

## 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 ( $\text{OH}$ ) cycling, ozone ( $\text{O}_3$ ) production, 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 ( $\text{C}_5\text{H}_8$ ) 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  $\text{yr}^{-1}$ , while Sindelarova et al. (2014) estimated around 411 Tg  $\text{yr}^{-1}$ . Reactions following emissions are complex, and are sensitive to other trace gases in the 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 (Emmerson et al. 2016). The sample of trees include 4 types of Eucalyptus, which are not representative of the hundreds of species that make up Australian forests. Additionally, 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 primary product of most BVOC (including isoprene) reactions, 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 southeast Australia (Emmerson et al. 2016). This chapter aims 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 evaluated against bottom-up a priori estimates and also briefly compared against 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 section 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 estimating biogenic isoprene emissions. However, 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 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. Bauwens et al. (2016) showed that isoprene emissions were overestimated by up to a factor of 3 over Australia. 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 posteriori data with their a priori (model) data and independent data sets. For Australia they found emissions ranging from  $26\text{--}94 \text{ Tg C yr}^{-1}$ , with MEGAN based estimates ranging from  $38 \text{ Tg C yr}^{-1}$ – $94 \text{ Tg C yr}^{-1}$ , and a posteriori emissions of  $36 \text{ Tg C yr}^{-1}$ . In this thesis I focus on the analysis of a top-down emissions estimate compared against MEGAN, along with how changed emissions affect modelled ozone levels.

### 3.1.3 Top-down isoprene emissions estimates

In the remote troposphere HCHO production is dominated by methane oxidation, while in the continental boundary layer production is largely due to non-methane VOCs (NMVOCs) (Abbot 2003; Kefauver, Filella, and Peñuelas 2014). This leads to a causal relationship between enhanced HCHO concentrations and NMVOC emissions at low ( $< 1 \text{ km}$ ) altitudes. NMVOCs are generally short lived ( $< 1 \text{ hr}$ ), 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 it 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. Continental tropospheric HCHO enhancement can be directly linked to biogenic isoprene emissions. Since 1997, when GOME measurements were first used to measure HCHO over Asia, satellites have been used to provided a total column measurement of HCHO, enabling 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, 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):

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

where  $\epsilon$  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 was reduced to 3-25% after applying satellite based improvements. More recently Kaiser et al. (2018) 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). 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 ( $\Omega$ ) enhancement above a background level ( $\Omega_0$ ) 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  $\Omega$  ( $\Omega_{sat}$  and  $\Omega_{sat,0}$ ) to determine  $\hat{E}_{isop}$ :

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

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

The calculation requires reaction rates and yields from isoprene to HCHO, which can be determined most readily using chemical modelling. The method for calculating isoprene emissions from HCHO is laid out in 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, high wind speeds can transport precursors, or low  $\text{NO}_x$  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 midday tropospheric columns of HCHO and GEOS-Chem midday biogenic isoprene emission rates, and then this relationship is applied to satellite measured HCHO total columns to derive a new isoprene emissions estimate. Daily modelled values averaged between 13:00-14:00 LT are used to match the overpass time of the Aura satellite. Then the slope is calculated using the reduced major axis (RMA) regression between the a priori isoprene emissions (those from GEOS-Chem,  $E_{GC}$ ) and tropospheric HCHO columns in each model grid box each month. There is very little HCHO above the tropopause, so differences between total and tropospheric column are negligible. In this work total and tropospheric column HCHO are used interchangeably and referred to using  $\Omega$ .

### 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 ( $\Omega_{OMI}$ ; 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  $\Omega_{OMI}$  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. A full description of the creation of this smearing filter is given in Section 2.7.3.
4. GEOS-Chem modelled biogenic emissions of isoprene ( $E_{GC}$ ) along with biogenic columns of HCHO ( $\Omega_{GC}$ ), 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  $\Omega_{OMI}$  and  $S$  then form the basis for the 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 3.2.7.
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 and SPS). Results are examined in Section 3.3.
7. GEOS-Chem is run using the a posteriori emissions (see Section 3.2.9), and HCHO, O<sub>3</sub>, isoprene, and NO<sub>x</sub> outputs are compared to a priori values and to campaign and satellite measurements where possible (Sections 3.3.3).

### 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 ( $\Omega_{OMI}$ ), 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  $\Omega_{OMI}$ , we require filters for non-biogenic sources. These masks are described in Section 2.7, and a brief recap is provided here. Methane oxidation is a major part of the formation of background HCHO; however, the linear regression used to estimate isoprene emissions effectively ignores this source, which means a methane filter is not required. Anthropogenic, pyrogenic, and smoke influence masks are created from three satellite products: NO<sub>2</sub>

