

2.6.1	Outline	77
2.6.2	Creating new shape factors	79
2.6.3	Creating new AMF using GEOS-Chem	80
2.6.3.1	Updating shape factors	80
2.6.3.2	Recalculating the AMF using PP code	80
2.6.3.3	Saving the AMF with satellite pixels	81
2.6.4	Vertical columns from AMF	81
2.6.5	Reference sector correction	81
2.6.6	Corrected vertical columns	82
2.6.7	Binning the results daily	85
2.6.8	Difference between original and corrected OMI HCHO columns	86
2.7	Filtering Data	89
2.7.1	Pyrogenic filter	91
2.7.2	Anthropogenic filter	98
2.8	Process schematic	106
2.9	Data Access	107
3	Biogenic Isoprene emissions in Australia	109
3.1	Introduction	109
3.1.1	Aims	110
3.1.2	Existing emissions estimates	110
3.1.3	Top-down isoprene emissions estimates	111
3.1.3.1	Bayesian inversion	111
3.1.3.2	Linear inversion	112
3.2	Methods	113
3.2.1	Outline	113
3.2.2	Masks and reprocessed satellite HCHO	114
3.2.3	GEOS-Chem emissions	116
3.2.4	Relationship between isoprene emissions and formaldehyde	116
3.2.5	Calculation of modelled slope	118
3.2.6	Calculation of Emissions	120
3.2.7	Accounting for smearing	124
3.2.7.1	Calculation of smearing	124
3.2.7.2	Sensitivity to smearing	126
3.2.7.3	Smearing length scale	129
3.2.7.4	NO _x dependence	130
3.2.8	Running GEOS-Chem using a posteriori emissions	130
3.3	Results	135
3.3.1	A posteriori emissions	135
3.3.2	Modelled impacts of reduced isoprene emissions	144
3.3.2.1	Implications for HCHO	144
3.3.2.2	Implications for ozone	144
3.3.2.3	Trends	146
3.3.3	Comparison with in situ measurements	146
3.4	Uncertainty	150
3.4.1	Top down emissions	151
3.4.2	Model Uncertainty	152

3.4.3	Satellite Uncertainty	156
3.4.3.1	Fitting error from the OMI retrieval	160
3.4.3.2	Uncertainty in AMF calculations	160
3.4.3.3	Uncertainty of HCHO background	160
3.4.4	Sensitivity to vertical column recalculation	160
3.4.5	Sensitivity to filtering	161
3.5	Conclusions and implications	161
4	Stratospheric ozone intrusions	163
4.1	Foreword	163
4.2	Introduction	163
4.3	Data and Methods	165
4.3.1	Ozonesonde record in the Southern Ocean	165
4.3.2	Model description	168
4.3.3	Characterisation of STT events and associated fluxes	169
4.3.4	Biomass burning influence	171
4.3.5	Classifying synoptic conditions during STT events	172
4.4	STT event climatologies	173
4.5	Simulated ozone columns	176
4.6	Stratosphere-to-troposphere ozone flux from STT events	181
4.6.1	Method	181
4.6.2	Results	182
4.6.3	Comparison to literature	182
4.7	Sensitivities and limitations	187
4.7.1	Event detection	187
4.7.2	Flux calculations	188
4.8	Conclusions	189
4.9	Contributions and Acknowledgements	190
5	Summary and Concluding remarks	193
5.1	Outcomes	193
5.2	Isoprene emissions	195
5.3	Ozone over Australia	195
5.4	Outputs and future work	197
Bibliography		199

Chapter 3

Biogenic Isoprene emissions in Australia

3.1 Introduction

Biogenic volatile organic compounds (BVOC) affect the oxidative capacity of the atmosphere and are largely driven by what type of vegetation is in the area (Kefauver, Filella, and Peñuelas 2014). In the troposphere, BVOC emissions affect hydroxyl radical (OH) cycling, ozone (O_3) and secondary organic aerosol (SOA) production, and methane lifetime. Australian forests are strong emitters of isoprene, the primary BVOC emitted globally (Guenther et al. 2006; Messina et al. 2016). Poor measurement coverage of isoprene, isoprene products, and isoprene emissions within Australia means that emissions are poorly understood. The lack of measurements makes it difficult to estimate the subsequent atmospheric processes. Isoprene (C_5H_8) is relatively difficult to measure due to its high reactivity and short lifetime.

Emission models used to derive estimates of isoprene fluxes are based on understanding the emissions from different plant species (phenotypes) in varying conditions. Guenther et al. (2012) estimated global biogenic isoprene emissions at roughly 535 Tg yr^{-1} , while Sindelarova et al. (2014) estimated around 411 Tg yr^{-1} . Reactions following emissions are complex, and are sensitive to other trace gases in the ambient atmosphere. Uncertainties in several important products such as ozone and SOA are increased due to both isoprene measurement difficulties and its complicated subsequent chemical mechanisms. Isoprene emissions may be overestimated in Australia since they are based on measurements taken from a few young trees (Winters et al. 2009) that may emit more than older trees (Emerson 2016). The sample of trees include 4 types of Eucalyptus, which are not representative of the hundreds of species that make up Australian forests, and how these species depend on biological and meteorological stresses is unclear (Winters et al. 2009; Fortems-Cheiney et al. 2012). Emissions estimates are often used as boundary conditions for atmospheric chemistry models and improving these estimates for Australia is one goal of this thesis.

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

technique very attractive. Top down techniques have informed isoprene emission inventories in North America (Abbot 2003; Palmer et al. 2003; Palmer et al. 2006; Millet et al. 2006; Millet et al. 2008), South America (Barkley et al. 2013), Europe (Dufour et al. 2008; Curci et al. 2010), Africa (Marais et al. 2012), Asia (Fu et al. 2007; Stavrakou et al. 2014), India (Surl, Palmer, and Abad 2018), and even globally (Shim et al. 2005; Fortems-Cheiney et al. 2012; Bauwens et al. 2016). In this thesis I apply the technique solely focusing on Australia for the first time.

3.1.1 Aims

Recent work suggests that modelled emissions may be overestimated in Australia (Emmerson et al. 2016). This ~~work tries~~^{southeast} to improve the understanding of isoprene emissions over the whole of Australia, clarifying the spatial distribution of bias and how these biases impact modelled chemistry. I estimate isoprene emissions in Australia using a top-down technique based on OMI HCHO measurements and GEOS-Chem modelled yields. This a posteriori top-down estimate is then evaluated against sparse available ground-based measurements. The GEOS-Chem model is modified to run with the a posteriori isoprene emissions to determine potential impact on modelled chemistry. Goodness of fit between in situ, satellite, and modelled HCHO is determined before and after scaling emissions estimates.

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

3.1.2 Existing emissions estimates

The Model of Emissions of Gases and Aerosols from Nature (MEGAN) is one of the most widely used sources for biogenic isoprene. Along with other models that rely on measured plant emission rates, it is poorly calibrated for Australian conditions. Emissions of isoprene (C_5H_8) appear to be overestimated ~~in~~ within Australia (Sindelarova et al. 2014; Stavrakou et al. 2014; Emmerson et al. 2016). Although the lack of measurements of isoprene emission rates in Australia makes this overestimation difficult to characterise. Stavrakou et al. (2015) ~~showed that~~ saw isoprene emissions ~~were~~ overestimated by a factor of 2-3 in January. Emmerson et al. (2016) suggest isoprene emissions are estimated 2-6 times too high compared against available measurements of isoprene concentrations. They also show that no blanket increase or decrease in emission factors is appropriate for the entire southeast of Australia. They compared modelled data to campaign measurements from multiple sites over different seasons and found that scaling emissions did not universally improve model outputs.

Recently Bauwens et al. (2016) estimated isoprene emissions with a top-down technique using the IMAGESv2 global CTM. They calculated emissions that create the closest match between model and satellite vertical columns, and compared these a

where (Australia-wide)?
is this for MEGAN?

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

clarify the two
MEGAN estimates

3.1.3 Top-down isoprene emissions estimates

In the remote troposphere HCHO production is dominated by methane oxidation, while in the continental boundary layer production is largely due to non-methane VOCs (NMVOCs) (Abbot 2003; Kefauver, Filella, and Peñuelas 2014). This leads to a causal relationship between enhanced HCHO concentrations and NMVOC emissions at low (< 1 km) altitudes. NMVOCs are generally short lived (< 1 hr), and the most prominent of these is isoprene. Isoprene is emitted and enters the atmosphere in the gas phase, where it begins a complex series of reactions. HCHO is produced with high yield in many reactions beginning with isoprene oxidation, discussed in more detail in Section 1.3.3, and has a lifetime of a few hours (Kefauver, Filella, and Peñuelas 2014).

Top-down estimates determine emissions of a particular species through careful analysis of the measurable products of that species. This generally takes advantage of longer-lived products that may reach an equilibrium in the atmosphere. For isoprene this is done through examination of atmospheric HCHO enhancement, which is a major product of isoprene oxidation ~~that takes place after emission~~. Since 1997, when GOME measurements were first used to measure HCHO over Asia, satellites have been used to provide~~d~~ a total column measurement of HCHO, providing isoprene emissions estimation by top-down methods (Thomas et al. 1998; Palmer et al. 2001; Bauwens et al. 2016). Using satellite information to improve estimates of biogenic emissions has been highlighted as a valuable use of satellite derived datasets (Streets et al. 2013). Here NASA's OMHCHO product based on measurements from the OMI instrument onboard the Aura satellite (see Section 2.3) is the basis for a top-down biogenic isoprene emission estimate over Australia.

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

3.1.3.1 Bayesian inversion

Bayesian inversion optimises model parameters in order to minimise the difference between model output and an (ideally) independent dataset such as satellite measurements. Emissions of isoprene (and other precursors to HCHO) will form part of the set of model parameters that are adjusted to make the model HCHO output most closely match satellite measurements. These inversions can be set up to account for effects from transport and allow source attribution (e.g. Curci et al. 2010; Fortems-Cheiney et al. 2012).

In general~~x~~, a model (the forward model) is used to determine the relationship between HCHO (y) and the state variable x , which represents isoprene emissions (and other variable parameters of interest):

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

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

This method was used by Shim et al. (2005) to optimise isoprene emissions in areas with high HCHO concentrations. They showed model underestimation of isoprene emissions by 14-46%, which when corrected reduced bias between GOME HCHO measurements and GEOS-Chem model output by 3-25%. More recently Kaiser2018 showed a 40% bias in MEGAN isoprene emissions over the southeast US using a Bayesian inversion based on OMI HCHO.

An advantage (over the linear method described below) of the Bayesian method is that it can account for pyrogenic and anthropogenic emissions, as these form part of the state variable x . However, biases may still arise due to errors in modelled emission estimation (Curci et al. 2010). More limiting is the fact that the Bayesian method is computationally expensive due to the requirement that model runs take place using many permutations of changed input parameters. In this work I do not use the Bayesian method due to the computational costs surpassing the resources available.

3.1.3.2 Linear inversion

The linear technique is the less complicated of the two, and is performed in this thesis. Vertical columns of HCHO from satellite and modelled yield from isoprene allow the inference of local (grid space) isoprene emissions (Palmer et al. 2003; Millet et al. 2006; Marais et al. 2012; Bauwens et al. 2016). In Australia the effective molar HCHO yield from isoprene has not been extensively studied, while in other continents this value varies from 1-3 depending on local NO_x concentrations (e.g. Palmer et al. 2006; Millet et al. 2006; Bauwens et al. 2016; Surl, Palmer, and Abad 2018). The primary assumption of the linear inversion technique is that HCHO and its precursors (primarily isoprene) are in a linear steady state relationship. This allows one to link isoprene emissions to HCHO measurements using production and loss rates. Essentially a linear relationship between total column HCHO (Ω) enhancement above a background level (Ω_0) and isoprene emissions (E_{isop}) is determined:

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

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

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

Probably not the most appropriate place for this sentence – move to where you discuss yield in more detail.

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

The calculation requires reaction rates and yields from isoprene to HCHO, which can be determined most readily using chemical modelling. The method for calculating isoprene emissions from HCHO is laid out in Palmer et al. (2003), taking into account the expected lifetime and reaction rates of the precursor VOCs and HCHO. In their work, isoprene emissions fluxes over the US were derived using the Global Ozone Monitoring Experiment (GOME) satellite instrument. The method has since been applied to many regions using OMI, GOME, GOME-2, and SCIAMACHY satellite data (e.g., Abbot 2003; Barkley et al. 2013; Stavrakou et al. 2014; Surl, Palmer, and Abad 2018). In this thesis I apply the technique solely focusing on Australia for the first time.

The linear inversion assumes fast HCHO yield from isoprene and no precursor transport, which is unrealistic in certain scenarios^{for example}, e.g. high wind speeds can transport precursors, or low NO_x concentrations can slow HCHO production (Palmer et al. 2006; Surl, Palmer, and Abad 2018). Filtering out data that do not match assumptions is required but can limit the utility of this technique, and leads to some dependence on environmental factors. Uncertainties in the technique are discussed in more detail in Section 3.4.1. Nonetheless, a major benefit is that the simple nature of the inversion requires very little computational power after acquiring satellite and model datasets, even over large amounts of gridded data. This allows an inversion using more than 8 years of satellite and model data, capturing inter-annual variability over all of Australia. With the computational resources available this would not have been possible using the Bayesian inversion.

3.2 Methods

I broadly follow the method of Palmer et al. (2001) to create a biogenic isoprene emissions estimate over Australia. A relationship is modelled between biogenic-only mid-day tropospheric columns of HCHO and GEOS-Chem midday biogenic isoprene emission rates, and then this relationship is applied to satellite measured HCHO total columns to derive a new isoprene emissions estimate. Daily modelled values averaged between 13:00-14:00 LT are used to match the overpass time of the Aura satellite. Then the slope is calculated using the reduced major axis (RMA) regression between the a priori isoprene emissions (those from GEOS-Chem, E_{GC}) and tropospheric HCHO columns in each model grid box each month. There is very little HCHO above the tropopause, so differences between total and tropospheric column are negligible. In this work I refer to both total and tropospheric column HCHO using Ω .

are used interchangeably and referred to

3.2.1 Outline

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

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

model background values (columns over the remote Pacific) are used to correct the vertical columns, which is explained in Section 2.6.5.

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

3.2.2 Masks and reprocessed satellite HCHO

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

In order to determine biogenic HCHO enhancements from Ω_{OMI} , we require filters for non-biogenic sources. These masks are described in Section 2.7, and a brief recap is provided here. While one source of HCHO production is methane oxidation, the linear regression used to estimate isoprene emissions effectively ignores this source as part of the background, which means a methane filter is not required. Anthropogenic,

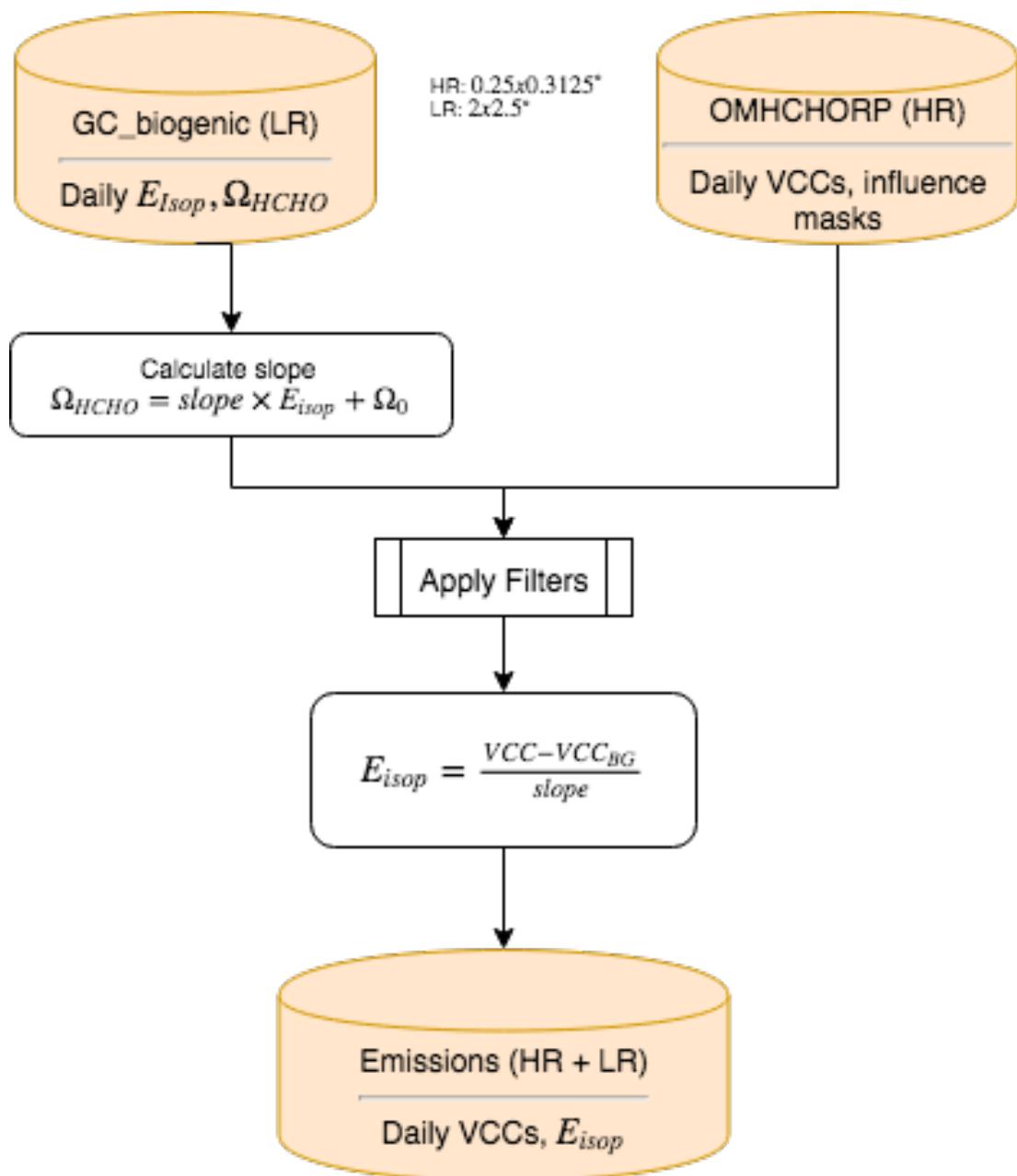


FIGURE 3.1: Top-down isoprene emissions estimate formation using OMHCHORP and biogenic GEOS-Chem outputs.

pyrogenic, and smoke influence masks are created from three satellite products: NO₂ from OMNO2d, fire counts from MOD14A1, and AAOD from OMAERUVd, respectively.

