HCHO, H is the layer height (for each grid box). Note that HCHO density is determined from the outputted mixing ratio:  $N_{HCHO} = C_{HCHO} \times N_a$ .

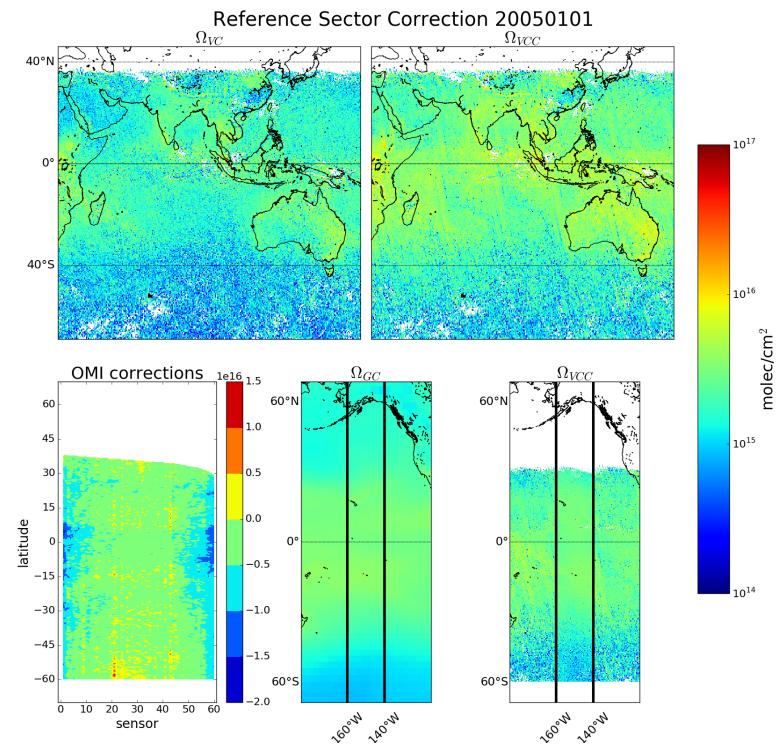
$S_\sigma(\sigma)$  is then stored in HDF-EOS5 format, to be used in conjunction with the satellite measurements to calculate an AMF as shown in equation 4.9. As the GEOS-Chem V10.01 output is in bitpunch format, the code to read the data and create the shape factor is written in IDL, which has many procedures and functions which are already written to handle reading this format (provided by GAMAP). The code is provided in supplementary TODO: put code into supplement section.

For each OMI slant column, a new AMF is calculated using  $S_\sigma(\sigma)$  and the provided scattering weights  $\omega(\sigma)$  using equation 4.9. This integral is applied in python by taking the sum of  $S_\sigma(\sigma) \times \omega(\sigma) \times d\sigma$  for each  $\sigma$  determined at 72 levels in GEOS-Chem, with the provided  $\omega$  interpolated linearly to these same levels. An example of these interpolations is shown in figure TODO: interpolation figure with symbols at original points and interpolated line overplotted for both functions over hPa. Globally this reprocessing changed the AMF by TODO: global total percent difference in AMF. In total this caused TODO: total column HCHO change globally/yearly In Summer over Australia the global AMF difference was TODO: Difference summers only. This changed Australia's HCHO amounts from TODO: X to Y Tg per year plus minus one std.

#### 4.3.8 Determining and applying the pacific ocean reference sector normalisation

As is done in [Abad2015], a reference sector defined over the Pacific ocean is used to correct OMI instrument degradation. This correction is calculated based on satellite measurements over the pacific ocean reference sector; between 140°W and 160°W, covering every latitude. Corrections are made using apriori HCHO columns in the same reference sector modelled using GEOS-Chem. The apriori reference sector of HCHO vertical columns (VCs) is created by GEOS-Chem using 15 minute time resolution, with 2 by 2.5° latitude by longitude resolution. These simulated values use the GEOS-Chem output averaged between 1300 and 1400 local time at each grid box, in order to match the overpass time of OMI. The longitudinal average is taken within the apriori reference sector, as corrections are assumed to be longitudinally