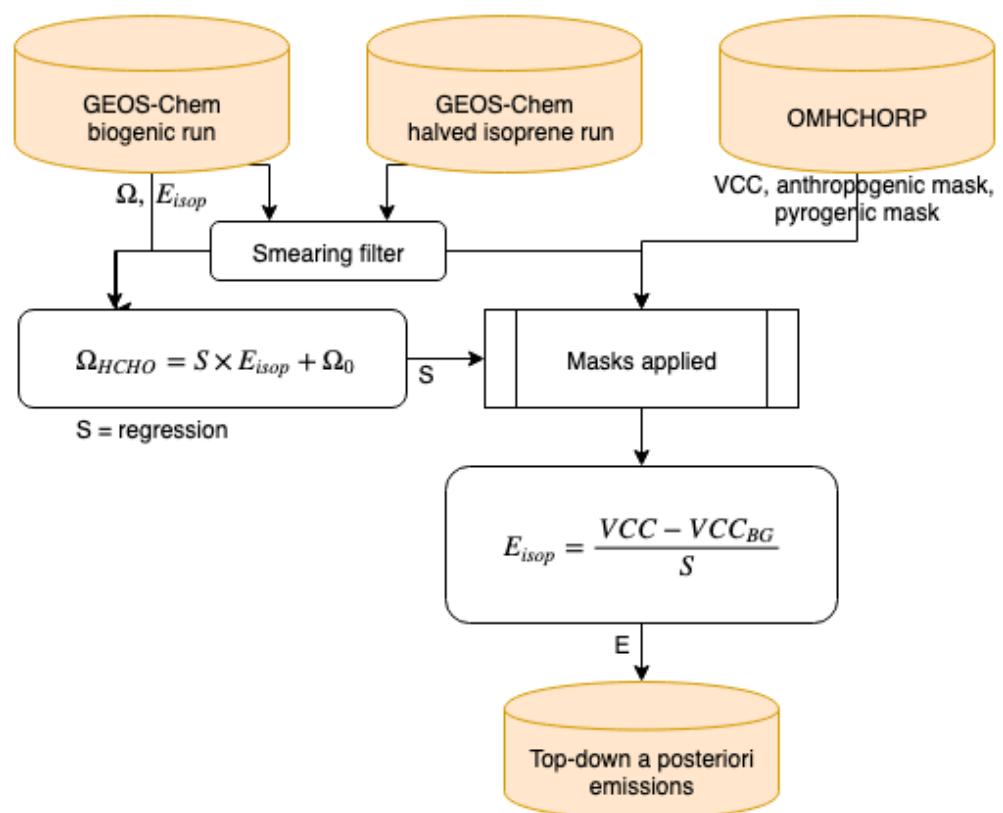


FIGURE 3.1: Top-down isoprene emissions estimate formation using OMHCHORP and GEOS-Chem outputs.

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  $0.25^\circ \times 0.3125^\circ$  horizontal resolution.
2. Influence from transported smoke plumes is removed by flagging grid squares 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 NO<sub>2</sub> tropospheric column amounts, masking any grid squares with greater than  $2 \times 10^{15}$  molec cm<sup>-2</sup> on any particular day, along with grid squares where the yearly average is above  $1.5 \times 10^{15}$  molec cm<sup>-2</sup>.

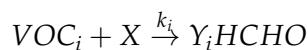
The recalculated corrected vertical columns are saved to OMHCHORP dataset both before and after applying the filters to allow filter analysis.

### 3.2.3 GEOS-Chem emissions

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

### 3.2.4 Relationship between isoprene emissions and formaldehyde

Tropospheric HCHO production is primarily due to the oxidation of VOC precursor species ( $VOC_i$ ). Background concentrations are driven by methane; a longer lived ( $\sim 8$  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). HCHO is then 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 ( $\Omega$ ; molec cm<sup>-2</sup>) can be linearly related to isoprene emissions. 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<sub>x</sub> concentrations (e.g., Palmer et al. 2006; Millet et al. 2006; Bauwens et al. 2016; Surl, Palmer, and Abad 2018).

The isoprene to HCHO relationship is derived using several assumptions that are outlined here. 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  $\Omega$  has had time to stabilise. The second assumption is that loss (with loss rate constant  $k_{HCHO}$ ) 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  $\sim 200$  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 by applying a smearing filter described in Section 2.7.3.

Another assumption is that  $\Omega$  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  $\Omega$  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  $\Omega$  (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 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). The linear relationship between isoprene emissions and  $\Omega$  is determined by equating  $P_{HCHO}$  and  $L_{HCHO}$  from Equations 3.4 and 3.3, substituting 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:  $S \equiv \frac{Y_{isop}}{k_{HCHO}}$ .

### 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  $\Omega_{GC}$  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 cm<sup>-2</sup> s<sup>-1</sup>) at 13:00 LT daily are extracted, along with  $\Omega_{GC}$  ( molec cm<sup>-2</sup>) output.
2. Filtering removes gridded output on days where grid squares are likely to be affected by smearing (see Section 3.2.8).
3. A reduced major axis regression slope is determined between  $\Omega_{GC}$  and  $E_{GC}$  using a month of modelled output (one value per day) for each grid square (e.g., see Figure 3.2)

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. High  $S$  values suggest increased sensitivity to isoprene emissions. 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 and HCHO levels in those areas. The regression coefficients also vary spatially (top right), and some areas show little correlation. It is likely that this is due to weather, transport, and a lack of local emission sources. The slopes in the bottom 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 described in Section 2.7.3. Due to the 2° × 2.5° horizontal resolution of GEOS-Chem, calculations over coastal grid boxes that are mostly oceanic are often discarded as the change in HCHO is not dominated by emissions of isoprene, as is assumed for equation 3.7.

One issue with slope calculation is potential transport (also known as smearing), either of isoprene transported in from outside the local grid box (before any HCHO is formed), or of HCHO formed by local emissions but transported out of the local grid box. The effects of this are dealt with using a smearing filter (see Section 2.7.3). 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, presumably due to higher biogenic isoprene emissions over summer, making transport more noticeable on windy days. Additionally, in winter, the HCHO column is relatively