1. The fire mask is created daily using non-zero MODIS fire counts over the prior 2 days that occur in local or adjacent grid squares at 0.25° × 0.3125° horizontal resolution.
2. Influence from transported smoke plumes is removed where OMI aerosol absorption optical depth (AAOD, from OMAERUVd) is greater than 0.03.
3. A filter for anthropogenic influence is created daily using OMNO2d NO₂ tropospheric column amounts, masking any grid squares with greater than 2 × 10¹⁵ molec cm⁻² on any particular day, along with grid squares where the yearly average is above 1.5 × 10¹⁵ molec cm⁻².

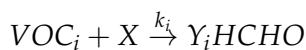
The recalculated corrected vertical columns are saved to OMHCHORP dataset both before and after applying the filters, to allow filter analysis.

3.2.3 GEOS-Chem emissions

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

3.2.4 Relationship between isoprene emissions and formaldehyde

Tropospheric HCHO production is primarily due to the oxidation of VOC precursor species (VOC_i). Background concentrations are driven by methane; a longer lived (~ 8 yr) VOC. Over continental land masses, the variability in HCHO is driven by shorter lived precursor emissions (Chance et al. 2000; Palmer et al. 2003). The intermediate steps are considered negligible as HCHO is produced quickly from short-lived intermediates:



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

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

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

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

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

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

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

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

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

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

Finally, we assume isoprene emissions are driving changes in Ω (as assumed elsewhere, e.g. Palmer et al. 2003; Millet et al. 2008; Marais et al. 2014; Stavrakou et al. 2015) and lump other terms together:

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

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

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

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

3.2.5 Calculation of modelled slope

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

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

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

Something missing in this sentence

Fig 3.2: top right never discussed. I would remove or replace with similar plot for one of the boxes used in your sample regressions. I would also re-order plots in fig and discussion here to:
 1. Examples of correlations used to calculate slopes
 2. R^2
 3. S that comes out of this.

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

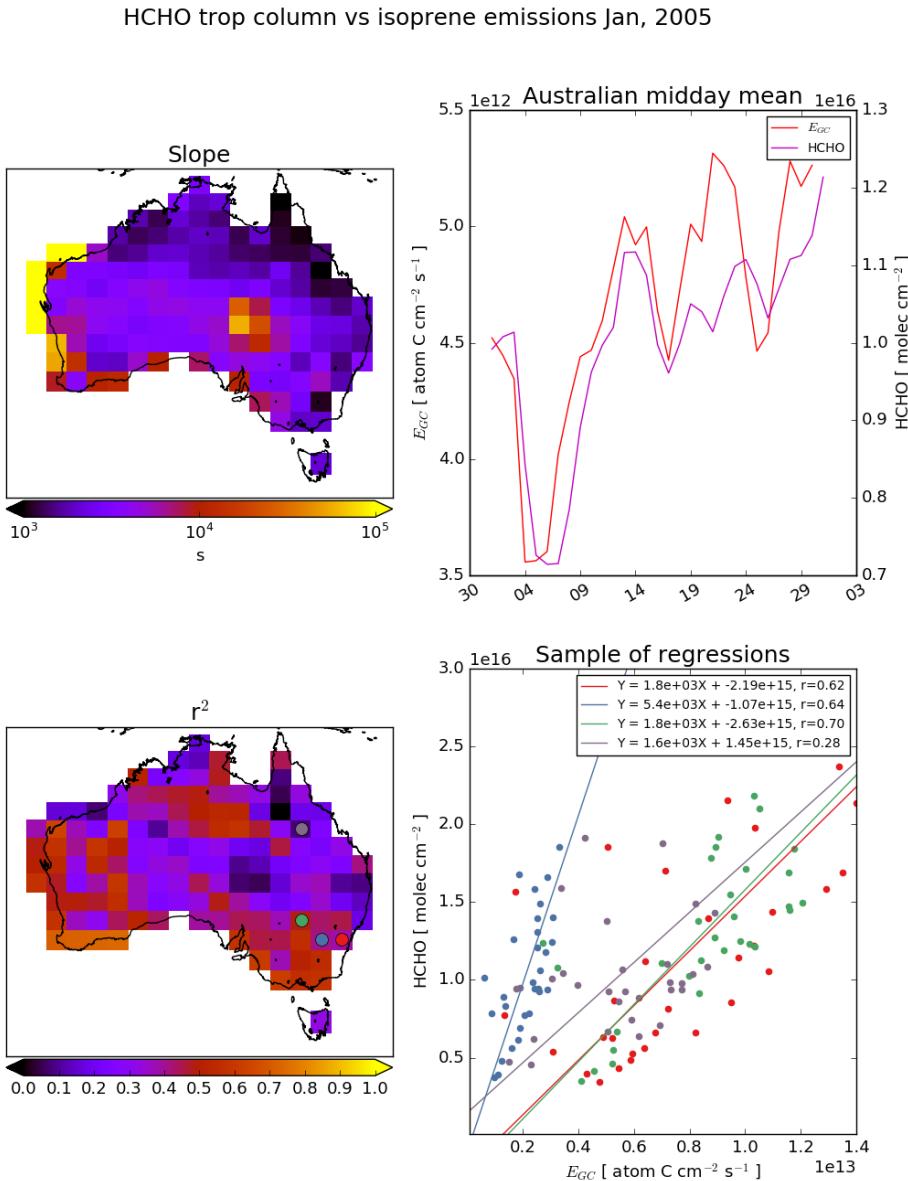


FIGURE 3.2: Top left: RMA slope between modelled tropospheric column HCHO and isoprene emissions (E_{GC}) using midday (13:00–14:00 LT) values over for January 2005, per grid square at $2^\circ \times 2.5^\circ$ horizontal resolution. Top right: Australia-wide average of midday emissions and tropospheric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

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

There are two simple ways to determine the modelled background HCHO, one of which involves running the model with no isoprene emissions. Since we have assumed variation in HCHO columns only depends on isoprene emissions, our background term is theoretically identical to this simulated HCHO. The other method uses HCHO over the remote Pacific at matching latitudes and times, which emulates how the background is determined for the satellite measured HCHO. Figure 3.4 shows GEOS-Chem total column HCHO with and without isoprene emissions along with amounts over the remote Pacific at the same latitudes. The difference in Ω_{GC} over Australia with no isoprene emissions, and Ω_{GC} over the remote Pacific (bottom right panel) shows the difference (~ 1 to $3 \cdot 10^{-10}$ molec cm $^{-2}$) between these methods in an example averaged month (January, 2005). This difference is relatively small, and may be due to non-isoprene HCHO precursors. For consistency with the satellite data, we determine backgrounds using the remote Pacific. Background HCHO for any latitude in this thesis is calculated by averaging longitudinally (140°W to 160°W) the matching latitudes over the remote Pacific.

3.2.6 Calculation of Emissions

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

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

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

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

not sure I follow.. Isn't the filter only for the isop:HCHO relationship? Why would anthro emissions matter?

Does this paragraph belong in the section about "Calculation of modelled slope"? If yes – make the link to the slope clear at the start. If no, move or add an appropriate section heading.

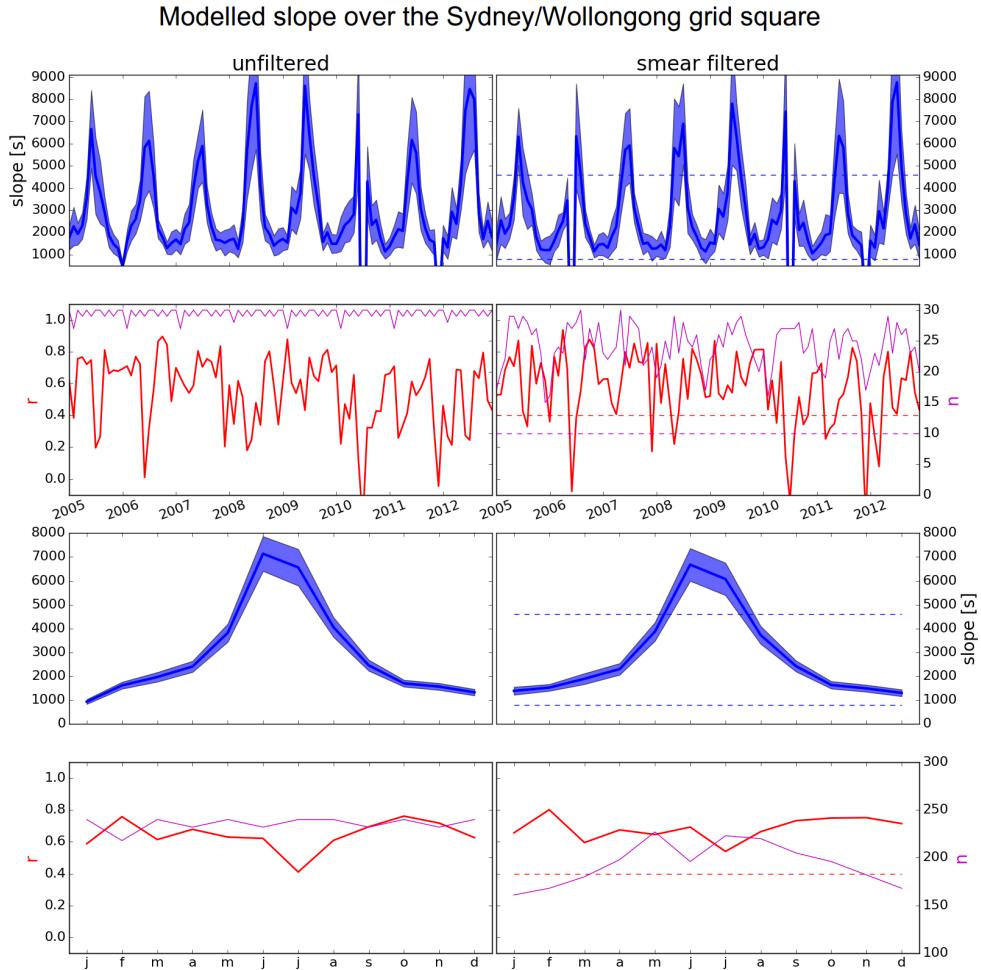


FIGURE 3.3: Row 1: monthly slope along with 95% confidence interval both before (left) and after (right) applying the smear filter for the model grid square containing Sydney over 2005-2012. Limits used in creation of the smear filter are shown with dashed lines. Row 2: regression coefficient and data-point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

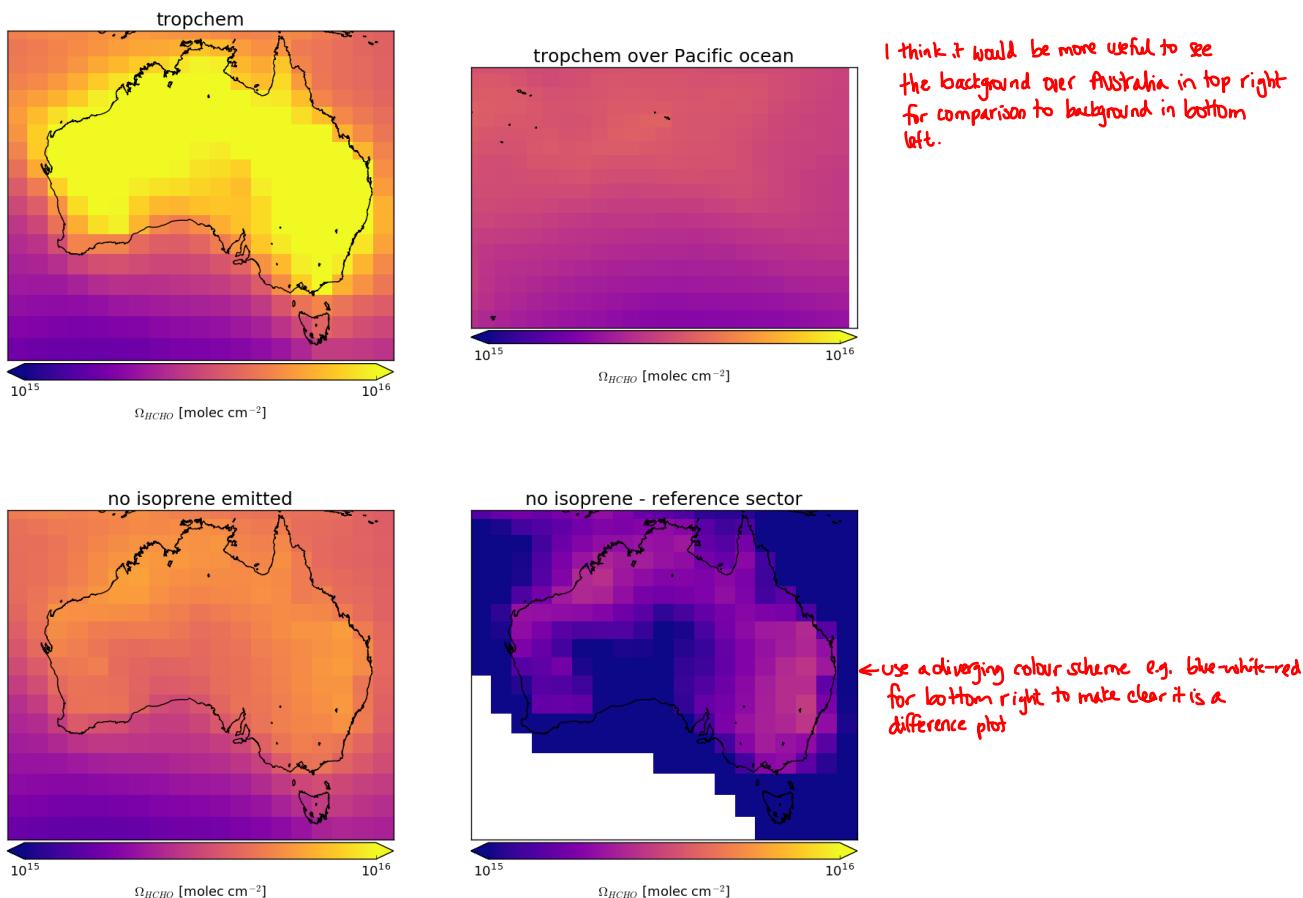


FIGURE 3.4: Top left: Total column HCHO over Australia using the standard tropchem GEOS-Chem simulation. Top right: As top left except over the remote Pacific region at southern mid-latitudes. Bottom left: ~~Background defined~~ As top left except using the no isoprene emissions GEOS-Chem simulation.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

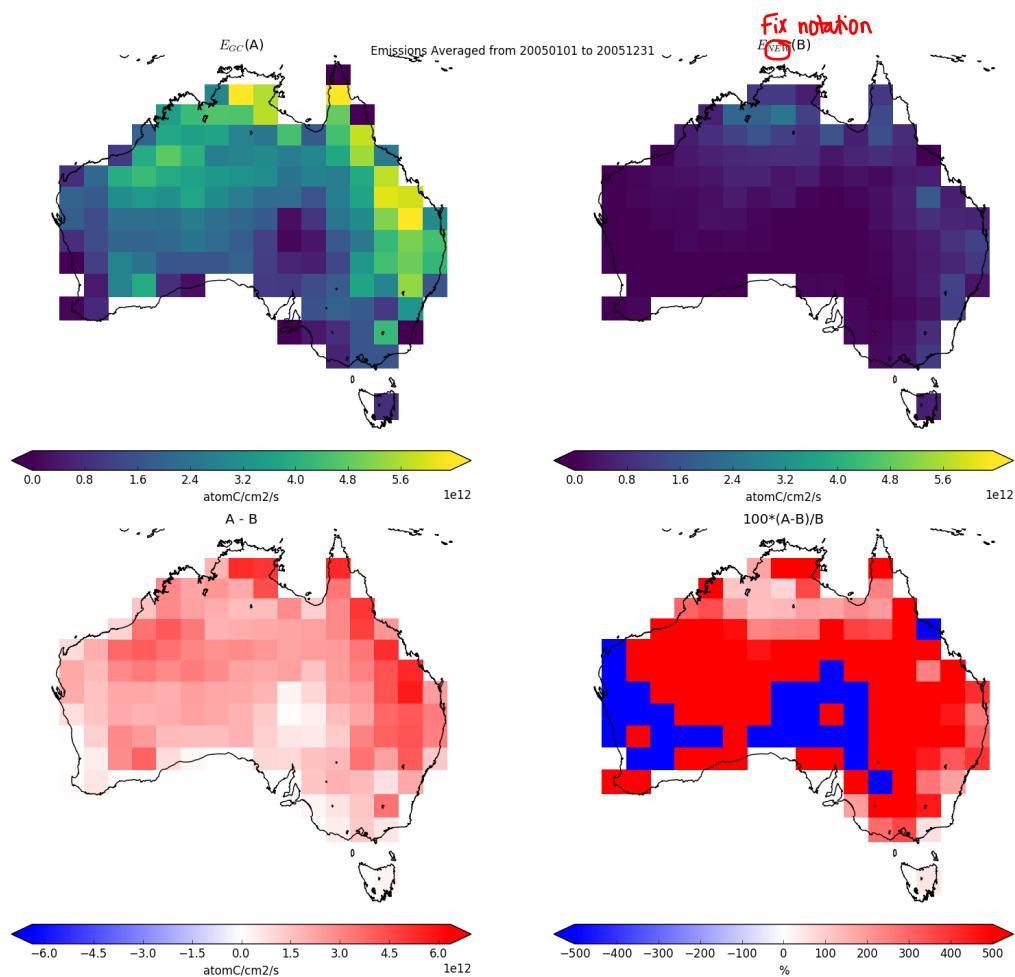


FIGURE 3.5: Top row: isoprene emissions from GEOS-Chem (a priori, left) simulation and top-down (a posteriori, right) calculations averaged over the month of January, 2005. Bottom row: the absolute (left) and relative (right) differences between the two.