invariant. The modeled reference sector is interpolated latitudinally in for use in the OMI measurement correction array creation. Figure 4.9 the simulated reference sector VCs as an example, calculated on January 1st 2005. In this figure the vertical resolution is increased from  $2^\circ$  to  $0.36^\circ$ , through linear interpolation, in order to form 500 vertical bins which are used in correcting the satellite data. Each day, good satellite measurements taken over the reference sector are used to determine a correction array. The correction is based on the difference between measured slant column and the modeled slant column within the reference sector. The model does not produce slant columns associated with each measurement, however one is created by multiplying the VC with the associated slant column's AMF.



**Figure 4.9:** Example of remote pacific reference sector correction (RSC) using 8-day average measurements and one month modelled data.  $\Omega_{VC}$  shows the uncorrected vertical columns, while  $\Omega_{VCC}$  shows the corrected vertical columns. OMI corrections shows the correction applied globally based on latitude and OMI track number(sensor).  $\Omega_{GC}$  shows the GEOS-Chem modelled HCHO VC over the RSC, with  $\Omega_{VCC}$  showing the corrected VC over the same area.

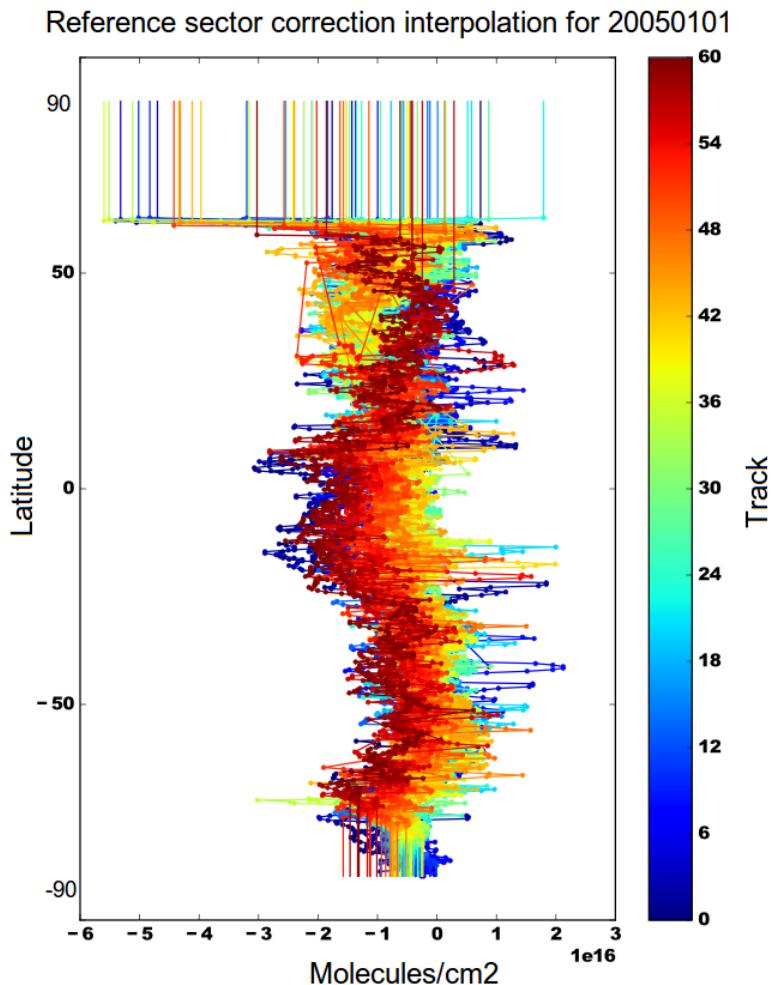
For OMI swaths, each row of measured data contains 60 ‘Across track’(track) measurements. The track index ( $i$ ) relates a the measurement to one of the 60 columns of data. Corrections for each measurement are calculated by taking the difference between the measured slant column and the apriori slant column as follows:

$$\text{Correction}(i, j) = SC_{HCHO}(i, j) - VC_{GEOS-Chem}(\text{lat}(j)) \times AMF_{OMI}(i, j) \quad (4.15)$$

where  $j$  represents a latitude index and  $VC_{GEOS-Chem}(lat)$  represents the apriori reference sector vertical column HCHO at the latitude corresponding to  $j$ . Note that the correction is in molecules per  $\text{cm}^2$ .