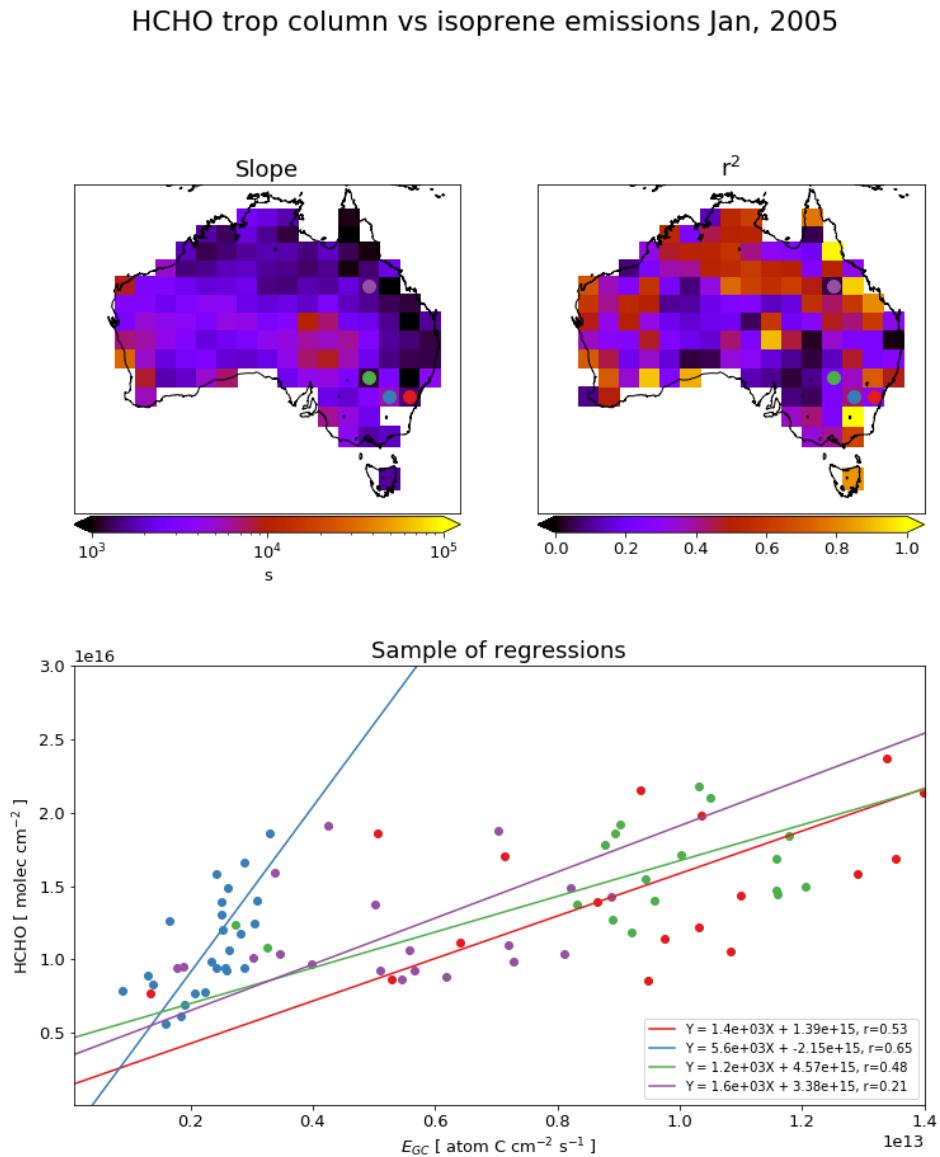


FIGURE 3.2: Top left: RMA slope between modelled tropospheric column HCHO and isoprene emissions ( $E_{GC}$ ) using midday (13:00–14:00 LT) values over for January 2005, per grid square at  $2^\circ \times 2.5^\circ$  horizontal resolution. Top right: Squared RMA correlation coefficient for regression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

more affected by anthropogenic precursors, which are not halved between the two runs. This may lower the smearing filter signal over Sydney and other densely populated areas. 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 the 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.

### 3.2.6 Modelled background HCHO

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 the simulated HCHO without isoprene emissions. 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 between  $\Omega_{GC}$  over Australia with no isoprene emissions (bottom left), and  $\Omega_{GC}$  over the remote Pacific mapped over Australia (top left) is shown for an example month (January, 2005). This difference is relatively small ( $\sim 1$  to  $3 \times 10^{15}$  molec cm $^{-2}$ ), 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 ( $140^{\circ}\text{W}$  to  $160^{\circ}\text{W}$ ) the matching latitudes over the remote Pacific.

### 3.2.7 Calculation of Emissions

Top-down emissions estimates are calculated using OMHCHO (see Section 2.3) slant columns and an updated AMF calculated using code written 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 ( $\Omega_{OMI}$  and  $\Omega_0$ ).  $\Omega_0$  is calculated using  $\Omega_{OMI}$  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

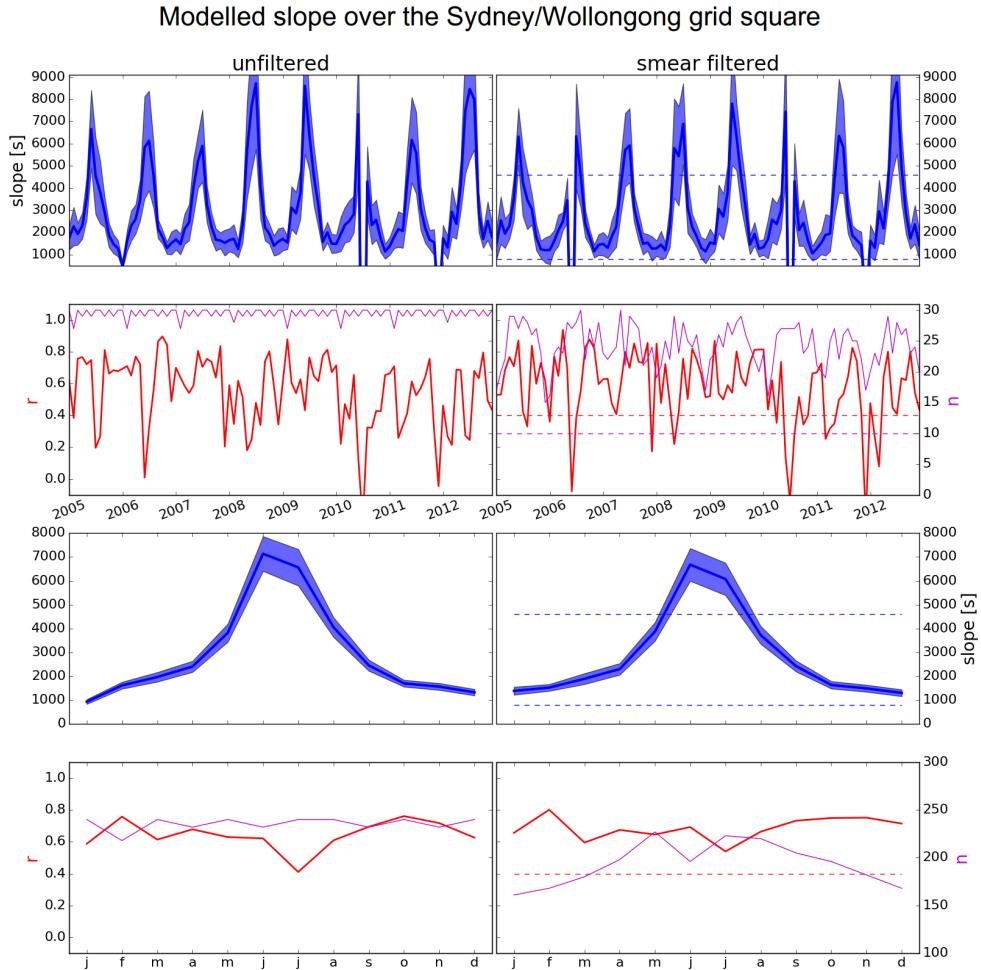


FIGURE 3.3: Row 1: monthly slope along with 95% confidence interval both before (left) and after (right) applying the smear filter for the model grid square containing Sydney over 2005-2012. Limits used in creation of the smear filter are shown with dashed lines. Row 2: regression coefficient and data-point 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.