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

3.2.7 Accounting for smearing

In high NO_x ($>\sim 1 \text{ ppb}$) environments, isoprene has a lifetime on the order of 30 minutes, and HCHO can be used to map isoprene emissions with spatial resolution from 10-100 km² (Palmer et al. 2003). In low NO_x conditions, isoprene has a longer lifetime (hours) and may form HCHO further from the source area (Liu2017_hpald; Fan and Zhang 2004; Liu et al. 2016b). Over Australia, NO_x levels are generally low and smearing is therefore expected to be important. Smearing limits the horizontal resolution of the linear top-down inversion process, as a finer resolution increases sensitivity to transport. Horizontal transport *smears* the HCHO signal so that its source location would need to be calculated using wind speeds and loss rates (Palmer et al. 2001; Palmer et al. 2003). Smearing is a measure of how much HCHO in a given grid box was produced from isoprene emitted in a different (upwind) grid box. Smearing affects emissions estimates as HCHO enhancements downwind of where precursor emissions occurred lead to misinterpretation of local emissions. Smearing-affected grid squares are filtered out prior to application of Equation 3.7.

Order is off in this paragraph.
You say smearing is a problem
for Australia before you describe
what it is. Re-read and re-order

3.2.7.1 Calculation of smearing

Smearing has been analysed in several publications (e.g., Martin et al. 2003; Palmer et al. 2003; Millet et al. 2006; Stavrakou et al. 2009; Marais et al. 2012; Barkley et al. 2013; Zhu et al. 2014; Wolfe et al. 2016; Surl, Palmer, and Abad 2018), and is often calculated using the method used in this thesis, as first described in Palmer et al. (2003). This involves using two model runs, one of which has isoprene emissions scaled globally by a constant (generally from 0.5 to 2). From Section 3.2.4, Equation 3.7 states that the modelled slope (S) is the yield of HCHO per C of emitted isoprene divided by HCHO loss rate ($S = \frac{Y_{isop}}{k_{HCHO}}$). Using two runs of GEOS-Chem with differing isoprene emissions but otherwise identical we have:

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

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

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

Publication	Source	min. (s)	max. (s)	type ^a	Region
Palmer et al. (2003)		1270	2090	Range	North America ^b
Marais et al. (2012)			4000	Limits	Africa
Barkley et al. (2013) ^c		1300	1800	Limits	South America
Surl, Palmer, and Abad (2018)		2200	4900	Range	India
In This Thesis		800	4600	Limits	Australia

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

b: Slope range for summer only.

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

not change between the two runs:

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

Equations 3.9 may then be combined as follows:

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

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

Potential smearing is masked by checking ~~thresholds~~ a daily modelled value for $\hat{S} \approx Y_{isop}/k_{HCHO}$ against thresholds. By assuming that midday HCHO lifetime ($\tau = 1/k_{HCHO}$) typically falls within 1.5 to 4 hrs (as seen in the USA; e.g., Palmer et al. 2006; Wolfe et al. 2016) and isoprene-to-HCHO yield (HCHO per isoprene carbon emitted) lies within the range 0.2 to 0.4 (scenarios estimated in Palmer et al. (2003)), one can set a simple bound on \hat{S} of $[0.2 \times 1.5, 0.4 \times 4]$ hrs or 1080 to 5760 seconds. As NO_x levels across Australia are relatively low, and lower NO_x levels reduce the prompt yield (Palmer et al. 2003; Wolfe et al. 2016), I reduce the threshold range by 20% and round to the nearest hundred leading for bounding range of 800 to 4600 for \hat{S} . This range strikes a balance between unlikely modelled yields and how much data is lost to filtering. Table 3.1 compares the smearing filter for Australia used in this thesis to typical slopes used in previous work for other regions.

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

see comment next page

and total (or molar) yield. First generation yield refers to the amount of HCHO produced per unit isoprene consumed by initial oxidation. Total yield refers to the time-dependent yield of HCHO over multiple oxidation stages (Wolfe et al. 2016). In this work, yield (Y_{isop}) is not calculated as the slope ($S = Y_{isop}/k_{HCHO}$) can be used instead.

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

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

← not clear why you are discussing it if it is not used.

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

This whole section (between []) doesn't fit here. Maybe somewhere but not here. Very jarring to define \hat{S} after then 'jump to' lifetimes then go back to \hat{S} .

To determine which model grid boxes are affected by smearing, we follow Marais et al. (2012). GEOS-Chem is run with and without isoprene emissions halved, then Equation 3.11 ($\hat{S} = \frac{\Delta \Omega_{HCHO}}{\Delta E_{isop}}$) provides \hat{S} . Here Δ represents the difference (daily 1300-1400 LT) between default and scaled runs. If \hat{S} sits outside the 800-4600 range then we remove that grid square day from both S and subsequent a posteriori calculations. A relatively large change in Ω_{HCHO} compared to local emissions ($\hat{S} > 4600$) suggests HCHO production is from non-local isoprene emissions. Alternatively, a relatively low value of \hat{S} ($\hat{S} < 800$) suggests emissions from the local grid square are being exported before they form HCHO.

3.2.7.2 Sensitivity to smearing

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

no hyphen: data loss

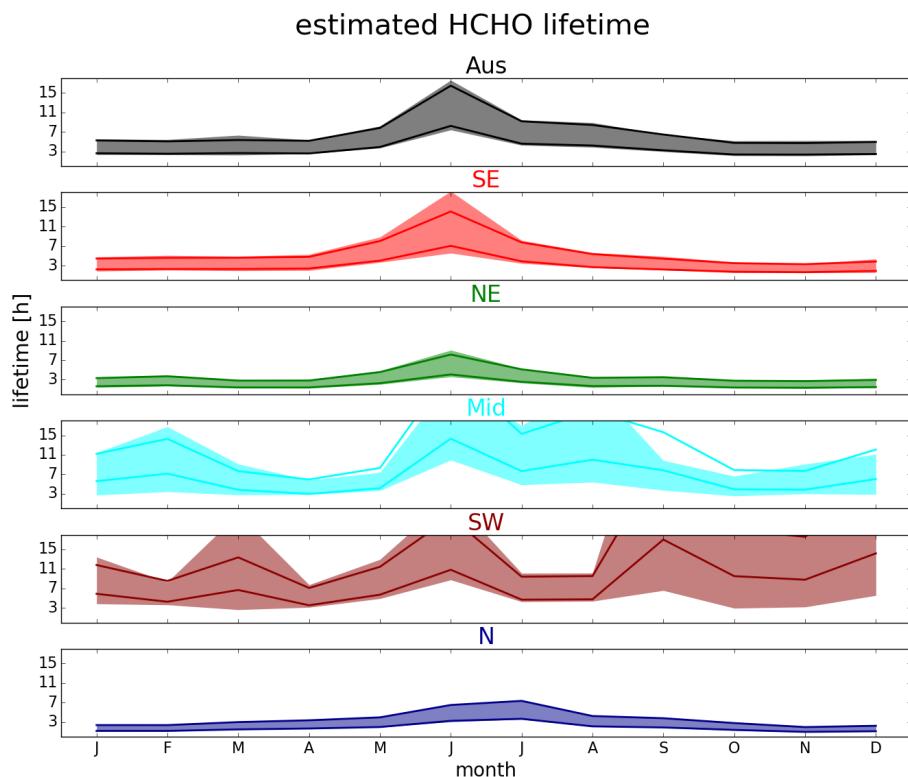


FIGURE 3.6: Monthly area averaged HCHO lifetime (τ in hours), with IQR shaded. Solid lines show lifetime assuming yield is 0.2, and 0.4 (higher and lower lines respectively). Coloured by regions shown in

Figure 3.12.

Weird to reference a later figure.
Maybe just move all the lifetime
shift into results (not really a method anyhow)

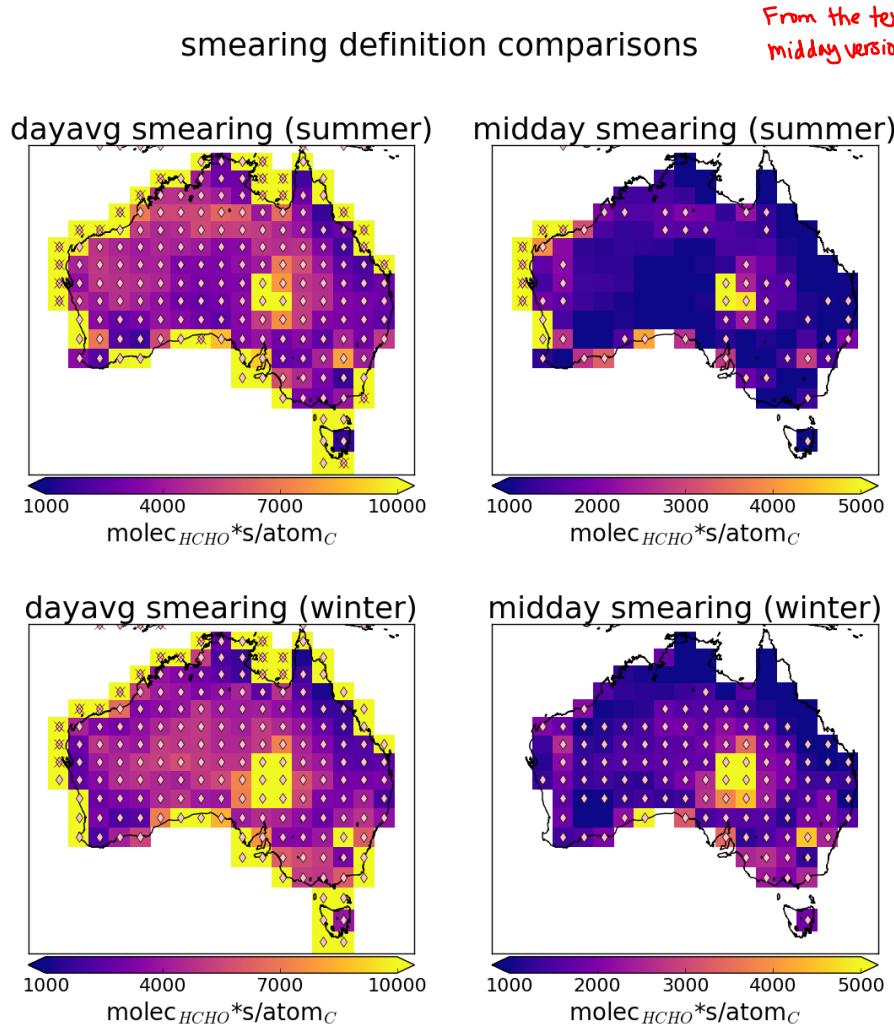


FIGURE 3.7: Seasonally averaged smearing (\hat{S} , see text) in summer (DJF) and winter (JJA) from 2005. Diamonds represent grid squares which have had at least 1 day removed due to the smearing filter, and red crosses show where more than 45 days ($\sim 50\%$) have been removed.

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

(Continued: your figure for τ isn't consistent w/ your assumption in the text (US-4h). overall, I think it does more to raise questions about the validity of the method than anything else. I'd put it in a file somewhere in case you need some of it later, but otherwise drop from this already long chapter.)

3.2.7.3 Smearing length scale - I don't understand the point of this section and suggest you remove it. It is not essential to your methods, and many of the assumptions are hard to justify. (Continued @ left)

The expected horizontal transport (prior to reaction) of a precursor can be calculated using the smearing length (Palmer et al. 2003). The distance travelled (L) downwind by a precursor (i) before forming HCHO can be estimated through:

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

where U is wind-speed. Palmer et al. (2003) further define a smearing length scale $L_{s,i}$ as the distance downwind where most of the precursor is completely transformed into HCHO. Here we use the simplification of $L_{s,isop} \approx \frac{U}{k_{HCHO}}$ as isoprene loss rates (2 h^{-1}) exceed those of HCHO ($0.25-0.67 \text{ h}^{-1}$) by a factor of 3-8. These numbers come from assuming HCHO lifetime of 1.5-4 h, and isoprene lifetime of 30 m [TODO: cite]. Throughout most of the year, over most of Australia, monthly averaged wind speeds do not exceed 20 km h^{-1} (http://www.bom.gov.au/jsp/ncc/climate_averages/wind-velocity/ [accessed Feb., 2019]), although daily wind speeds will be highly variable. This means that a reasonable upper limit for the smearing length scale is $20/0.25 = 80 \text{ km}$ over Australia. Grid boxes used in top-down estimation of isoprene emissions describe rectangles with both side lengths approximately 200 km, so generally I do not expect to be overly impacted by smearing. The estimated loss rate of HCHO in GEOS-Chem is up to three times higher in summer and along the north and eastern regions associated with denser forest regions, when compared against other regions. This is largely due to loss rates being proportional to concentrations, and leads to less smearing sensitivity over areas of high isoprene emissions and HCHO concentrations.

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

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

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

you seem to use "loss rate" for both k and $k[HCHO]$ - be precise.

think you mean rate constant?

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

3.2.7.4 NO_x dependence

NO_x concentration directly affects the fate of VOCs in the atmosphere, influencing HCHO production by isoprene. In low NO_x environments, reported HCHO yields from isoprene are around 0.2 - 0.3 C per C (or 100-150 molar %), while in high NO_x environments this value becomes two to three times higher (Palmer et al. 2003; Wolfe et al. 2016). Some values for HCHO yield from prior literature are shown in Table 3.2.

The effect of NO₂ on smearing can be seen in Figure 3.9. This plot shows how smearing over Australia compares to satellite NO₂, with smearing distributions binned by NO₂ both with and without filtering for smearing. At lower NO₂ levels the smearing is often 2-4 orders of magnitude above the upper threshold. This abruptly decreases at around 5×10^{14} molec cm⁻² NO₂. There is also a higher number of data points below the lower threshold before that same NO₂ ^{value} ~~level~~, showing that transport is a bigger issue at NO₂ < 5×10^{14} molec cm⁻². Due to the coarse resolution of the model, many regions with low NO₂ also have low isoprene emission ~~levels~~, and this filtering does not remove ~~too~~ many data points from biogenic-dominated regions.

3.2.8 Running GEOS-Chem using a posteriori emissions

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

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

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

*built into
the parameterisation
of the WRFAN
model*

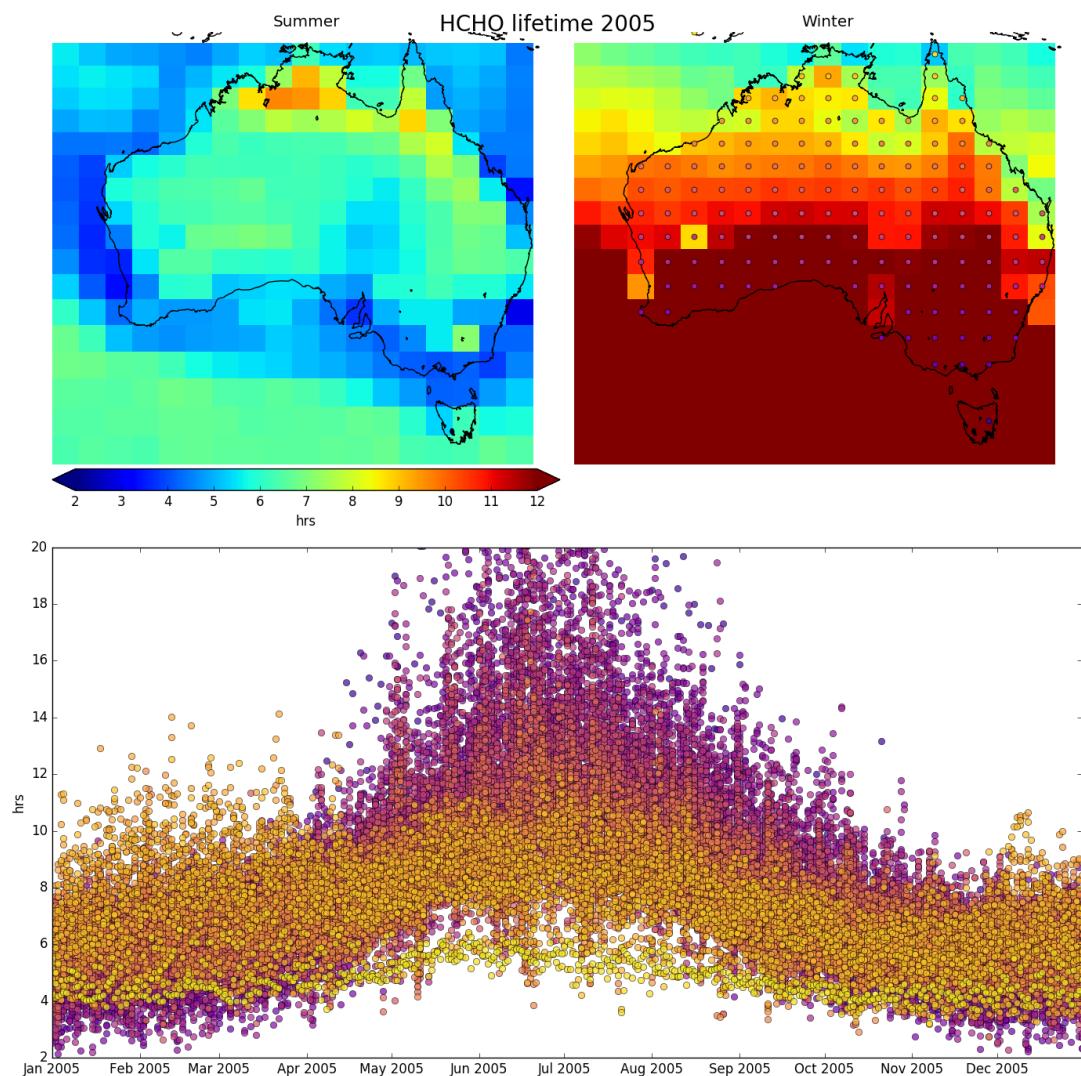


FIGURE 3.8: Top left, right: summer (Jan., Feb.) and winter (JJA) averaged daily surface HCHO lifetime (τ). Bottom panel: τ over the year, coloured by location (see dots in top right panel).

TABLE 3.2: Isoprene to HCHO yields and lifetime.