The reference sector correction is independently calculated for each of the 60 tracks, at each latitude where a good satellite measurement exists which used that track. The  $\text{Correction}(i, lat(j))$  function is determined by binning corrections for each track into 500 equidistant latitude bands.

Due 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 4.10 shows an example of the 60 track corrections for January 1st 2005, the points are satellite measurements and the lines are the interpolations for each track.



**Figure 4.10:** Example of track correction interpolations for January 1st 2005, points represent satellite slant column measurements, with lines interpolating and extrapolating along the latitudinal direction.

Another way to look at this correction is given in the OMI corrections panel of

figure 4.9, which has the sensors along the x axis, and latitude on the y axis, and shows how for this example 8-day period, the corrections are distributed with more negative values towards the left or right sensors, especially in the tropics.

One correction 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:

$$VCC(i, j) = \frac{SC_{HCHO}(i, j) - Correction(i, lat(j))}{AMF(i, j)} \quad (4.16)$$

Finally, for each day, the good satellite measurements are averaged into our own latitude longitude resolution bins along with the associated corrected SC, VC, VCC, AMF, and bin entry count. The bin entry count is used to create an 8-day average out of the one day averages, as it is the daily mean multiplied by the daily count summed over 8 days divided by the total count for each bin.

### 4.3.9 Estimation of error or uncertainty

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

- a Fitting error from the OMI retrieval.
- b Uncertainty in AMF calculations.
- c Uncertainty of HCHO background.

a) is available in the OMI product and reduced through spatial and temporal averaging. Another method for examining uncertainty of OMI is to analyse the standard deviation of the HCHO columns over the remote pacific. If we assume there is no HCHO variation from background levels over any 8-day period, then this method infers variations in the measuring instrument, and can be used as a metric for uncertainty as done in [DeSmedt2012]. TODO: uncertainty calculation on remote pacific OMI.

b) and c) are determined through an analysis of GEOS-Chem output, validated against the total column of HCHO at Wollongong using FTIR measurements from the (TODO: Nicholas Jones roof HCHO citation here). TODO: uncertainty calculations. Compare this error estimate with that of [Curci2010], where the error in b) and c) are respectively found to be 30% and 15% based on their analysis of CHIMERE.

### 4.3.10 Comparison with in-situ measurements

TODO: Describe Wollongong FTIR and junk Analyse comparison of gridbox with instrument!

# Chapter 5

## Isoprene Emissions in Australia

### 5.1 GEOS-Chem isoprene mechanisms

#### 5.1.1 Outline

The isoprene reactions simulated by GEOS-Chem were originally based on (TODO: read+citet: Horowitz1998) (TODO: check wiki that this is true also). This involved blahblah(read that paper). The mechanism was subsequently updated by [Mao2013], who change the isoprene nitrates yields and add products based on current understanding as laid out in [Paulot2009a, Paulot2009b].

Formation of isoprene nitrates have an effect on ozone levels through  $\text{NO}_X$  sequestration, and the yields and destinies of these nitrates is analysed in [Paulot2009a]. They use anion chemical ionization mass spectrometry (CIMS) to determine products of isoprene photooxidation. In a chamber with clean air and high NO concentrations, isoprene photooxidation is initially driven by OH addition, followed by  $\text{NO}_X$  chemistry (150 min - 600 min), and finally  $\text{HO}_X$  dominated chemistry. The yields of various positional isomers of isoprene nitrates is estimated, and pathways of their oxidation products is shown and used in the GEOS-Chem isoprene mechanism ([Paulot2009a, Mao2013]).