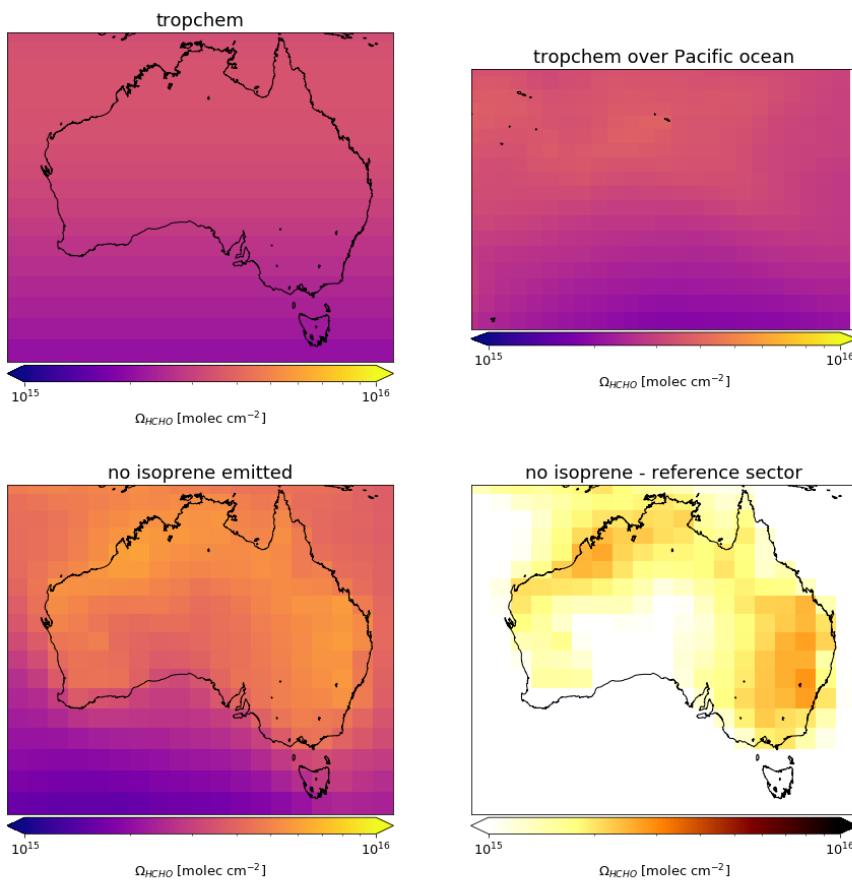


FIGURE 3.4: Top left: Total column HCHO over Australia using the standard tropchem GEOS-Chem simulation. Top right: As top left except over the remote Pacific region at southern mid-latitudes. Bottom left: Background  $\Omega_{HCHO}$  defined using the no isoprene emissions GEOS-Chem simulation. Bottom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

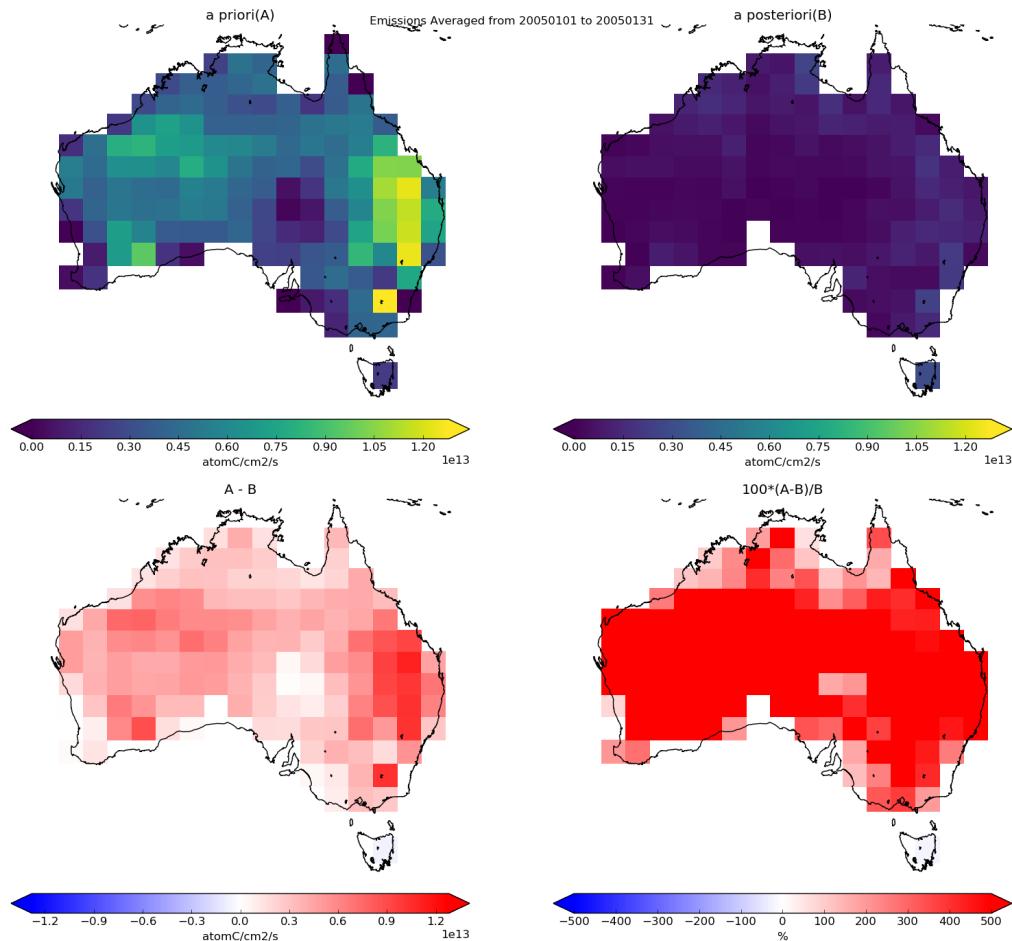


FIGURE 3.5: Top row: isoprene emissions from GEOS-Chem (a priori, left) simulation and top-down (a posteriori, right) calculations averaged over the month of January, 2005. Bottom row: the absolute (left) and relative (right) differences between the two.

a priori exceeds 500% of the a posteriori in many regions; however, the major absolute differences occur in the east between Sydney and Brisbane. Analysis of the full record is discussed in the results (Section 3.3).