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

a Atkinson and Arey (2003): Table 2, Yield from Isoprene reaction with OH, two values are from two referenced papers therein.

b Palmer et al. (2003): lifetimes assume [OH] is 1e15 mol cm⁻³.

c (Lee et al. 2006b): Calculated through change in concentration of parent and product linear least squares regression. Estimates assume 20° C conditions.

d Wolfe et al. (2016): “prompt yield”: change in HCHO per change in isop₀. $[isop]_0 = [isop] \exp(k_1[OH]t)$; where k_1 is first order loss rate. Effectively relates HCHO abundance with isoprene emission strength.

e Dufour et al. (2008): One-day yields from oxidation modelled by CHIMERE, using MCM reference scheme.

group all the “High” together.
Any idea what “high” and “low” mean in this context?
I’d put 1 ppbv below High and 0.1 ppb w/ low.

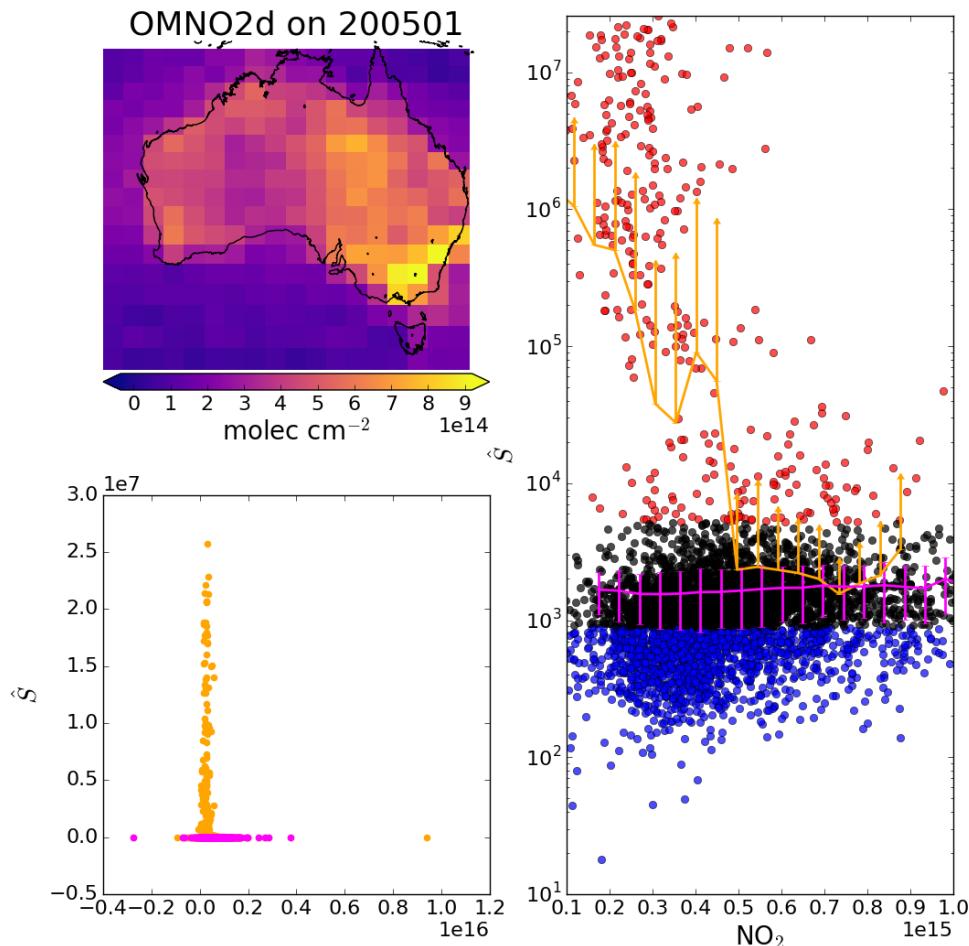


FIGURE 3.9: Top left: OMNO2d tropospheric NO₂ columns (NO₂: molec cm⁻²) averaged into $2^\circ \times 2.5^\circ$ horizontal bins for Jan 2005. Right: Scatter plot of NO₂ against smearing calculations from GEOS-Chem (\hat{S}), with points above and below the smearing threshold range of 900–5200 s coloured red and blue respectively. Points are binned by NO₂ with and without the smearing filter applied (orange and magenta respectively). Overplotted is the mean and standard deviation (error bars) within each bin. Due to the logarithmic Y scale we only show the positive direction of standard deviations for unfiltered data. Bottom left: Daily NO₂ scattered against smearing with (magenta) and without (orange) applying the smearing filter. This plot is a zoomed out version of the right panel.

I don't get anything from bottom left. I'd drop and just state the range of data not shown in right plot. Could probably also ditch the map to simplify the figure.

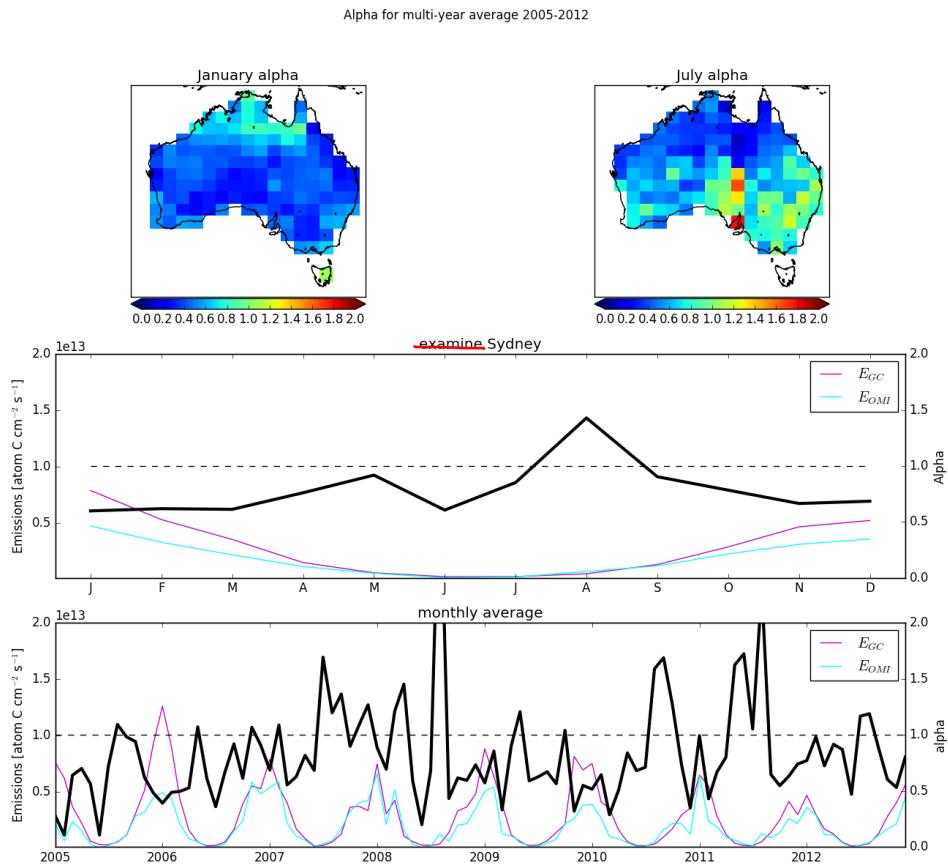


FIGURE 3.10: Row 1: α for the average January (left) and June (right) over 2005-2012. Row 2: E_{GC} (magenta, left axis), E_{OMI} (cyan, left axis), and α (black, right axis) multi-year monthly averages calculated for Sydney. Row 3: Monthly averages of the same terms in Row 2.

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

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

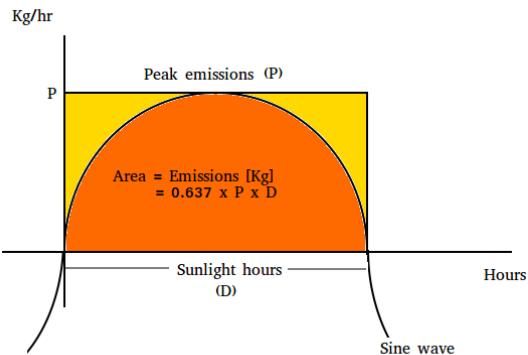


FIGURE 3.11: Total daily isoprene emissions (in kg) is represented by the area under the sine wave.

3.3 Results

~~Cover~~

Australia is roughly $7.7 \times 10^6 \text{ km}^2$, with heterogeneous environmental conditions. The results presented in this section are frequently split into five regions that are differentiated by colour, as shown in Figure 3.12. *Say something about why you chose these regions.*

3.3.1 A posteriori emissions

as described Figure 3.13 shows a priori emissions over Australia along with a posteriori emissions calculated in the prior sections. This figure shows the time series of seasonal area-averaged midday emissions and their absolute differences. The difference between a priori and a posteriori estimates differs between regions, but a seasonal overestimation peaking in summer can be seen in all regions.

Figure 3.14 shows the multi-year seasonal emissions for each region for the a priori and a posteriori emissions side by side. The a priori is approximately twice that of the a posteriori. Absolute differences are highest in spring and summer, when emissions are generally greatest. The exception is the northern region, where the largest overestimation occurs in spring, and large differences are seen in all seasons. Figure 3.15 shows the multi-year monthly mean and inter-quartile range of daily midday isoprene emissions estimates in each region. Months outside of May to August show the most difference between a posteriori and a priori. The most overlap is seen in the south-eastern region, where high summer emissions along with high variance occur in both the a priori and a posteriori. The highest variance is seen in both eastern regions, potentially due to diversity within the regions which include high density cities, large forests, and rural areas. Over the entirety of Australia the seasonal cycle of emissions is shown to be overestimated by the a priori. This overestimate by the a priori may be caused by some mixture of overly high emission factors and high emission sensitivity to temperatures. While most regions show similar overestimates, the northern region follows a different cycle of bias. Northern Australia appears to be overestimated throughout the year, with the lowest bias in early summer. One potential reason is that the wet season (November-April) interferes with satellite measurements

Cite Emmerson 2016. Also may be worth noting soil moisture and airmass Emmerson 2019

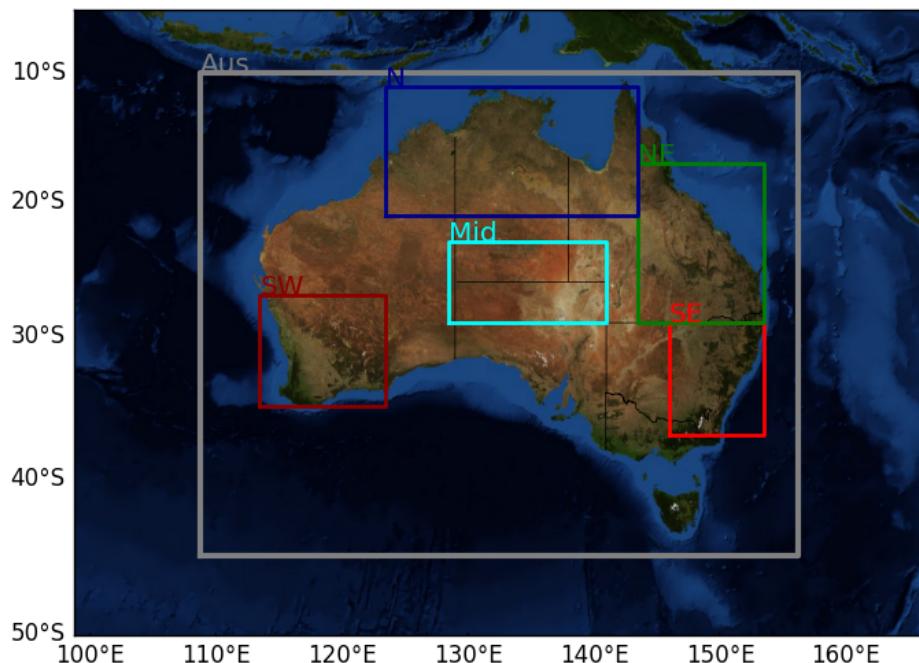


FIGURE 3.12: Sub-regions used in subsequent figures. Australia-wide averages will be black or grey, while averages from within the coloured rectangles will match the colour shown here. *Spell out abbreviations in caption.*

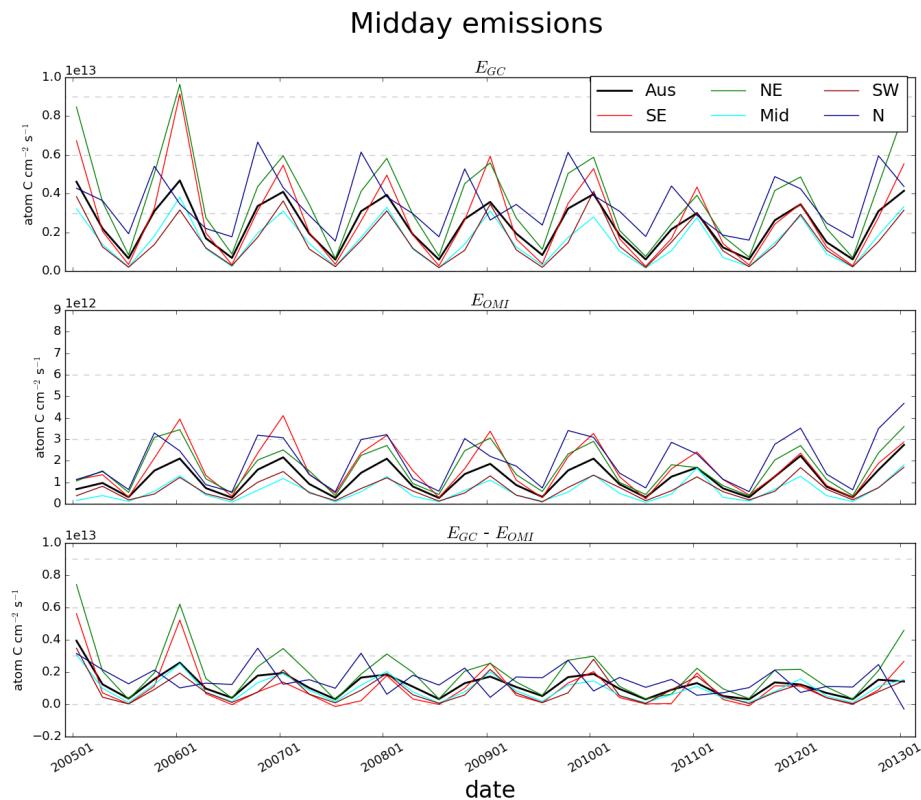
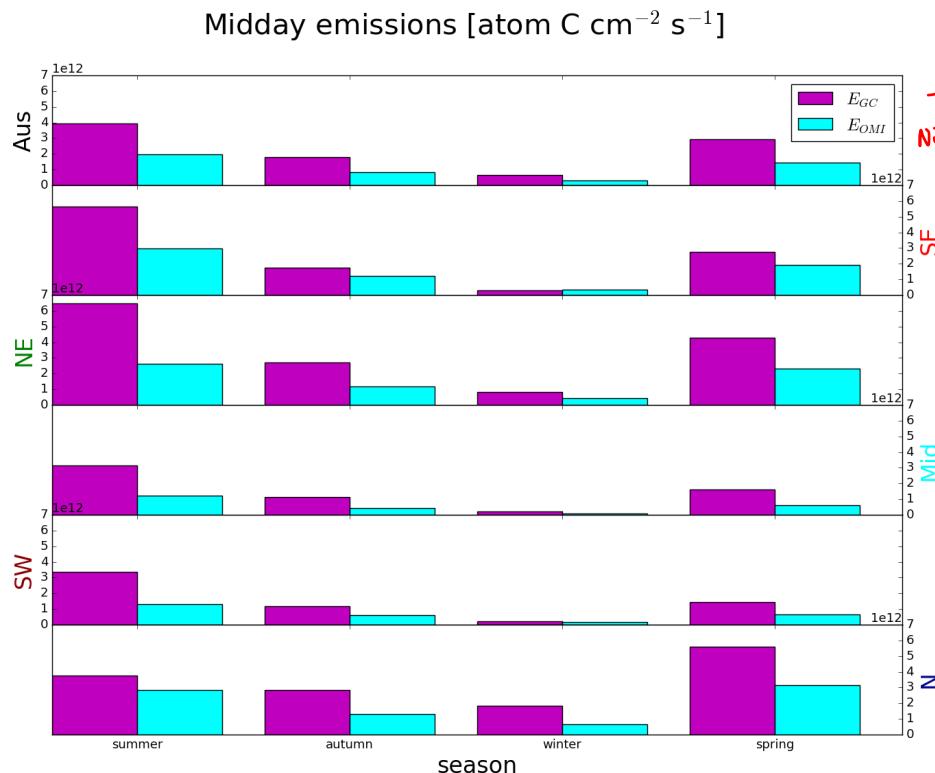


FIGURE 3.13: Row 1: Biogenic emissions of isoprene from GEOS-Chem (a priori, E_{GC}). Row 2: Emissions calculated using the OMI top down inversion (a posteriori, E_{OMI}). Row 3: Absolute differences between the first two rows. Midday emissions are averaged for each season (DJF, MAM, JJA, SON), and colours represent averaged areas from sub-regions shown in Figure 3.12. Grey dashed horizontal bars are added highlighting the scale between rows.



This is a useful figure!
Nice to see the differences simplified like this.

FIGURE 3.14: Regional multi-year seasonal mean a priori emissions (magenta) compared to a posteriori emissions (cyan). TODO: error bars show the regionally averaged uncertainty and the upper bound of the error bars includes potential bias from low satellite HCHO amounts.

due to increased cloud coverage, while also changing the ecosystem's response to sun-
light and temperature. This is backed up by the low summer OMI pixel count (before
filtering) in this region (see Section 3.4.3). Low measurement counts in summer in the
northern region could lead to a low bias in the a posteriori emissions estimate from
the drier regions being over-represented. Further analysis of emissions over northern
forests during the monsoonal seasons is appropriate before drawing conclusions here.