In low  $\text{NO}_X$  conditions, isoprene oxidises to yield 70% hydroxyhydroperoxides (ISOPOOH), which then oxidises to create dihydroxyperoxides (IEPOX) with OH recycling maintaining the OH levels in the atmosphere ([Paulot2009b]). In older models isoprene produced ISOPOOH which then titrated OH, however, the loss of OH has not been seen in measurements ([Paulot2009b, Mao2013]). The isoprene mechanism in GEOS-Chem has been updated to include OH regeneration from oxidation of epoxydiols and slow isomerisation of ISOPO<sub>2</sub> ([Mao2013]).

Under high  $\text{NO}_X$  conditions, isoprene undergoes OH addition at the 1 and 4 positions, becoming  $\beta$  (71%) or  $\delta$  (29%) hydroxyl peroxy radicals (ISOPO<sub>2</sub>). The  $\beta$ -hydroxyl reacts with  $\text{NO}_X$  and produces HCHO (66%), methylvinylketone (40%)

(MVK), methacrolein (26%), and  $\beta$ -hydroxyl nitrates (6.7%) (ISOPNB). The  $\delta$ -hydroxyl reacts with NO to form  $\delta$ -hydroxyl nitrates (24%) (ISOPND), and ISOPNB (6.7%). ISOPNB and ISOPND yield first generation isoprene at 4.7% and 7% respectively.

Under low  $\text{NO}_X$  conditions, ISOPO<sub>2</sub> may react with HO<sub>2</sub> to form ISOPOOH. In this case there is also production of HCHO (4.7%), MVK(7.3%), and MACR (12%). As stated in earlier; most ISOPOOH will form IEPOX (epoxydiols) after reacting with OH and lead to OH regeneration. The other mechanism in low  $\text{NO}_X$  environments is unimolecular isomerisation of ISOPO<sub>2</sub>. This leads to production of hydroperoxyaldehydes (HPALDS), which generally photolysse and have an OH yield of 100%. [Mao2013] show that a lower (factor of 50) rate constant for ISOPO<sub>2</sub> isomerisation leads to better organic nitrate agreements with ICARTT.

This update leads to more accurate modelling of OH concentrations ([Mao2012]) (todo read abstract and cite).

The updates to isoprene chemistry by [Mao2013] are the last before version 11, which was not used in this work.

The full current mechanism is described online at [http://wiki.seas.harvard.edu/geos-chem/index.php/New\\_isoprene\\_scheme](http://wiki.seas.harvard.edu/geos-chem/index.php/New_isoprene_scheme) (TODO: check this).

### 5.1.2 Emissions from MEGAN

MEGAN simulates biogenic emissions of various gases including isoprene, based on various meteorological, land cover, and plant type parameterisations.

One of the important parameters in Australia is the soil moisture activity factor ( $\gamma_{SM}$ ), which can have large regional affects on the isoprene emissions ([Sindelarova2014, Bauwens2016]). Generally if soil moisture is too low, isoprene emissions stop, however in many Australian regions the plants may be more adapted to lower moisture levels. (TODO: 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 ( $\theta$ ):

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

where  $\theta_w$  is the wilting point, and  $\theta_1$  determines when plants are near the wilting point. The wilting point is set by a land based database from [Chen2001], while  $\theta_1$  is set globally based on [Pegoraro2004]. (TODO: Check both these papers from

Guenther2006)

## 5.2 Isoprene emissions estimation

### 5.2.1 HCHO Products and yield

Australian forests are strong emitters of both isoprene and monoterpenes, which go on to form various products including but not limited to secondary organic aerosols, oxygenated VOCs (OVOCs), ozone, OH, and HO<sub>2</sub>. This production occurs over several steps, yields are often classed into at least two categories. First generation yield refers to the amount of HCHO produced per unit isoprene consumed by initial oxidation, total yield (sometimes molar yield) refers to time dependent yield of HCHO over multiple oxidation stages ([Wolfe2016]). [Wolfe2016] define prompt yield as the change in formaldehyde measurement per unit change in initial isoprene emissions. Some argue that isoprene emissions are overestimated, due to the fact that they are based on relatively few measurements of isoprene emission factors ([Winters2009, FortemsCheiney2012]) TODO: read and cite paper mentioned in Fortems.