One potential issue in this top down estimation technique is the low number of valid satellite measurements that may occur due to the higher solar zenith angles in winter and at higher latitudes. Another issue is that negative emissions are commonly calculated in areas wherever measured HCHO columns are lower than background amounts (as  $E_{OMI} = \frac{\Omega_{OMI} - \Omega_0}{S}$ ). These are set to zero, which increases the overall average emissions from  $4.4 \times 10^{11}$  atom C cm $^{-2}$  s $^{-1}$  to  $1.1 \times 10^{12}$  atom C cm $^{-2}$  s $^{-1}$ , due to the frequency of sub-zero column amounts (especially in winter) over low isoprene emission areas.

### 3.2.8 Sensitivity to smearing

The top-down isoprene emission estimate assumes enhanced HCHO is caused by local emissions, which is not always the case. A *smearing* filter is created (see Section 2.7.3) which filters out grid squares on days when transport is likely, and the effects on the modelled  $S$  values are shown here. 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 both spatially and temporally. Smearing is also sensitive to time of day, season, and latitude, as lower insolation results in slower photolysis. Figure 3.6 shows smearing and how frequently grid squares are filtered using the smearing filter. The smearing filter is more active in winter and spring, especially at higher latitudes. Grid squares along the east coast are filtered frequently in all seasons, as well as grid squares in the high smearing lake Eyre region in northeastern South Australia. During summer data loss from smearing is approximately 30% over the entirety of Australia, which increased to 40% over winter. The data loss is generally higher towards the north east and southern coastlines, and at lake Eyre.

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 emissions or unsuitable meteorological conditions.  $S$  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. This happens only infrequently, and only in locations with very limited isoprene emissions such as the lake Eyre basin.

### 3.2.9 Running GEOS-Chem using a posteriori emissions

After creating the a posteriori isoprene emissions estimate, GEOS-Chem is re-run with biogenic emissions scaled to match the new estimate. This is performed by applying a seasonal scaling factor  $\alpha$ , 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.  $\alpha$  is the ratio between the multi-year averaged monthly emissions from GEOS-Chem  $E_{GC}$  and the a posteriori  $E_{OMI}$ :

$$\alpha = \frac{E_{OMI}}{E_{GC}} \quad (3.9)$$

This seasonal scaling retains shorter time-scale variability and meteorological dependencies built into the parameterisations of the MEGAN model, while ensuring the multi-year monthly averaged emissions match. Initially  $\alpha$  is uniformly set to 1 globally. Where top down emissions exist and  $E_{GC}$  is non-zero, we set  $\alpha$  using Equation 3.9.  $\alpha$  is applied through the emissions module in GEOS-Chem where isoprene emissions are calculated. First, the new midday (13:00-14:00 LT) emissions (per grid box) are combined forming a multi-year monthly mean, which can be compared to the a priori equivalent. Missing values for  $\alpha$  when  $E_{GC}$  are zero are a negligible issue since

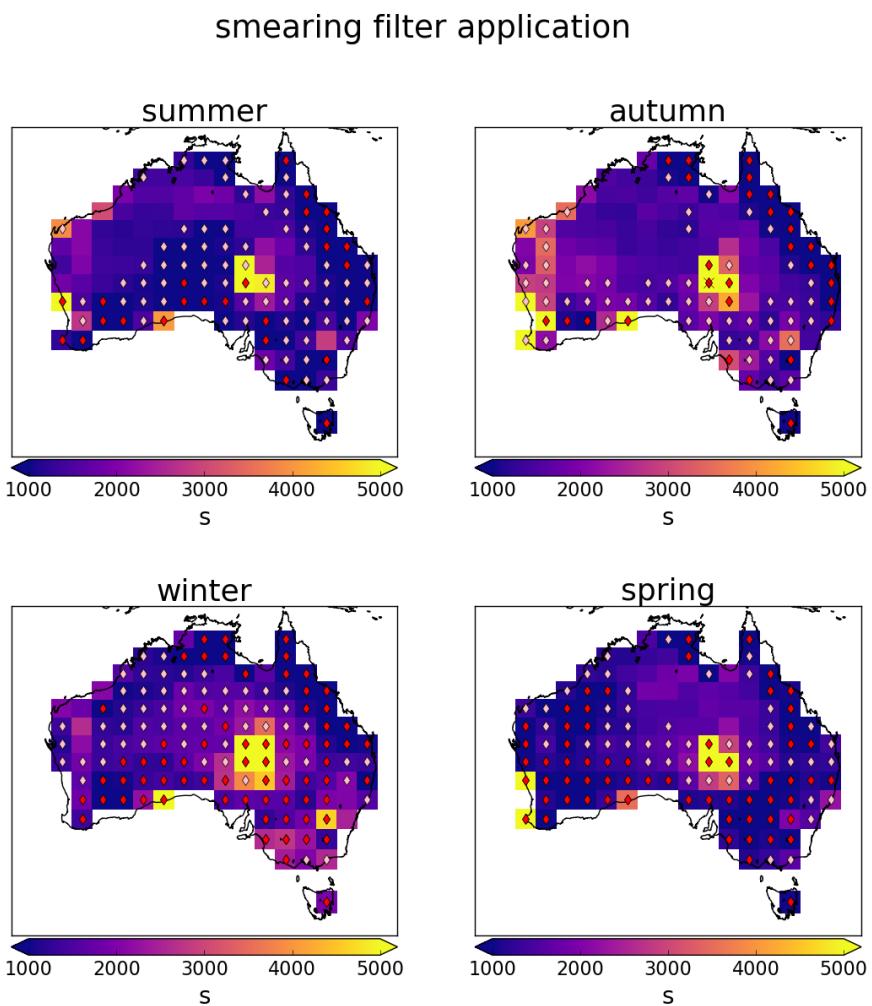


FIGURE 3.6: Seasonally averaged smearing ( $\hat{S}$ , see text) over 2005. Diamonds represent grid squares which have had at least 10 (pink) or 30 (red) days removed due to the smearing filter over the season. Red crosses show where the filter has removed all data.