There is only weak correlation between daily estimates of the a priori and a posteriori ($r < 0.2$), and a priori distributions show more variance (figure not shown). This is likely due to the filtering applied to satellite data (e.g., whenever cloud coverage exceeds 40%) which reduces the count and spread of a posteriori emission calculations. Figure 3.16 shows how the distributions of a posteriori emissions compare to a priori emissions in each region during summer months (DJF) with zeros removed from both distributions. This figure also shows the regressions between monthly averages of the same data. In the summer monthly averages, we can see the linear regression coefficient r ranging from 0.2 to 0.81 depending on which region is being compared. The highest correlations between a priori and a posteriori emissions are in the south^{east} ($r = 0.81$) and south^xwestern ($r = 0.79$) regions, followed by the north^{eastern}

is that an analysis you do here? if not, I would rephrase to suggest in situ measurements in both monsoon and non-monsoon seasons are needed.

Comment on what it means that same regions are highly correlated between prior/posterior. What I get from this is that MEGAN is actually capturing the most important sources of variability in these regions but overestimating the magnitude. What might that say about what components are most uncertain?

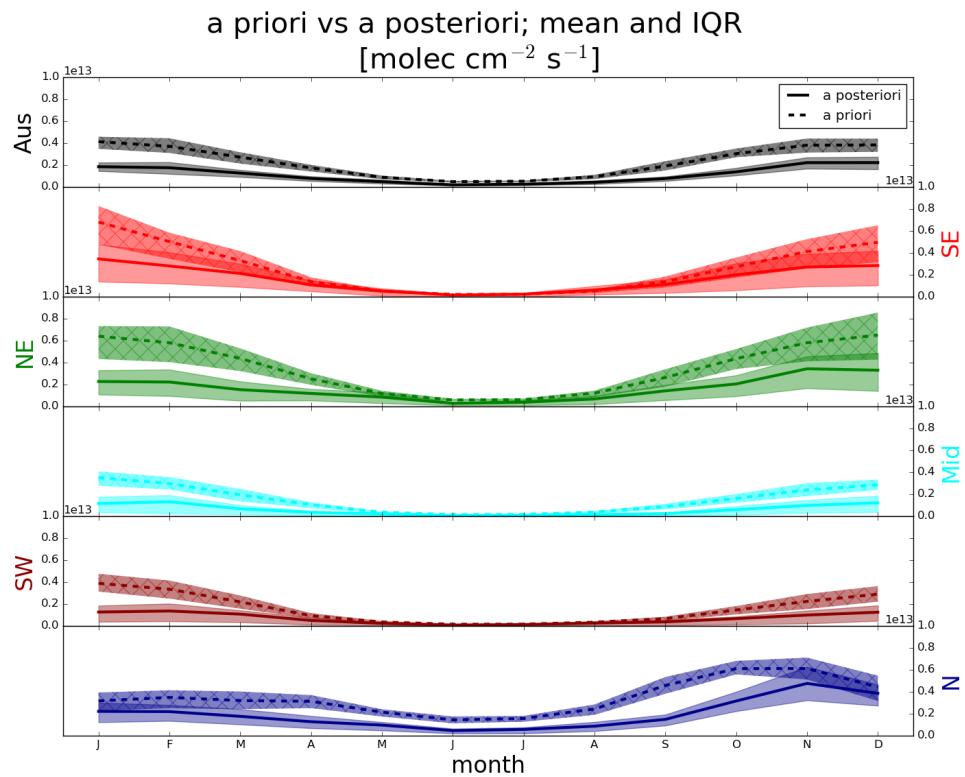


FIGURE 3.15: The multi-year monthly mean (lines) and IQR (shaded) of midday (13:00-14:00 LT) isoprene emissions estimates. A priori emissions are shown by the dashed lines and hatched shaded areas show the IQR. A posteriori emissions are shown using the solid lines, with IQR shown by unhatched shaded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

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

Estimate	Source	Year	notes	Reference
43(2)	bottom-up	2005-2010	(a) from this thesis	a → same for other notes.
19(2)	top-down	2005-2010	(b) from this thesis	
~ 80	bottom-up	1980-2010	(c) Sindelarova et al. (2014)	
26-94	both	2005-2013	(d) Bauwens et al. (2016)	

what is this? if uncertainty, use t

separate lines in table for their bottom-up and top-down

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

$(r = 0.61)$ region. This is likely due to isoprene emissions in these regions being dominated by the biogenic sources (large forests) that the top down emission estimate is based upon. Although the northern region also contains large areas of forest, the correlation is the worst. This could be due to misrepresented forest emissions in the model, exacerbated by unrepresented responses to moisture and poorly modelled environmental stresses in this region which lies within the tropics and undergoes monsoonal weather and intense heat and drought seasons.

is your earlier point about Sampling bias due to clouds relevant here? Also add refs to support your hypotheses.

as described in Section 3.2.3

Figure 3.17 shows the a priori daily emissions cycle from each region, and for comparison the estimated a posteriori emissions cycle over Australia. The conversion of midday a posteriori emissions ($\text{molec cm}^{-2} \text{s}^{-1}$) into Tg yr^{-1} involves integration over an assumed sinusoidal diurnal emission cycle. A priori emissions peak from approximately 11:00 LT to 16:00 LT, while outside these hours there is a non-sinusoidal drop in emissions to below the assumed a posteriori diurnal emission cycle. This means the conversion may be biased by this consistent difference between modelled a priori emission cycles and the assumed a posteriori cycle. This potential bias is not analysed further, and should be relatively small compared to other uncertainties. because it only affects emissions at their daily minima.

When comparing GEOS-Chem (MEGAN) to the a posteriori calculated using our top-down inversion, we find a decrease of from 45.4 Tg yr^{-1} to 19 Tg yr^{-1} (decrease of $\sim 58\%$). Table 3.3 compares annual Australian isoprene emissions from this work to previously published values. Our a posteriori estimate of 19 Tg yr^{-1} suggests isoprene emissions may be lower than any bottom up estimates, and is close to the top-down estimate of Bauwens et al. (2016) of 26 Tg yr^{-1} . Figure 3.18 shows how this decrease is distributed spatially, with E_{GC} and E_{OMI} in Tg yr^{-1} calculated as a multi-year mean. Across all of Australia we see large reductions of total emissions using the new top-down estimate.

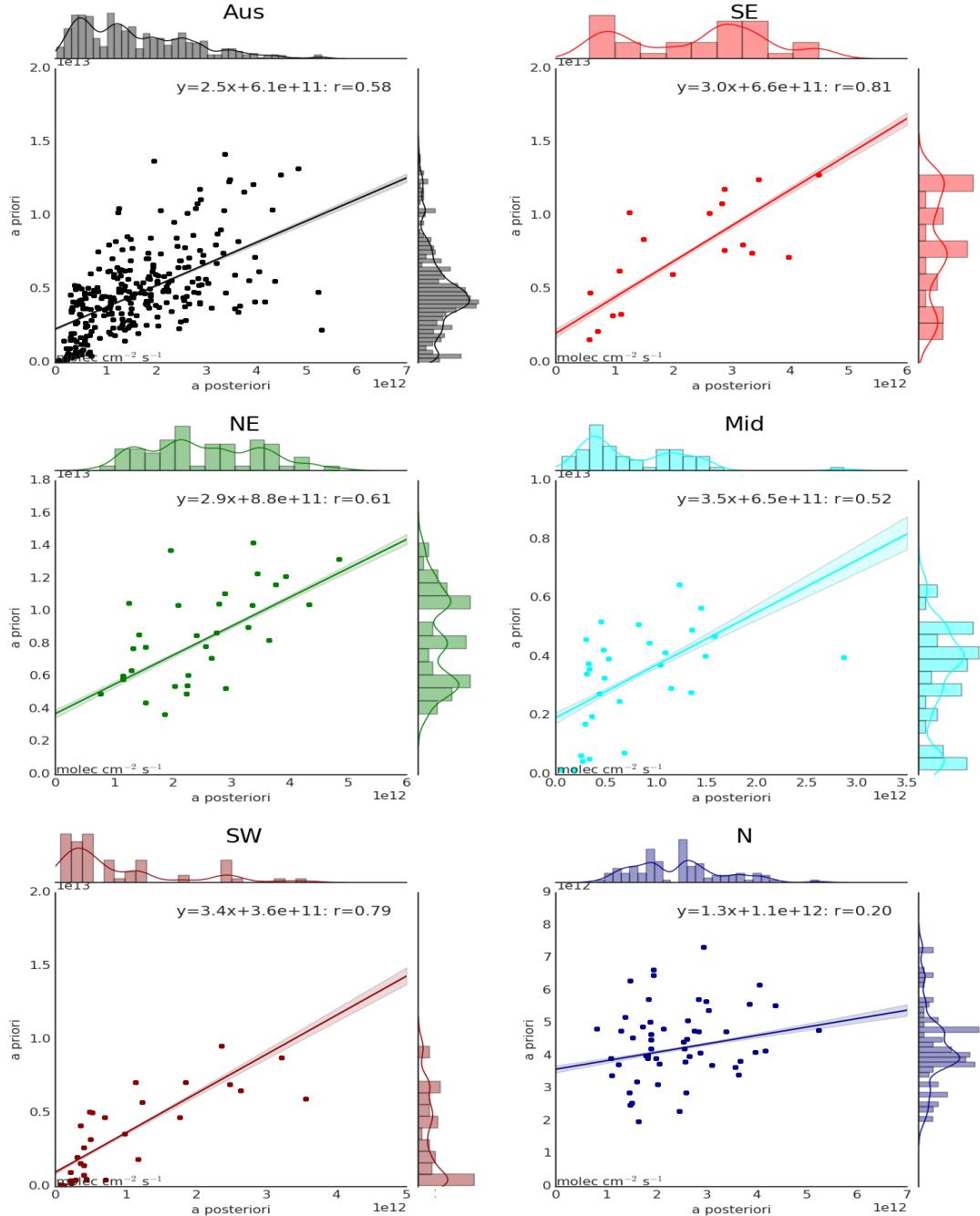


FIGURE 3.16: Scatter plot of a priori emissions against a posteriori using monthly averaged grid squares as regression datapoints. Data points are created using monthly averages (of midday emissions) for each grid box for each month of summer (DJF) within each region shown. Multiple years of data are used, meaning if a region has 10 grid boxes, the 8 years of data will add up to $10 \text{ boxes} \times 3 \text{ months} \times 8 \text{ years} = 240$ data points minus filtered and zero emission squares. Plots are coloured by regions matching those shown in Figure 3.12. The linear best fit regression is inset into each plot along with the line equation and regression coefficient. *Add sentence explaining distributions.*

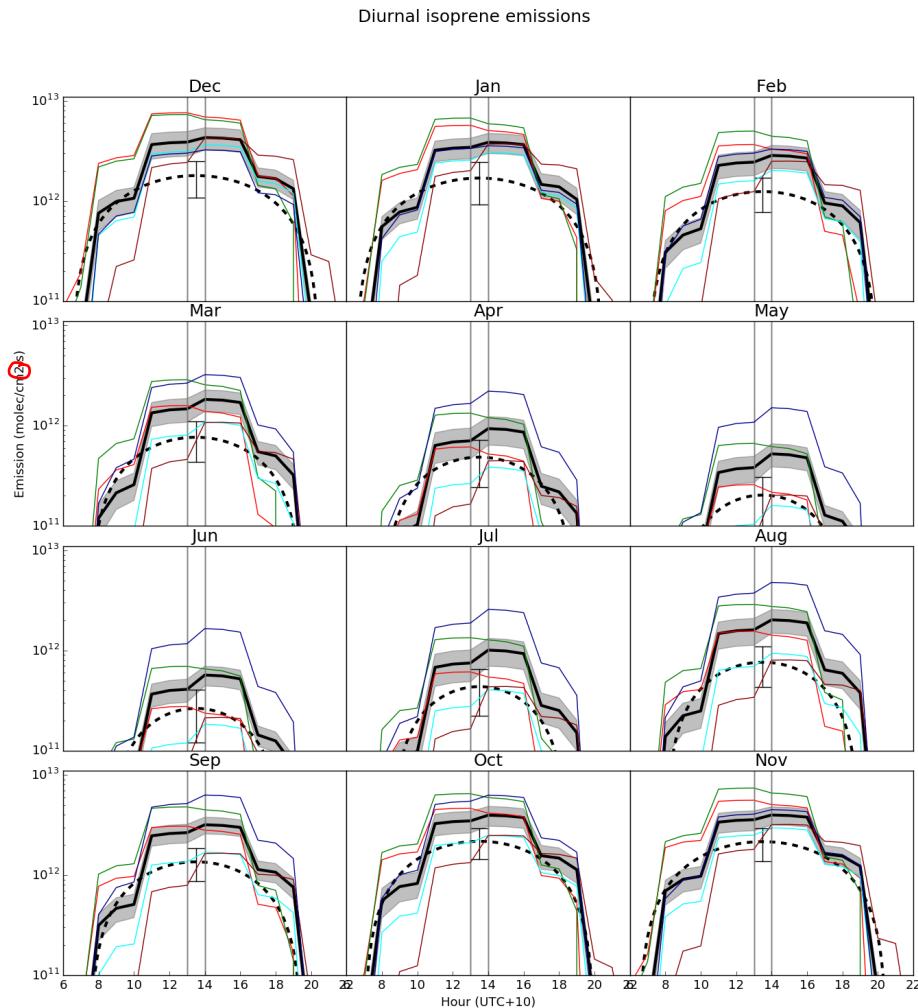


FIGURE 3.17: The diurnal cycle of GEOS-Chem a priori emissions (solid line) averaged by month into hourly bins over from 2005 to 2013 are shown against top-down a posteriori (dashed line) emissions. Standard deviations for the monthly average are shaded for the a priori, and shown with error bars at 13:30 LT for the a posteriori. The coloured lines show the a priori averaged over different regions (see Figure 3.12). Vertical bars are added at hours 13:00 LT and 14:00 LT, the overpass time for OMI. Top down emissions shown here are based on monthly mid-day emissions being the peak of a sine wave which drops to zero after and before daylight hours. (See section 3.2.8).

I find the vertical bars distracting rather than helpful, esp. as you already show the satellite overpass time with the error bars at 13:30. I would remove vertical bars and modify the caption so it explicitly says the error bars at 13:30 are shown at time of OMI overpass.

why a priori? If you are only going to show one, a posteriori would make more sense (results chapter..)

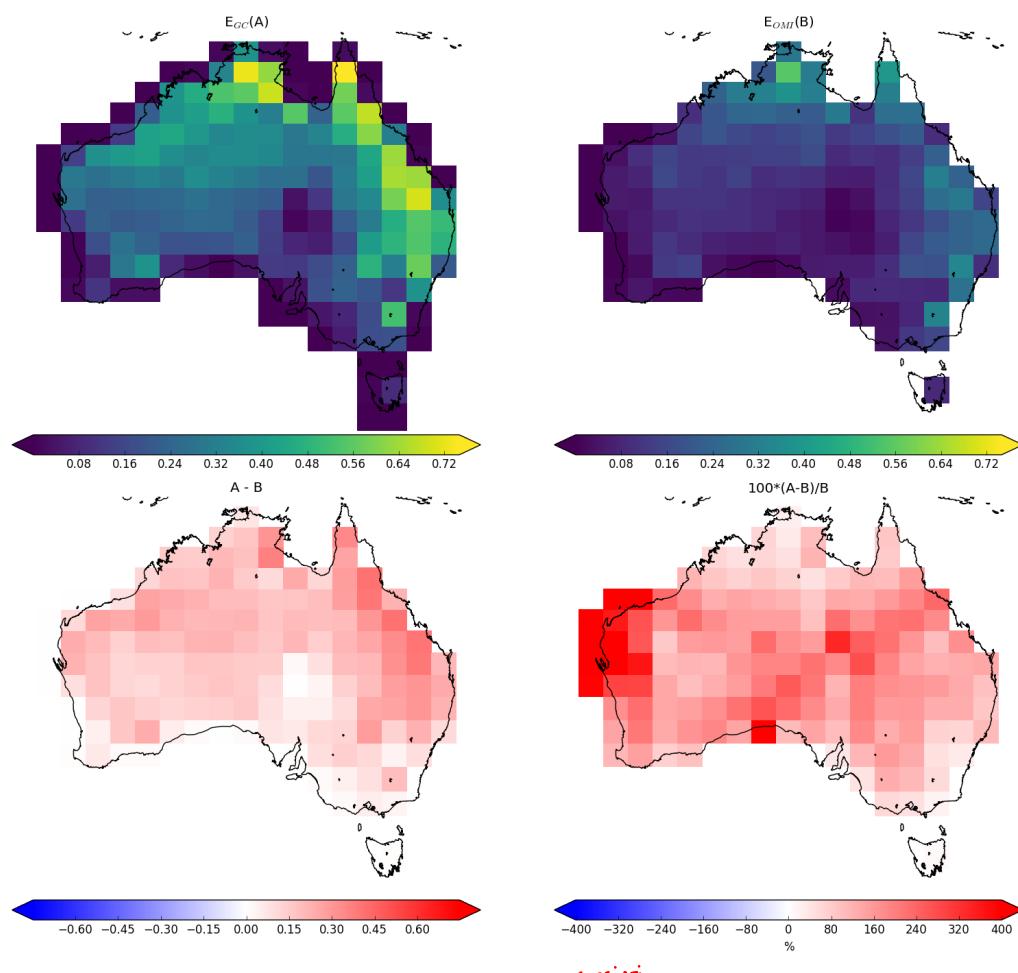


FIGURE 3.18: Top row: multi-year mean emissions in Tg yr^{-1} from E_{GC} (GEOS-Chem; running MEGAN) and E_{OMI} (top-down emissions) respectively. E_{OMI} uses an assumed sinusoidal daily cycle, with day-light hours prescribed for each month: see Section 3.2.6). Bottom left and right show the absolute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

*check section -
I thought it was 3.2.8*

3.3.2 Modelled impacts of reduced isoprene emissions

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

3.3.2.1 Implications for HCHO

After running GEOS-Chem with isoprene emission scaled to match the a posteriori, we first compare simulated Ω_{GC} and Ω_{GC}^α to Ω_{OMI} over January in 2005 (TODO: update to use all summers and winters) in Figure 3.19. In every region, Ω_{GC}^α is closer to Ω_{OMI} with biases decreasing from 50-120% down to 30-50%. These drive the creation of Ω_{GC}^α , and as we expect the relationship is much improved. The remaining differences are most likely driven by filtering and temporal averaging of the applied scaling factor α . Variances in satellite HCHO vary from X to Y %, with more variance in summer. Monthly variance of Ω_{GC} , Ω_{GC}^α , and Ω_{OMI} is shown in Figure TODO. Scaling isoprene emissions TODO has what affect on variances compared to those of OMI.