Isoprene production of HCHO depends on several factors, importantly NO<sub>X</sub> levels have a direct effect on the fate of VOCs in the atmosphere. At higher NO mixing ratios (at least a few hundred pptv), organic peroxy radicals (RO<sub>2</sub>) react mostly with NO. At low NO (less than 10's of pptv), reaction with HO<sub>2</sub>, other RO<sub>2</sub>, and isomerization dominate the fate of RO<sub>2</sub>. In low NO<sub>X</sub> environments, reported HCHO yields from isoprene are from XtoY%, while in high NO<sub>X</sub> environments this value is XtoY% TODO: these values from table. For monoterpenes the yields are around X, Y% for low, high NO<sub>X</sub> respectively. Emissions and yields for various species including some terpenes can be seen in table 5.1. [Wolfe2016] determine that going from NO<sub>X</sub> = 0.1 to 2.0 ppbv triples the prompt yield of HCHO, from 0.3 to 0.9 ppbv ppbv<sup>-1</sup> due to isoprene, while the background HCHO doubles. They determine prompt yield as the change in HCHO per change in ISOP<sub>0</sub>, using  $[ISOP]_0 = [ISOP] \exp(k_1[OH]t)$ ; where  $k_1$  is first order loss rate. This effectively relates HCHO abundance with isoprene emission strength. (TODO:and finish Wolfe2016 discussion paper for yields) TODO:go through atkinsonarey2003

Many of the HCHO yields from terpenoids are estimated through chamber studies which examine the products molecular mass and charge after mixing the compound of choice into a known volume of air. These conditions generally don't match those of the real world, where ambient air will have a cocktail of these compounds as well as various reactants.

A proton transfer reaction mass spectrometer (PTR-MS) can be used to de-

**Table 5.1:** HCHO yields from various species averaged over Australia during Summer.

Species	Emissions <sup>a</sup> (Tg C per month)	Lifetime <sup>b</sup>	HCHO Yield <sup>c</sup> (per C reacted)	HCHO production <sup>d</sup> %
Isoprene	Y	n minutes	0.x	10
$\alpha$ -Pinene	Y	n minutes	0.x	10
$\beta$ -Pinene	Y	n minutes	0.x	10
HCHO	Y	n minutes	1.0	10

a: Calculated using GEOS-Chem emissions over Australia in January 2005.

b:

c:

d: Production determined by dividing emission\*yield by the sum of all VOC emissions\*yields.

termine gas phase evolution of terpene oxidation products. This is done through analysis of mass to charge ratios ( $m/z$ ) which can be used identify chemical compounds.

Looking at Australian emissions from running GEOS-Chem and using yields provided by XYZ (another table) TODO: Fill out this table

Yields table looking at literature provided yields of HCHO.

### 5.2.2 Outline

Once we have vertical columns of biogenic HCHO we can infer the local (grid space) isoprene emissions using effective molar formaldehyde yield (around 2-3, or 1 in low  $\text{NO}_x$  conditions) ([Palmer2003, Marais2013, Bauwens2016]). If we assume there is fast HCHO yield, so that the effect of chemical transport is minimal, and that HCHO and isoprene are at steady states, then we can calculate local yield from our CTM. This yield is derived from both HCHO and isoprene, such as was used by [Millet2006] who produced a molar HCHO yield of 2.3 in north eastern USA. The calculations used to determine isoprene emissions over Australia are fully described in 5.2.3 and follow the method of [Palmer2003]. To calculate emissions we use a reduced major axis (RMA) regression between modelled average (from 1000-1200 LT) values of the loss rates and total columns, an example is shown in figure TODO: figure with RMA of these over whatever time and space I end up using.