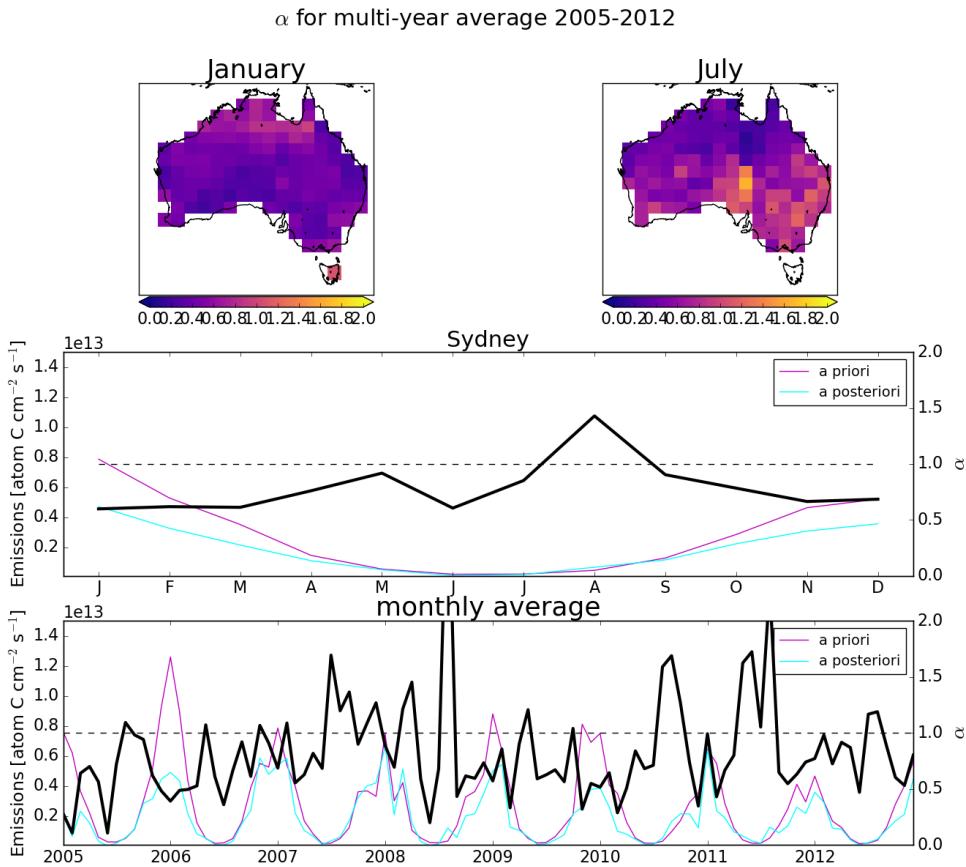


FIGURE 3.7: Row 1:  $\alpha$  for the average January (left) and June (right) over 2005-2012. Row 2: a priori (magenta, left axis), a posteriori (cyan, left axis), and  $\alpha$  (black, right axis) multi-year monthly averages calculated for Sydney. Row 3: Monthly averages of the same terms in Row 2.

the dominant discrepancies between estimates occur during summer when high emission rates are overestimated. Figure 3.7 shows  $\alpha$  for January and June averaged over 2005-2012, along with the time series of  $E_{GC}$  and  $E_{OMI}$  and  $\alpha$  calculated for Sydney and their multi-year seasonal average.

### 3.2.10 Conversion to emissions by kg

Top-down emission rates calculated in this work are in units of atom C cm<sup>-2</sup> s<sup>-1</sup>. 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 3.8 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

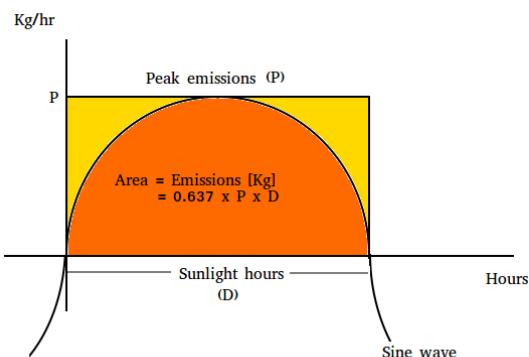


FIGURE 3.8: Total daily isoprene emissions (in kg) is represented by the area under the sine wave.

diurnal peak. GEOS-Chem emissions are similarly multiplied by area, but then integrated over time using hourly output to derive emissions in kg.

### 3.3 Results

Australia covers roughly  $7.7 \times 10^6$  km<sup>2</sup>, with heterogeneous environmental conditions. The results presented in this section are therefore frequently split into five regions that are differentiated by colour, as shown in Figure 3.9. These regions are large enough to reduce the uncertainty with at least 10 grid squares in each area, and small enough to be somewhat homogeneous at the national scale.

#### 3.3.1 A posteriori emissions

Figure 3.10 shows a priori emissions over Australia along with a posteriori emissions calculated as described 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 overestimate peaking in summer can be seen in all regions except the northern, which is generally overestimated but less so in summer.

Figure 3.11 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 overestimates occur in spring, and large differences are seen in all seasons. Uncertainty is summarised for the a posteriori using the mean monthly uncertainty per grid square within each region, divided by the square root of 24 (8 years times 3 months per season). Uncertainty is relatively small since we have 8 years of data combined into the seasonal mean. The potential bias (see Section 3.4.1) is added to the uncertainty and displayed using horizontal red bars. The highest over-estimations (occurring in spring in the Northern region, and summer in all others) lie outside the potential bias