Wollongong FTIR measurements (see Section 2.2.3.4) provide total column amounts TODO: get these and check process. This is the only non-satellite long-term measurement record of total column HCHO available in Australia and we use it here to examine trends and seasonality. The time series for HCHO is shown in figure TODO, along with GEOS-Chem output before and after updating isoprene emissions. Modelled HCHO in this figure only shows the grid square containing Wollongong, which is the same grid square that contains Sydney and a large amount of rural and forested area. This makes direct comparison between GEOS-Chem output and the FTIR suffer from representational error. TODO: discuss plot here

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

Figure TODO shows how Wollongong FTIR HCHO profiles compare against GEOS-Chem modelled HCHO before and after isoprene scaling is performed.

Figure TODO FTIR total columns against geos chem and geos chem scaled

3.3.2.2 Implications for ozone

Isoprene oxidation and can eventually lead to ozone formation, especially when isoprene enriched air masses mix with polluted urban air masses that contain high NO_x concentrations. Figure 5.3 shows surface level (up to ~150 m altitude) ozone concentrations over 2005 before and after scaling modelled isoprene emissions. Reducing isoprene emissions lowers surface ozone concentrations by TODO: XX to YY % in summer, and XX to YY % in winter. While the overall decrease in surface ozone is clear, there is no direct correlation between reduced emissions and changes in surface ozone concentrations found in either daily or monthly averaged model output. This suggests that changes in isoprene emissions affect ozone in non-local grid squares: i.e. ozone reductions are occurring downwind of emissions.

Note that this is not an independent validation as

refer to Fig. 3.21

?? Figure in
a different
chapter?

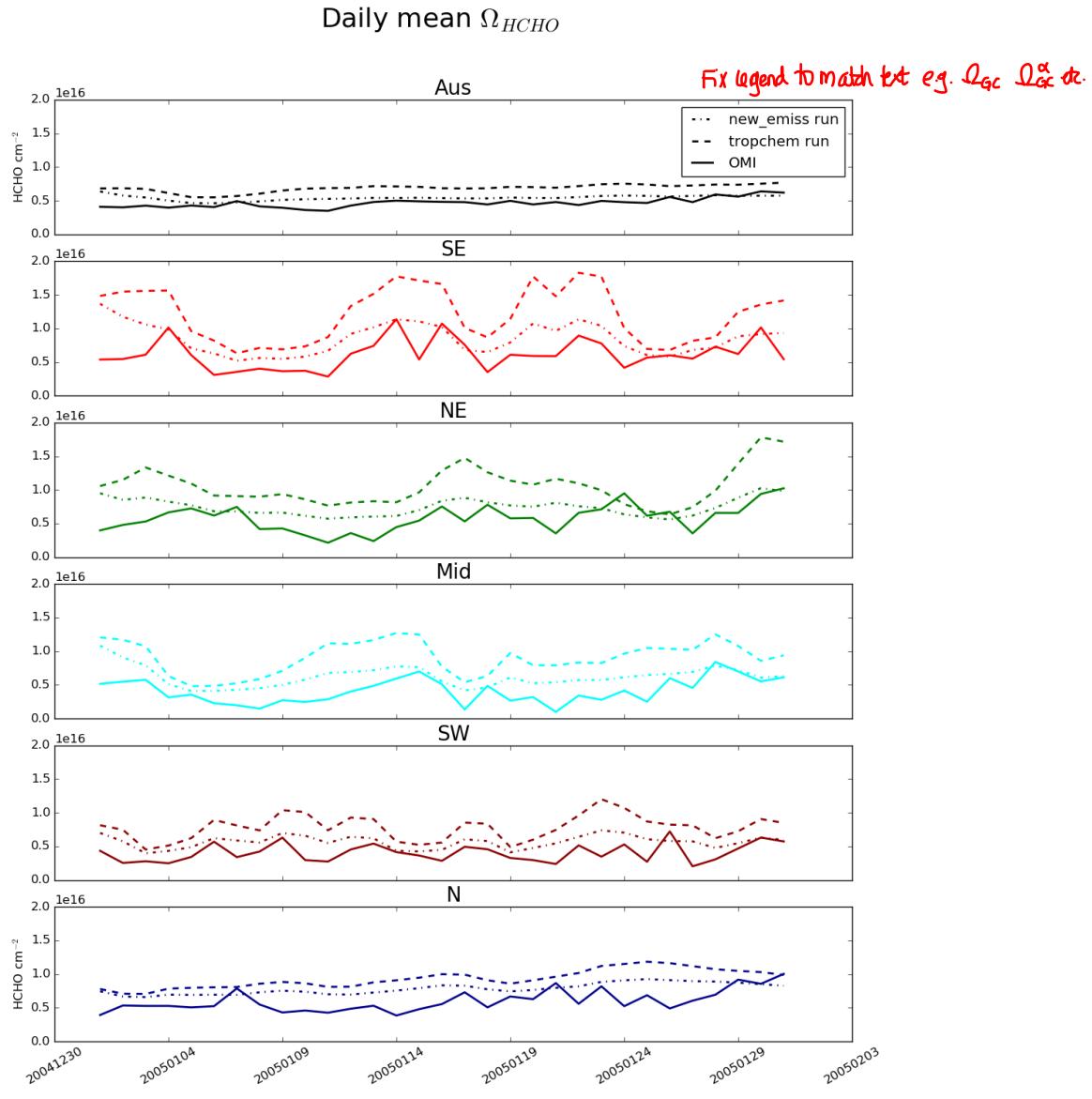


FIGURE 3.19: Daily mean total column HCHO amounts from GEOS-Chem with (new emissions run) and without (tropchem run) a posteriori scaled isoprene emissions, along with the original OMI HCHO columns. Each row shows the average over regions in Figure 3.12.

Downwind effects of isoprene emission are most likely to affect suburban fringes (e.g. western Sydney) of many Australian cities which are surrounded by vegetation. Outside of densely populated regions, Australia is likely to be NO_x limited and changes in VOC emissions will have less direct effects on ozone production. This means that cities may have higher sensitivity to changes in modelled isoprene emissions than shown here, since we are examining large area averages which are mostly non-urban. TODO: Notes about ozone exposure from State of the Environment report 2016.

TODO: Figure shows modelled surface ozone concentrations and their differences between model runs over an average summer (DJF).

would be good to add some refs to this paragraph and/or previous. There is a paper by Millet et al (maybe 2013ish?) that talks about transported isoprene having a delayed effect on urban ozone.

3.3.2.3 Trends

Figure 3.20 shows monthly deseasonalised a priori and a posteriori midday emission anomalies for each region. First the emissions are spatially averaged within each region to form a daily time series of midday emission rates. This is averaged into monthly data, and then the multiyear monthly mean is subtracted to form the anomaly time series. Any anomaly greater than three standard deviations from the mean is removed (crosses in Figure 3.20). An ordinary least squares linear regression is then performed to look for any significant trend. A trend is considered significant if the p-value from a Wald test (equivalent to a t-test) is less than 0.05. The same process is repeated for surface concentrations of isoprene, ozone, HCHO, and NO_x , with results summarised in Table 3.4. The midday surface isoprene concentrations show a small decline outside of the southwest region over the 8 year period from 2005-2012. This decline reduces in scale by approximately a factor of 2-3 after emissions are scaled down. None of the species shows any significant changes in trend due to scaling isoprene emissions. Some trends are no longer significant in surface HCHO, however none of the trends change sign. Since the scaling factor α is applied to each grid square seasonally and not changing each year, changing trends are not expected.

Given this result, the fact you don't expect a trend change anyway as you state and the fact that there are large NO_x/O_3 trends shown that have nothing to do with isoprene, I would:

- (1) Remove $\text{NO}_x, \text{O}_3, \text{HCHO}$ from table and text
- (2) move this to section 3.3.1

3.3.3 Comparison with in situ measurements

Comparison between ground-based measurements and large ($2^\circ \times 2.5^\circ$) averaged grid squares suffers from representation error. Figure 3.21 shows the SPS and MUMBA measurement sites, along with the extent of the $2^\circ \times 2.5^\circ$ relevant GEOS-Chem grid box. The grid box is the area over which GEOS-Chem outputs are averaged, a rectangle with edge lengths of roughly 200 km 2 . The urban footprint of Sydney and Wollongong can be seen, along with some ocean, forest, and rural regions, which will all affect the model output and other calculations averaged here. Due to high uncertainty in components of the top-down emissions estimate, temporal resolution is also limited. MUMBA, SPS1 and SPS2 each provide relatively on the order of one month of hourly or daily datapoints, which are compared in this section against surface level concentrations from GEOS-Chem before and after scaling the biogenic emissions.

Figure 3.22 shows GEOS-Chem output in the grid square containing Sydney overlaid on SPS measurement data. Superficially the comparison is not too bad between these two datasets. TODO: mean midday quantities for matching months. This figure shows at a glance how the measurements compare to modelled data, and also the

how do they compare?

Campaigns are introduced in earlier chapter, right? Reference it here.

Meaningless - quantity.

TABLE 3.4: Yearly trend in surface amounts before and after scaling isoprene emissions.

Region	a priori	a posteriori
Isoprene (ppb_{Cy})		
Aus	-.04	-.02
SE	-.15	-.07
NE	-.13	-.04
Mid	-.07	-.02
SW	-.01	-.01
N	-.06	-.02
Formaldehyde (ppb)		
Aus	-.01	-.00
SE	-.03	-.02
NE	-.02	-.01
Mid	-.01	-.00
SW	-.00	-.00
N	-.01	-.00
Ozone (ppb)		
Aus	.05	.07
SE	-.34	-.28
NE	-.08	-.05
Mid	-.05	-.04
SW	.15	.17
N	.18	.17
NO_x (pptv)		
Aus	-.48	-.44
SE	-4.25	-4.08
NE	-2.00	-1.82
Mid	-1.08	-.72
SW	-.22	-.35
N	2.3	1.83

Statistically significant (two sided test with $\alpha = 0.1$) trends are bolded.

Units are ppbv Carbon for isoprene, pptv for NO_x, ppbv for others, all rounded to two decimal places.

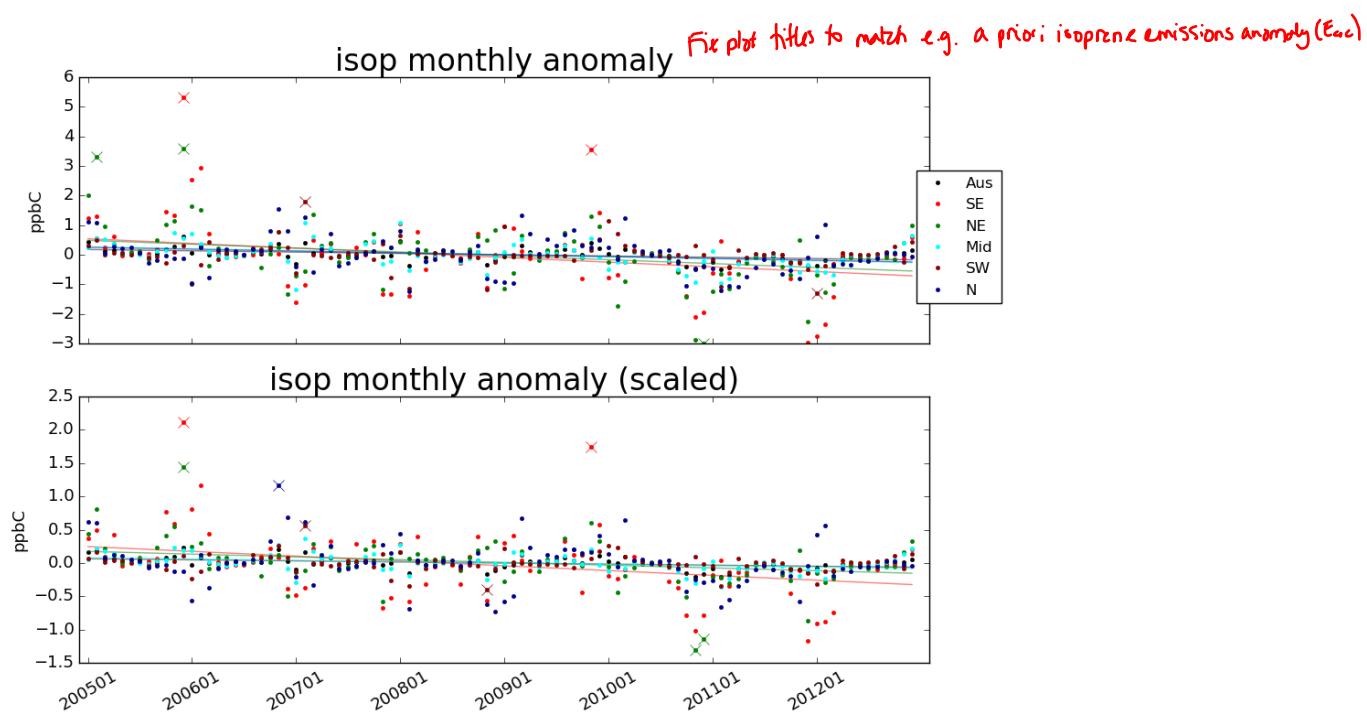


FIGURE 3.20: A priori (row 1) and a posteriori (row 2) emissions anomaly from multiyear monthly mean, split by region (see Figure 3.12).



FIGURE 3.21: grid box at $2^\circ \times 2.5^\circ$ containing SPS and MUMBA campaign data. *and Wollongong FTIR.*

Separate plots for SPS1 and SPS2 for HCHO, no need for top row.

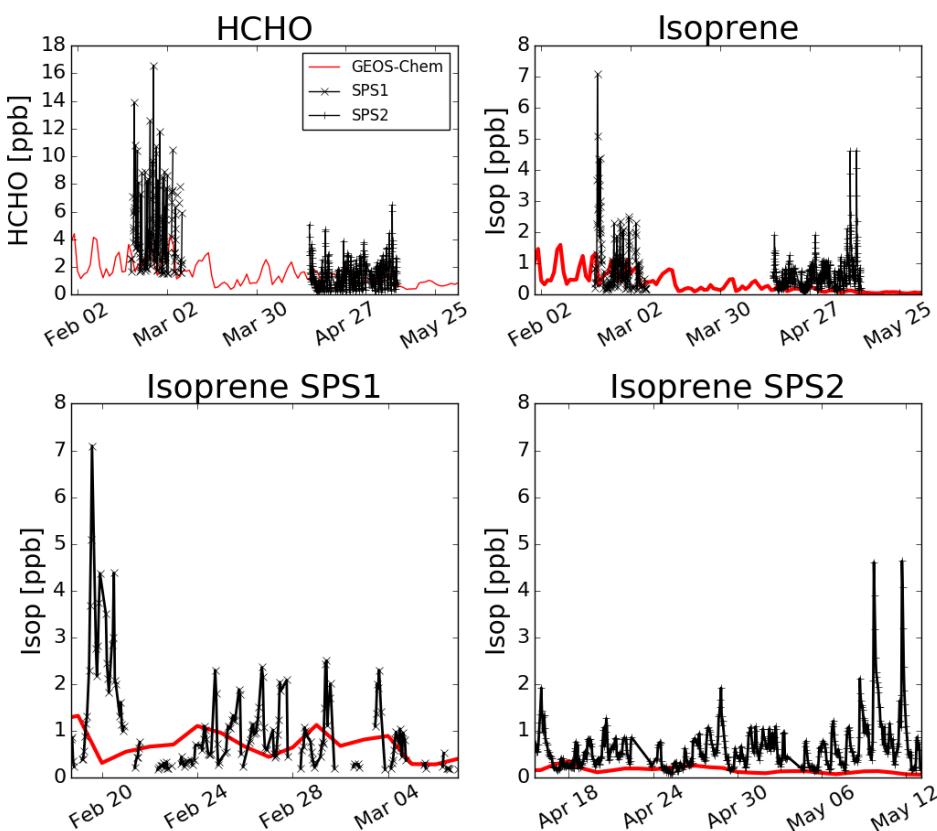


FIGURE 3.22: Comparison between GEOS-Chem HCHO concentrations in the grid square containing Sydney for the duration of the SPS 1 and 2 campaigns
 in surface air
 ↑ MUMBA?

limited nature of the temporal and spatial scale of available measurements. TODO: Add scaled output to plot, and discuss The SPS data is point-source and taken during the daytime when isoprene is higher, so it is very likely that GEOS-Chem HCHO and isoprene output is in fact too high since the daily average should not match the peak of the measurements.
 so you are not comparing hourly output? In that case show obs as daily averages, too.

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

3.4 Uncertainty

This section identifies and quantifies the overall uncertainties of calculating isoprene emissions using OMI HCHO observations and the GEOS-Chem model in the top-down method used in this chapter. However, these uncertainties lack verification

TABLE 3.5: Relative uncertainty estimates.

Region	ΔE_{OMI}	Summer			Winter		
		$\Delta\Omega_{OMI}$	ΔS		ΔE_{OMI}	$\Delta\Omega_{OMI}$	ΔS
Aus	60%	20%	40%		100%	100%	40%
SE	TODO: properly fill in this table						
NE							
...							

against measurements. Even as the top-down inversion performed in this chapter attempts to ~~work around~~ circumvent the lack of measurements over Australia, it suffers from the lack of independent observations against which it can be verified.