The background HCHO is assumed to be the HCHO measured in the remote pacific at the same time. This is determined from GEOS-Chem when calculating yield of HCHO from isoprene, and from the biogenic HCHO product when calculating isoprene emissions. (TODO: update this if that is not the final process)

Isoprene quickly forms HCHO in the atmosphere when in the presence of high levels of  $\text{NO}_x$ . However, over Australia  $\text{NO}_x$  levels are generally not high enough

**Table 5.2:** HCHO yields from various species, and lifetime against oxidation by OH.

Species	HCHO Yield (molar % )	Life vs OH	Notes	Source
Isoprene	315±50		High NO <sub>X</sub>	a
	285±30		High NO <sub>X</sub>	a
	225	35 min	High NO <sub>X</sub>	b
	150		Low NO <sub>X</sub>	b
	150		Low NO <sub>X</sub>	d
	450		High NO <sub>X</sub>	d
$\alpha$ -Pinene	28±3		Low NO <sub>X</sub>	c
	X±3		X NO <sub>X</sub>	d
	230±90		High NO <sub>X</sub>	a
	190±50		High NO <sub>X</sub>	a
	19	1 hour		b
$\beta$ -Pinene	65±6		Low NO <sub>X</sub>	c
	X±3		X NO <sub>X</sub>	d
	540±50		High NO <sub>X</sub>	a
	450±80		High NO <sub>X</sub>	a
	45	40 min		b
Methane	100	1 year		b
Ethane	180	10 days		b
Propane	60	2 days		b
Methylbutanol	.13(per C)	1 hour		b
HCHO	100	2 hour		b
Acetone	.67(per C)	10 days		b
Methanol	100	2 days		b

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<sup>-3</sup>.

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 ISOP<sub>0</sub>.  $[ISOP]_0 = [ISOP] \exp(k_1[OH]t)$ ; where  $k_1$  is first order loss rate. Effectively relates HCHO abundance with isoprene emission strength

e Calculated from satellite detected concentrations of HCHO.

f Calculated using PTR-MS and iWAS on SENEX campaign data.

and we must take extra care that we can account for the transport or 'smearing' caused by slower HCHO formation. Smearing sensitive grid boxes within the model can be detected by running the model with two uniformly differing isoprene emission levels, then finding the grid boxes where the changed HCHO column is greater than can be attributed to local emission difference. Using equation 5.2 with two different isoprene emission levels:

$$\hat{S} = \frac{\Delta \Omega_{HCHO}}{\Delta E_{ISOP}}$$

Consider halving the isoprene emitted globally and rerunning the model, if the local grid HCHO is reduced by much more than half (factoring yield) then you can infer sensitivity to non-local isoprene emissions. This 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 sensitivity is both spatially and temporally diverse, shown in figure TODO: is a picture of the smearing sensitivity over Australia.

Once the smearing sensitive grid squares are filtered out, application of equation 5.2 gives us an estimate of isoprene emissions across the nation.

Most recently a [Bauwens2016] undertook a similar process to what I am doing, although with slightly different focus, using the IMAGESv2 global CTM instead of GEOS-Chem. They calculate emissions which create the closest match between model and satellite vertical columns, and compare these postiori data with the apriori (satellite data) and independent data sets. (TODO: simple outline of what they did and how my focus is different, this paper will also need to be summarised in the LitReview)

### 5.2.3 Calculation of Emissions

As is done in [Palmer2003, Millet2006, Bauwens2016], we assume that HCHO, and Isoprene columns are in a steady state, with no horizontal transport. In these circumstances the emissions of precursors are easy to calculate as long as we know the molar HCHO yields ( $Y_i$ ) and effective chemical loss rates ( $k_i$ ):

$$\Omega_{HCHO} = \frac{1}{k_{HCHO}} \sum_i k_i Y_i \Omega_i = \frac{1}{k_{HCHO}} \sum_i Y_i E_i \quad (5.1)$$

We can infer the local (grid space) isoprene emissions ( $E_{isop}$ ) using effective formaldehyde yield from isoprene ( $Y_{isop}$ ).