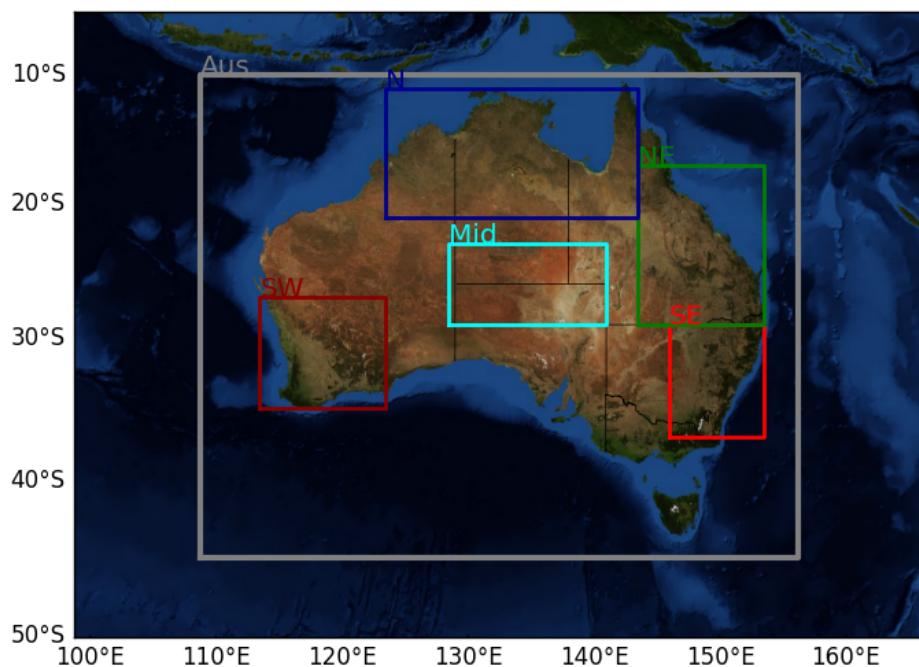


FIGURE 3.9: Sub-regions used in subsequent figures: Northern, North Eastern, South Eastern, South Western, and Middle. Australia-wide averages will be black or grey, while averages from within the coloured rectangles will match the colour shown here.

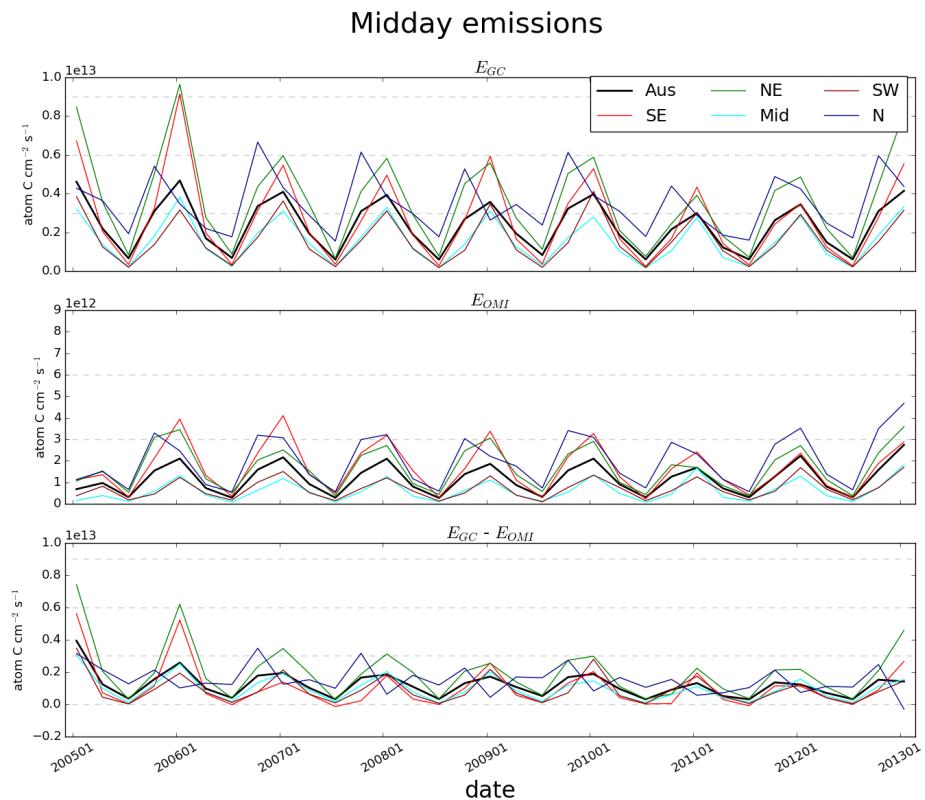


FIGURE 3.10: Row 1: Biogenic emissions of isoprene from GEOS-Chem (a priori,  $E_{GC}$ ). Row 2: Emissions calculated using the OMI top down inversion (a posteriori,  $E_{OMI}$ ). Row 3: Absolute differences between the first two rows. Midday emissions are averaged for each season (DJF, MAM, JJA, SON), and colours represent averaged areas from sub-regions shown in Figure 3.9. Grey dashed horizontal bars are added highlighting the scale between rows.

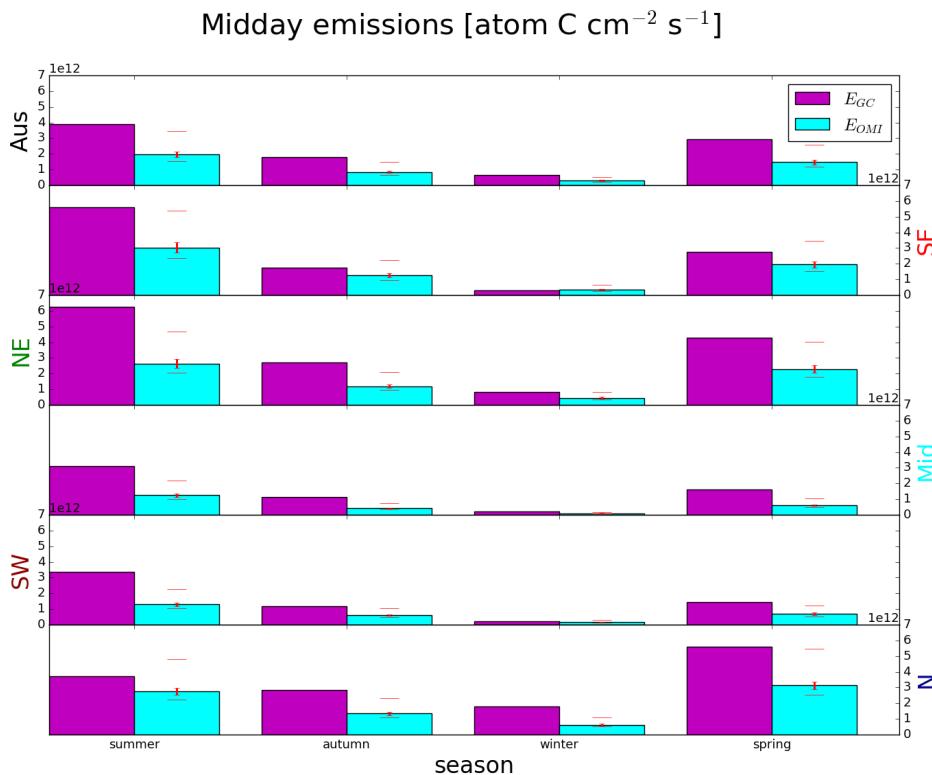


FIGURE 3.11: Regional multi-year seasonal mean a priori emissions (magenta) compared to a posteriori emissions (cyan). Error bars show the regionally averaged uncertainty. Additional horizontal dashes show the uncertainty plus effects from potential HCHO biases (discussed in Section 3.4.1) from satellite underestimation (40%) and monthly clear sky overestimation (13%).

caused by satellite underestimations of HCHO. However, many overestimated seasons within each region are within this wide range of potential bias and uncertainty. Some monthly grid squares are very uncertain ( $> 200\%$ ) and these are removed for this plot, causing an increase in a posteriori emissions of  $< 1\%$ . The cause for these highly uncertain grid squares is not analysed, but may be worth examining in future work.