The major source of uncertainty throughout the year comes from uncertainty in the modelled yield slope S (see Section 3.4.2); however, in winter, uncertainty from satellite column calculations becomes dominant (see Section 3.4.3). Monthly calculated slope uncertainty mostly lies within 30% to 50%, and this is not reduced through averaging. Uncertainty from each OMI satellite measurement is relatively large ($> 100\%$); however, averaging thousands of pixels in each grid square greatly reduces the monthly uncertainty. Uncertainty in satellite HCHO is seasonally dependent, with better signal during the summer. Reliable OMI measurements are less frequent (leading to higher uncertainty) at high solar zenith angles, which ~~worsen~~ increase with latitude and during winter. Table 3.5 shows the estimated uncertainty calculated in this work in summer and winter over each region described by Figure 3.12. The assumptions and calculations made to determine uncertainties in the top-down estimate (ΔE_{OMI}), the satellite column ($\Delta\Omega_{OMI}$) and the slope (ΔS) are described in the following subsections. *[and the results summarised briefly here.]*

*MOVE this part
(between []) to start of
paragraph.*

3.4.1 Top down emissions uncertainty

Sources of uncertainty

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

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

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

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

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

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

Which leads to the uncertainty estimation for ~~our~~^{the} a posteriori emissions as follows

$$\Phi \equiv \Omega_{OMI} - \Omega_{OMI,0}$$

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

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

ΔE_{OMI} is calculated using the uncertainty in underlying terms: ΔS , $\Delta\Omega_{OMI}$, and $\Delta\Omega_{OMI,0}$. For ΔS ($\Omega_{GC} = S \times E_{GC} + \Omega_{OMI,0}$ from equation 3.7) I use variance in the monthly linear regression of modelled isoprene emissions and column HCHO, shown in Section 3.4.2. For $\Delta\Omega_{OMI}$ and $\Delta\Omega_{OMI,0}$, uncertainty comes from instrument fitting uncertainty, modelled AMF uncertainty, and uncertainty in the background correction terms, which are described and calculated in Section 3.4.3.

Figure 3.23 shows relative uncertainty over each region of Australia in monthly bins. Uncertainty in the southern regions increases between May and July due to increased error in the satellite measurements. Northern regions are impacted less by the seasonal satellite error, which generally is 10 – 20% lower than the error from S . Figure 3.24 shows the spatial distribution of relative uncertainty in the a posteriori in summer and winter. Here the effects of satellite uncertainty at higher latitudes (especially in winter) are shown as a general increase with latitude is apparent.

3.4.2 Model Uncertainty

E_{OMI} depends partly on the product it is trying to improve, as modelled yield is based on GEOS-Chem run with MEGAN emissions. The uncertainty in the RMA regression slope between model HCHO (Ω_{GC}) and emissions (E_{GC}) is used to estimate ΔS in Equation 3.16. Here I use the ratio of the upper bound of the 95% confidence interval (CI_{UB}) over S to represent the relative uncertainty.

$$\frac{\Delta S}{S} = \frac{CI_{UB}}{S} - 1 \quad (3.17)$$

The confidence interval for each month is based on the covariance matrix between Ω_{GC} and E_{GC} , and the critical t-statistic considering n to be the days in the month and α to be 0.025. For example, if the interval upper bound is 30% higher than the slope, relative uncertainty is set to 0.3 (or 30%).

This is a simple method of approximating the uncertainty of this term, only accounting for monthly uncertainty of the slope calculation. It does not take into account uncertainty in the underlying model, nor uncertainties arising from temporal or spatial resolution issues, which would be difficult to quantify.. Figure 3.25 shows the relative uncertainty in S over Australia, and for each region. There is little discernible seasonality to the relative error in S , which generally ranges from 0.3 – 0.4 (30 – 40%). For comparison, Palmer et al. (2006) found $\frac{\Delta S}{S}$ to be 30%, after comparison with another chemical model and in situ measurements. To improve understanding of uncertainty in S would require further analysis of GEOS-Chem yield over Australia, including how it responds to environmental and meteorological parameters, and how representative this modelled quantity is compared to measurements.

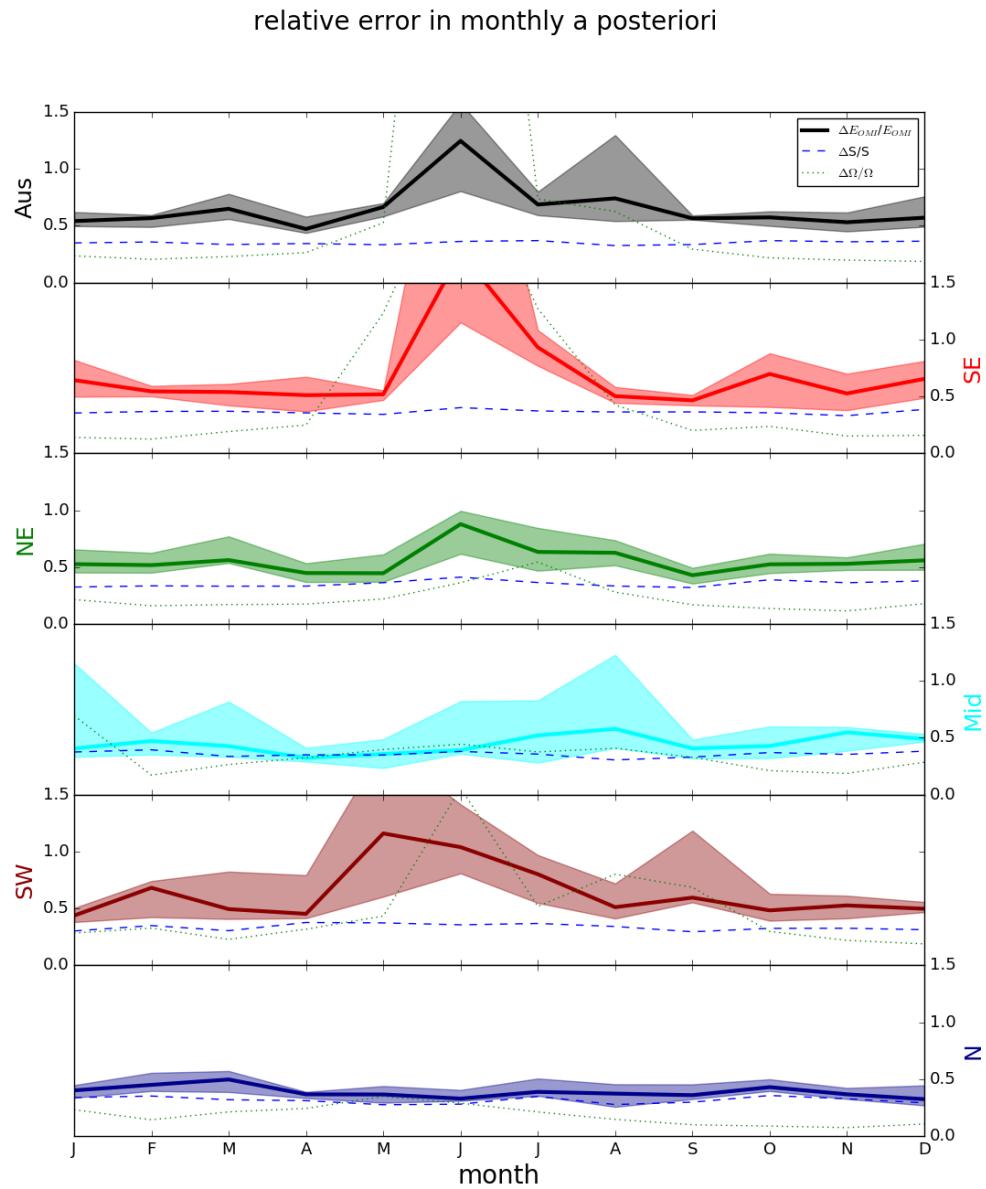


FIGURE 3.23: Median and inter-quartile range of multi-year monthly relative uncertainty in the a posteriori. Median relative uncertainty in S and Ω are added as dashed and dotted lines respectively.

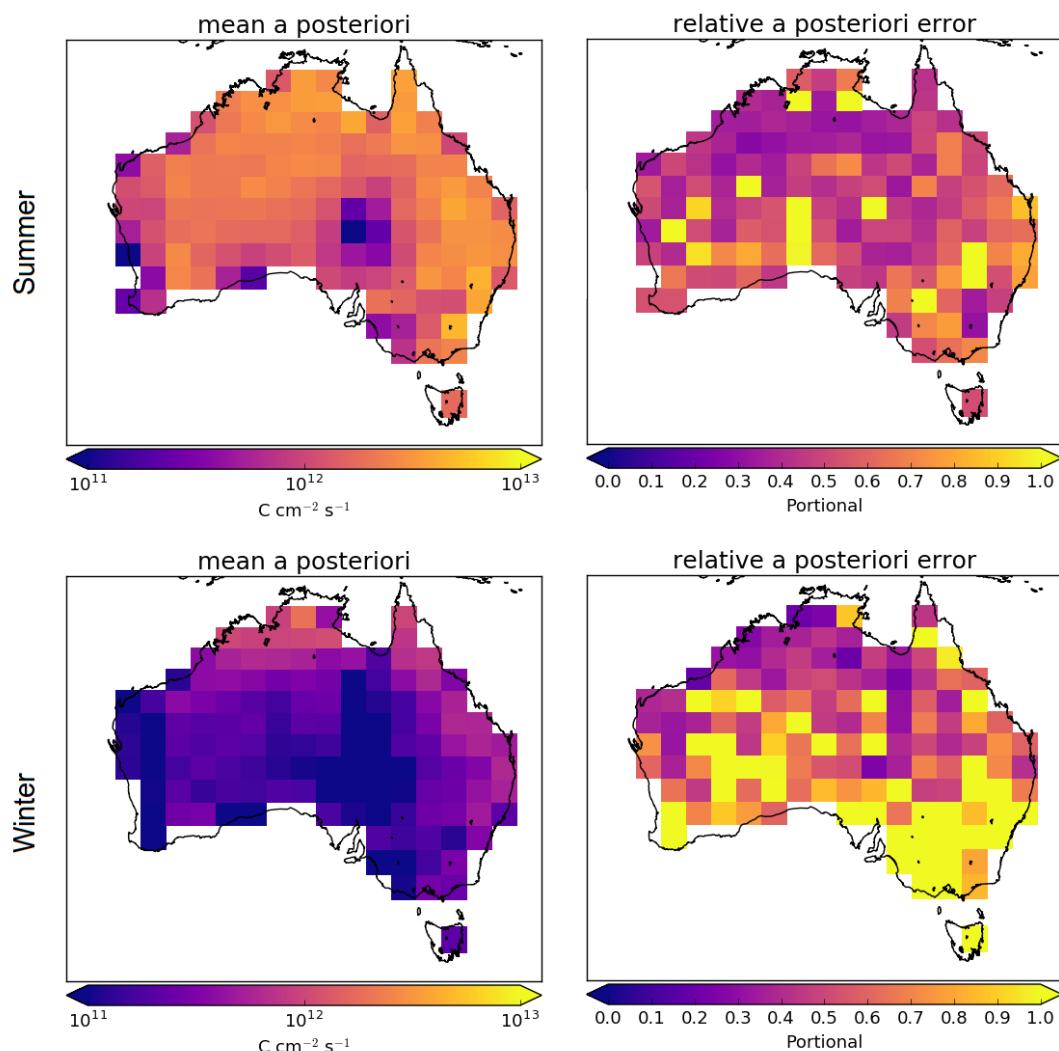


FIGURE 3.24: Summer (DJF, top row) and winter (JJA, bottom row) a posteriori emissions (left column) and ~~its~~ relative error (right column).

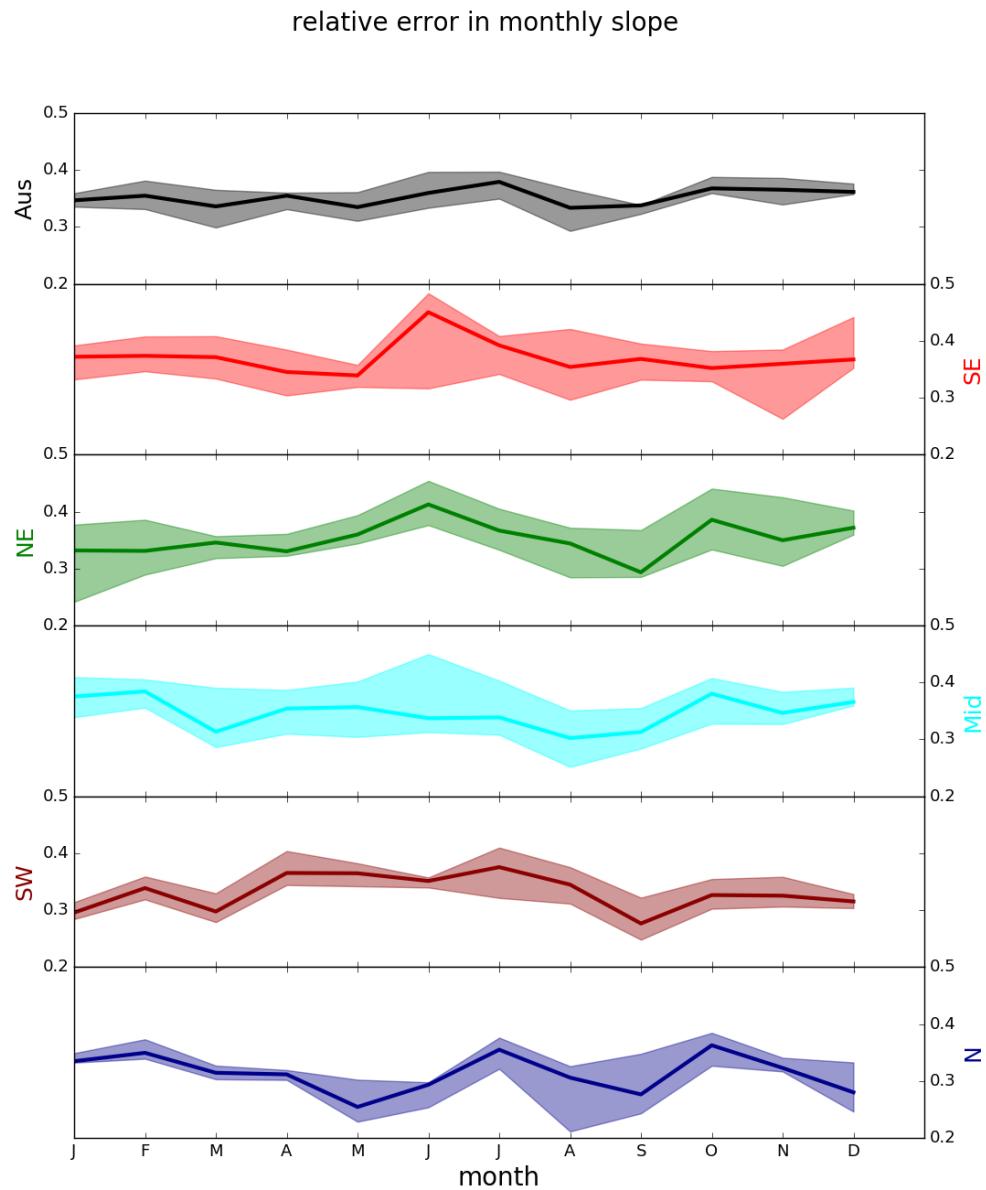


FIGURE 3.25: Median and inter-quartile range of monthly binned uncertainty in S .

Filtering for spatial smearing (see Section 3.2.7) reduces the number of data points making up ~~the~~ regression slope S . The process generally improves the linear relationship, and where it does not a multi-year average (or in the worst cases no value at all) is used in lieu of monthly S data. The thresholds for the smearing filter are based on literature ~~studying~~ values from other countries, which may prove to be unsuitable within Australia. To improve the understanding of smearing in Australia, a better approximation of HCHO lifetimes and yields, and NO_x seasonality and regional concentrations, is required. The filtering process is accounted for in the calculation of ΔS , and the overall effect on uncertainty is discussed. *where?*

Model biases are not analysed in this thesis, except to note that they would impact both preliminary OMI calculations and the modelled slope. Insufficient independent measurements in Australia ~~make it impossible to~~ mean that any quantified uncertainty ~~would require un-~~ reasonable extrapolation to apply at a national scale.

Hesitant to state this as we are extrapolating other things...

3.4.3 Satellite Uncertainty

Corrected vertical columns of HCHO from the OMI product are calculated using Equation 2.23: $\Omega = \frac{SC - RSC}{AMF}$. Error in satellite HCHO columns is determined by error in the three terms SC, RSC, and AMF:

Fitting error from the OMI retrieval

Fitting error represents the uncertainty in the DOAS technique used to estimate HCHO concentrations. Fitting error is provided in the OMHCHO product. This error is ascribed to the SC term.

Uncertainty in AMF calculations

Air mass factors model ~~how~~ the satellite instrument ~~is~~ vertically ~~sensitive~~ ^{im} to slant path measurements, and uncertainties arise predominantly from uncertain cloud parameters (Palmer et al. 2006). AMF uncertainty can be determined through comparison of GEOS-Chem output to independently measured HCHO columns. Here we use 30% as a rough estimate of error in this term (Palmer et al. 2006), since measurements over Australia are lacking. This error is ascribed to the AMF term.