$$\Omega_{HCHO} = S \times E_{isop} + B \quad (5.2)$$

Where  $B$  is the background HCHO, and  $S = Y_{isop}/k_{HCHO}$  is determined monthly as

the regression between  $k_{HCHO} * \Omega_{HCHO}$ , and  $k_{isop} * \Omega_{isop}$ . For an initial estimate of the effective yield from simulated data: we use  $k_{HCHO}$ ,  $k_i$ ,  $\Omega_{HCHO}$ , and  $\Omega_i$  outputs from a standard run of GEOS-Chem - which provides one data point per day. This gives us a value for  $Y_i$  resolved to our  $2^\circ$  by  $2.5^\circ$  horizontal resolution, which is entirely based on the model, and can be compared against the yield calculated using OMI derived  $\Omega_{HCHO}$ . Using our measurements of the biogenic HCHO column ( $\Omega_{OMIHCHO}$ ) recalculated from the OMI satellite product, we derive yield using the same method and determine our new top-down emissions estimates. Figure ()TODO: figure with RMA of these over whatever time and space I end up using.) shows the modelled isoprene emissions and column concentrations along with the RMA regression line - which leads to an estimated yield of TODO

This works if there is fast HCHO yield, so that the effect of chemical transport is minimal. The background HCHO is calculated using measurements in the remote pacific at the same time and latitude. Table 5.1 shows the average yield calculated for Australia. (TODO: this table and some notes)

### 5.2.4 Calculation of smearing effect

TODO: Smearing scale length,  $\hat{S}$  formula, and results of calculations in here.

### 5.2.5 Extrapolating the circadian cycle

Isoprene emissions occur with regular daily cycles caused by things like local temperature, sunlight, drought, and other environmental factors (TODO: find/cite eucalypt isoprene paper, daily cycle plot if can find).

(TODO: following stuff, add some basic plots and error analysis eventually also) Using a model of the daily isoprene emissions fit to the offset determined by satellite HCHO based estimates, we produce a high temporal resolution isoprene emissions inventory. During days with more than one HCHO column measurement we can more confidently fit the cycle. For example EOS AURA's OMI measurements from 2004 can be combined with MetOp-A's GOME2 after October 2006, with daily overpasses by OMI and GOME2 at 1345 and 0930 respectively. This allows a better retrieval of the daily amplitude of isoprene emissions.

### 5.2.6 Comparison with MEGAN

TODO: Direct comparison here, maps of differences for some metrics(monthly average,?). comparison of model run results using different inventory shown in section (reference here)

### **5.3 New estimates affects on the Australian atmosphere**

# **Appendix A**

## **Appendix A**

### **A.1 Grid Resolution**

Level	z(km)	eta
1	0.058483	0.99250000
2	0.188515	0.97745600
3	0.320427	0.96237000
4	0.453892	0.94728500
5	0.588963	0.93220000
6	0.725697	0.91711600
7	0.864132	0.90203100
8	1.004320	0.88694800
9	1.146330	0.87186400
10	1.290220	0.85678100
11	1.436040	0.84169800
12	1.583870	0.82661600
13	1.758950	0.80902100
14	1.988400	0.78640000
15	2.249380	0.76126500
16	2.517070	0.73613400
17	2.791920	0.71100600
18	3.074460	0.68587800
19	3.439220	0.65447100
20	3.895550	0.61679000
21	4.374520	0.57911500
22	4.879030	0.54144900
23	5.412570	0.50379500
24	5.979530	0.46615300
25	6.585330	0.42852800
26	7.236820	0.39092700
27	7.943070	0.35334900
28	8.845840	0.30985400
29	9.936460	0.26358700
30	11.020800	0.22377200
31	12.085500	0.19006100
32	13.133500	0.16151300
33	14.169600	0.13728700
34	15.198200	0.11669500
35	16.221700	0.09919110
36	17.242600	0.08431270
37	18.727200	0.06655890
38	20.835700	0.04764090
39	23.020500	0.03381390
40	25.306800	0.02375520
41	28.654300	0.01434240
42	34.024400	0.00658754
43	40.166100	0.00281639
44	47.135400	0.00110910
45	54.833600	0.00039878
46	63.052800	0.00012699
47	72.180100	0.00002763

**Table A.1:** The 47 level vertical grid used by GEOS-Chem