How does this compare to your Willmington FTIR comparison?

prob need to remind reader that GEOS-Chem used to create AMF

Uncertainty of HCHO background

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

These sources of error can be reduced through spatial and temporal averaging, as they are assumed to be unbiased. Uncertainty is reduced by the square root of the number of pixels averaged over ~~our~~ ^{each} $2^\circ \times 2.5^\circ$ grid squares for each day or month. For example, daily averaging reduces pixel uncertainty by a factor of 2-4. Figure 3.26 shows the pixel counts in each region before and after applying filters. Winter has lower pixel count for southern regions, with approximately 50-60% fewer good pixels through

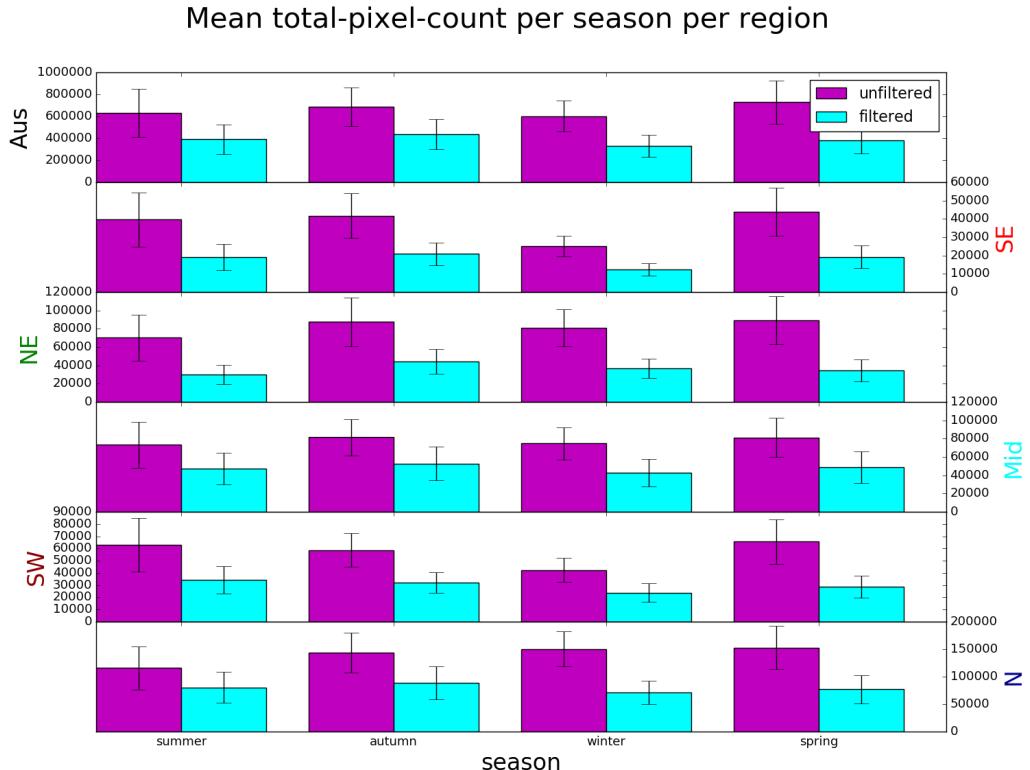


FIGURE 3.26: Mean and standard deviation (vertical error bars) of total pixel counts per region per season, before (magenta) and after (cyan) applying smearing, pyrogenic, and anthropogenic filters.

May, June, and July. The lowest pixel counts occur in the south~~east~~ in winter, likely due to a mix of ~~the~~ relatively high solar zenith angles~~x~~, and filtering due to anthropogenic emissions. Northern regions have lower pixel counts in the summer, most likely due to increased cloud coverage which limits satellite measurement capabilities. *as discussed previously.*

Calculation of uncertainty in the OMI vertical column HCHO ($\Omega = \frac{SC - RSC}{AMF}$) is performed using quadrature equations 3.13 and 3.14. Error in the slant column (ΔSC) is combined with our assumed relative AMF error ($\frac{\Delta AMF}{AMF}$) of 30%, and background error to calculate $\Delta\Omega_{OMI}$ (and $\Delta\Omega_{OMI,0}$) as follows:

$$\begin{aligned} \Delta(SC - RSC) &= \sqrt{(\Delta SC)^2 + (\Delta RSC)^2} \\ \Delta\Omega &= \Omega \sqrt{\left(\frac{\Delta(SC - RSC)}{(SC - RSC)} \right)^2 + \left(\frac{\Delta AMF}{AMF} \right)^2} \\ \frac{\Delta\Omega}{\Omega} &= \sqrt{\frac{(\Delta SC)^2 + (\Delta RSC)^2}{(SC - RSC)^2} + \left(\frac{\Delta AMF}{AMF} \right)^2} \end{aligned} \quad (3.18)$$

The RSC term is described in Chapter 2 Section 2.6.6. ΔRSC for each latitude is based

TABLE 3.6: Uncertainties in satellite total column HCHO.

uncertainty	location	source	notes
40%	North America	(a)	mostly due to cloud interference
26%	North America	(b)	with cloud fraction less than 20%
30%-40%	global	(c)	GOME-2 instrument
> 60%	Mid-latitude	(c)	GOME-2 instrument in winters
1%-10%	Australia	this work	monthly uncertainty at $2^\circ \times 2.5^\circ$
50%-100+%	Australia	this work	in winter at higher latitudes

a: Millet et al. (2006) and Palmer et al. (2006)

b: Millet et al. (2008)

c: De Smedt et al. (2008) and De Smedt et al. (2012)

are all others OMI then? Specify

on the standard deviation of corrections over the remote Pacific. Negative columns can occur where slant column amounts are lower than RSC, and these are not removed so as not to introduce a bias. When monthly averages are less than zero, relative error is set to 100%. This only impacted the uncertainty calculations in winter for the non-northern regions, where occasional highly negative absolute uncertainty was seen where Ω approached 0.

Figure 3.27 shows the relative uncertainty in monthly satellite columns for each sub-region and all of Australia. Uncertainty in winter at higher latitudes is greatly increased due to lower pixel counts, lower absolute column amounts, and higher fitting error.

Uncertainty in satellite HCHO ($\Delta\Omega$) from literature and calculated here is listed in Table 3.6. De Smedt et al. (2012) found satellite HCHO uncertainty to be 30 – 40% for the GOME-2 instrument by combining slant column systematic and random errors. For mid latitude winters they saw an excess of 60% uncertainty. OMI measurements will have similar amounts of uncertainty; however, the array of detectors provide more pixels which can be averaged to reduce this uncertainty.

In order to calculate the bias or systematic error, an understanding of biases in the underlying terms is required, since there is little in the way of comparable measurements. OMI has been shown to underestimate observed HCHO by up to 40%. For example OMI underestimated aircraft measurements by 37% in Guyana (Barkley et al. 2013). OMI underestimation ranges from 20-40% when compared against ground based remotely sensed profiles (Zhu et al. 2016; De Smedt et al. 2015). Satellite HCHO may also suffer from $\sim 13\%$ overestimation when taking monthly averages due to only measuring on relatively cloud-free days (Surl, Palmer, and Abad 2018). Since our a posteriori is linearly related to the satellite HCHO, any bias is directly transferred. We note therefore that our product may be biased by up to 40% too low, and this output may be 13% too high, which gives us a potential bias factor (β) of $1/0.6 \times 1/1.13 = 1.47$. This may be complicated further if the satellite bias over Australia does not match the bias over the remote Pacific at corresponding latitudes. However we can not quantify bias over Australia due to insufficient measurements. GEOS-Chem biases would affect the recalculation of HCHO, but they are assumed to be minimal in this thesis. Lacking sufficient suitable measurements to estimate

where? Also Zhu is based on aircraft not ground-based.

I definitely wouldn't say that! Safer to say they cannot be quantified and so are not included here.

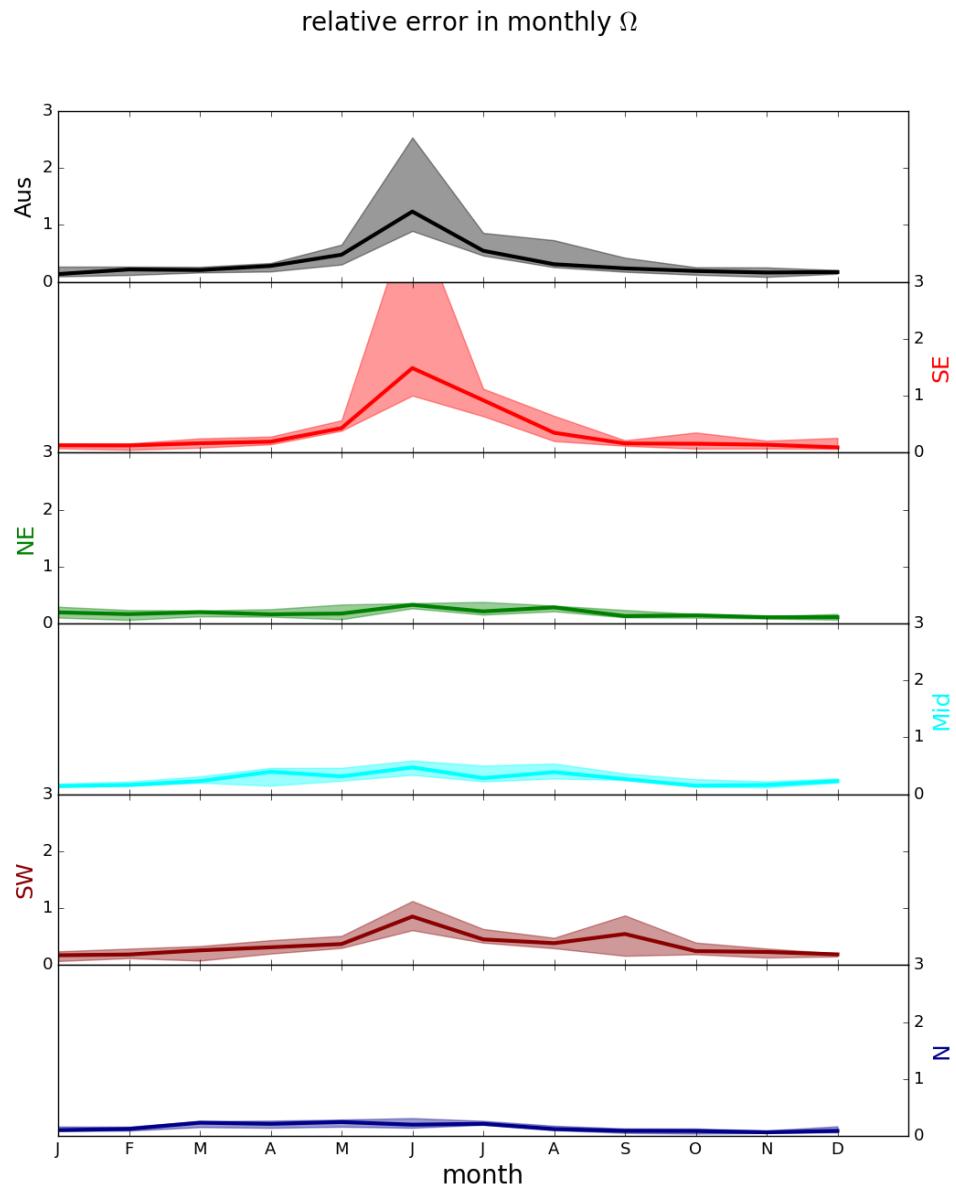


FIGURE 3.27: Median and inter-quartile range for monthly binned relative uncertainty in satellite vertical columns.

satellite bias over Australia, the potential bias factor $\beta = 1.47$ is applied to the mean emissions in each month when calculating the upper bound of our uncertainty range.

3.4.3.1 Fitting error from the OMI retrieval

I don't really understand why you have each of these subheadings twice, once at start of section and again here. Makes things clunky and leads to redundancy. Can you move true up to where you mentioned them before?

Provided with the OMI product is the measurement of uncertainty in each pixel, calculated by the Smithsonian Astrophysical Observatory from the back scattered solar radiation fit (Gonzalez Abad et al. 2015; Abad et al. 2016). This is used as the ΔSC in Equation 3.18. The relative fitting error per pixel ($\frac{\Delta SC}{SC}$) ranges from around 20% to 150%, and is higher where low amounts of HCHO are detected, and at higher solar zenith angles (i.e. at high latitudes).

3.4.3.2 Uncertainty in AMF calculations

Palmer et al. (2006) calculate the error in AMF through combining estimates of error in the UV albedo database ($\sim 8\%$), model error based on in situ measurements, cloud error (20 – 30% (Martin et al. 2003)), and aerosol errors (< 20%), totalling AMF error of around $\sim 30\%$ (calculated in quadrature). Compare this error estimate with that of Curci et al. (2010), where the error in AMF calculations and background columns are respectively found to be 30% and 15% based on their analysis of CHIMERE. Millet et al. (2008) also examine this uncertainty and determine an overall uncertainty (1σ) of 25 – 27% in HCHO vertical columns with calculated AMFs where cloud fraction < 0.2. Based on these prior works, I assume relative uncertainty in the AMF to be 30%

3.4.3.3 Uncertainty of HCHO background

The RSC, or background correction, is based on differences in the remote Pacific between daily HCHO slant columns measured by OMI and monthly averages from GEOS-Chem. The correction for each pixel is determined per latitude and OMI track; however, here we use a couple of conservative simplifications to estimate the error in this term. For each SC, the RSC is set to the mean correction matching the SC latitude over all tracks. For each day, the ΔRSC is set to the standard deviation of the RSC over Australian latitudes (45°S to 10°S) in all tracks. These terms are used in Equation 3.18. TODO: print some example rerr for RSC from omhchorp.uncertainty, summarise here. For comparison, the background error is assumed to be 15% in Curci et al. (2010) following (Dufour et al. 2008).

3.4.4 Sensitivity to vertical column recalculation

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

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

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

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

3.4.5 Sensitivity to filtering

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

3.5 Conclusions and implications

Very few ground based measurements of BVOCs concentrations and emissions are available in Australia. Emissions from models use largely unverified extrapolations for emission factors in Australia, leading to overestimated isoprene emissons. This leads to uncertainty and error when modelling atmospheric ozone and other trace gases. also add references

In this chapter I created and test^{ed} an isoprene emissions estimate (a posteriori) based on OMI satellite measurements of HCHO, a high-yield product of isoprene oxidation. To test how the a posterior emissions affect ozone and HCHO, I run GEOS-Chem with a seasonal (multi-year monthly averaged) gridded ($2^\circ \times 2.5^\circ$) scaling factor applied to the a priori emissions. Uncertainty in the primary components of the top-down calculation of emissions is calculated where possible, and potential biases are noted.

My top down biogenic isoprene emission estimate shows that MEGAN overestimates emissions in summer by up to a factor of 5, and total yearly Australian emissons are reduced from 45.4 Tg yr^{-1} to 19 Tg yr^{-1} (decrease of $\sim 58\%$). The overestimation is spatially and temporally diverse and leads to model overestimation of both HCHO and ozone. Running GEOS-Chem using scaled emissions reduced the overall HCHO overestimation from 50-120% down to 30-50% (TODO: update that number for summer season rather than the 2005 January it is based on). Scaling GEOS-Chem emissions also lowers simulated surface ozone concentrations by $\sim 5\%$. A posteriori uncertainty is analysed (where possible), and shown to be on the order of 50% in summer, with a potential bias coming from satellite data. The primary uncertainty in the a posteriori emissions comes from the monthly modelled isoprene to HCHO yield ($\sim 30\% - 50\%$), although at higher latitudes in winter, satellite uncertainty becomes restrictively high ($> 100\%$).

The full story behind model overestimation ^{reason for the} still unclear, and both global and Australian emissions estimates for isoprene range widely. ^{it} This is in part due to the MEGAN emission models' sensitivities to inexact parameters including leaf area indices, plant functional type emission factors, and meteorological factors. In Australia, a mixture of

Things you did go in past tense in Conclusions.
(fix throughout)

I don't remember seeing that - still a todo?

poorly defined emission factors (e.g. Emmerson et al. 2016), unaccounted for soil moisture (e.g. Sindelarova et al. 2014) and poorly understood forest responses to meteorological stresses (TODO: cite papers talking about clay and poor moisture responses) is likely driving uncertainty and model biases. In the US, bias between OMI and ~~in situ measurements~~ is as high as 40%, but bias ^{across} in Australia cannot be determined as there ~~is~~ are not enough ^{independent observations}. Ground-based VOC, ~~OVC~~, NO_x, and ozone measurements over large areas timed to coincide with satellite overpass times would be help quantify performed to handle the issue of unknown satellite biases while additionally providing some constraints for bottom-up models. In the northern region, where forests are affected by monsoonal seasons, characterisation of forest emissions and their response to sunlight, temperature, and moisture are even more important in order to improve modelled emissions. This is due to the reduced opportunity to get relatively cloud-free satellite measurements of the atmosphere at these times. In the future an improved soil moisture map and response parameterisation could be tested using top-down emissions estimates (such as is provided in this chapter) as a measure of model correction. I don't really understand what you mean by this and don't think it is a strong way to end the chapter.

add Emmerson 2019

Could mention using other satellites (GOME-2, TROPOMI) and evaluating resolution-dependence (things you originally planned to do).

Chapter 4

Stratospheric ozone intrusions

4.1 Foreword

This chapter is unchanged (excepting titles, labels, and numbering) from work published by me into the *Atmospheric Chemistry and Physics journal: Greenslade2017* *a paper on which I was the first author.*
Full reference here.

4.2 Introduction

Tropospheric ozone constitutes only 10% of the total ozone column but is an important oxidant and greenhouse gas which is toxic to life, harming natural ecosystems and reducing agricultural productivity. Over the industrial period, increasing tropospheric ozone has been estimated to exert a radiative forcing (RF) of 365 mWm^{-2} (Stevenson et al. 2013), equivalent to a quarter of the CO₂ forcing (Forster et al. 2007). While much tropospheric ozone is produced photochemically from anthropogenic and natural precursors, downward transport from the ozone-rich stratosphere provides an additional natural source of ozone that is particularly important in the upper troposphere (Jacobson and Hansson 2000, and references therein). The contribution of this source to overall tropospheric ozone budgets remains uncertain (Škerlav, Sprenger, and Wernli 2014), especially in the southern hemisphere (SH). Models show that stratospheric ozone depletion has propagated to the upper troposphere (Stevenson et al. 2013). However, work based on the Southern Hemisphere Additional OZonesonde (SHADOZ) network suggests stratospheric mixing may be increasing upper tropospheric ozone near southern Africa (Liu et al. 2015; Thompson et al. 2014). Uncertainties in the various processes which produce tropospheric ozone limit predictions of future ozone-induced radiative forcing. Here we use a multi-year record of ozonesonde observations from sites in the southern hemisphere extra-tropics, combined with a global model, to better characterise the impact of stratospheric ozone on the tropospheric ozone budget in the southern hemisphere.

Stratosphere-to-troposphere transport (STT) primarily impacts the ozone budget in the upper troposphere but can also increase regional surface ozone levels above the legal thresholds set by air quality standards (Danielsen 1968; Lelieveld et al. 2009; Lefohn et al. 2011; Langford et al. 2012; Zhang et al. 2014; Lin et al. 2015). In the western US, for example, deep STT events during spring can add 20–40 ppbv of ozone to the ground-level ozone concentration, which can provide over half the ozone needed to exceed the standard set by the U.S. Environmental Protection Agency (Lin et al. 2012; Lin et al. 2015). Another hotspot for STT is the Middle East, where surface