Figure 3.12 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 and soil moisture (Emmerson et al. 2016;

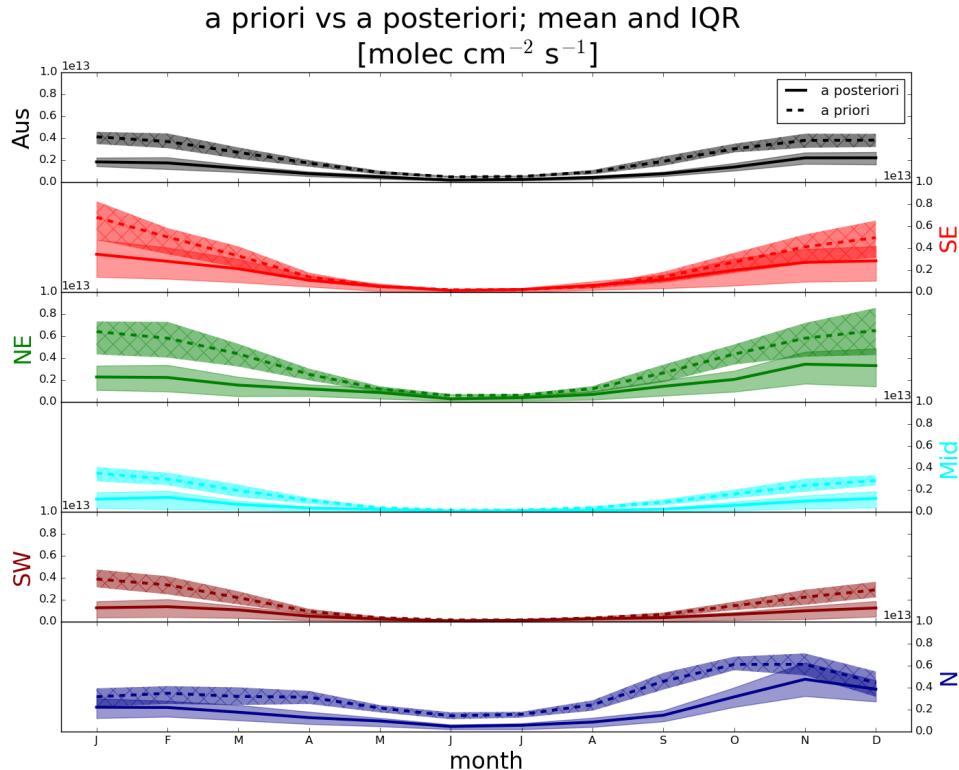


FIGURE 3.12: The multi-year monthly mean (lines) and IQR (shaded) of midday (13:00-14:00 LT) isoprene emissions estimates. A priori emissions are shown by the dashed lines and hatched shaded areas show the IQR. A posteriori emissions are shown using the solid lines, with IQR shown by unshaded shaded areas. Colours denote the region over which the monthly average was taken, as in Figure 3.9.

Emmerson et al. 2019).

While most regions show similar overestimates, the northern region of Australia follows a different cycle of bias. Northern Australian emissions appear 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 due to increased cloud coverage, while also changing the ecosystem's response to sunlight and temperature (e.g., Surl, Palmer, and Abad 2018). This is evidenced 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. In-situ measurements in both monsoon and non-monsoon seasons are required before robust conclusions can be drawn in this region.

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 3.13 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, the linear regression coefficient  $r$  ranges 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 southeast ( $r = 0.81$ ) and southwest ( $r = 0.79$ ) regions, followed by the northeast ( $r = 0.61$ ) 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. In part this is 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 (Emmerson et al. 2016; Surl, Palmer, and Abad 2018; Emmerson et al. 2019). If the same analysis is performed over winter months, the regression in the northern region is greatly improved (to  $r = 0.66$ ), while other regions remain approximately the same. If the modelled emissions are correct, this suggests that satellite measurements are not capturing representative samples of monthly grid square averages in summer, likely because of cloud coverage, and the summer inversion in this region may be biased.

When comparing GEOS-Chem (MEGAN) to the a posteriori calculated using our top-down inversion, we find a decrease of  $\sim 46\%$  from  $39 \text{ Tg yr}^{-1}$  to  $21 \text{ Tg yr}^{-1}$ . Table 3.1 compares annual Australian isoprene emissions from this work to previously published values. Our a posteriori estimate of  $19 \text{ Tg yr}^{-1}$  suggests isoprene emissions may be lower than any bottom up estimates, and is close to the lowest estimate of Bauwens et al. (2016) of  $26 \text{ Tg yr}^{-1}$ . Figure 3.14 shows how this decrease is distributed spatially, with  $E_{GC}$  and  $E_{OMI}$  in  $\text{Tg yr}^{-1}$  calculated as a multi-year mean. Across all of Australia we see large reductions of total emissions using the new top-down estimate.

### 3.3.1.1 Diurnal emissions

Figure 3.15 shows the a priori daily emissions cycle for Australia compared to the estimated a posteriori emissions cycle over Australia. The conversion of midday a posteriori emissions ( $\text{ molec cm}^{-2} \text{ s}^{-1}$ ) into  $\text{Tg yr}^{-1}$  involves integration over an assumed sinusoidal diurnal emission cycle as described in Section 3.2.10. 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 diurnal emission cycles and the assumed a posteriori cycle. This potential bias is not analysed further, and should be relatively small compared to other uncertainties as it only affects emissions towards the daily minima.

### 3.3.1.2 Trends

Figure 3.16 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. They are averaged into

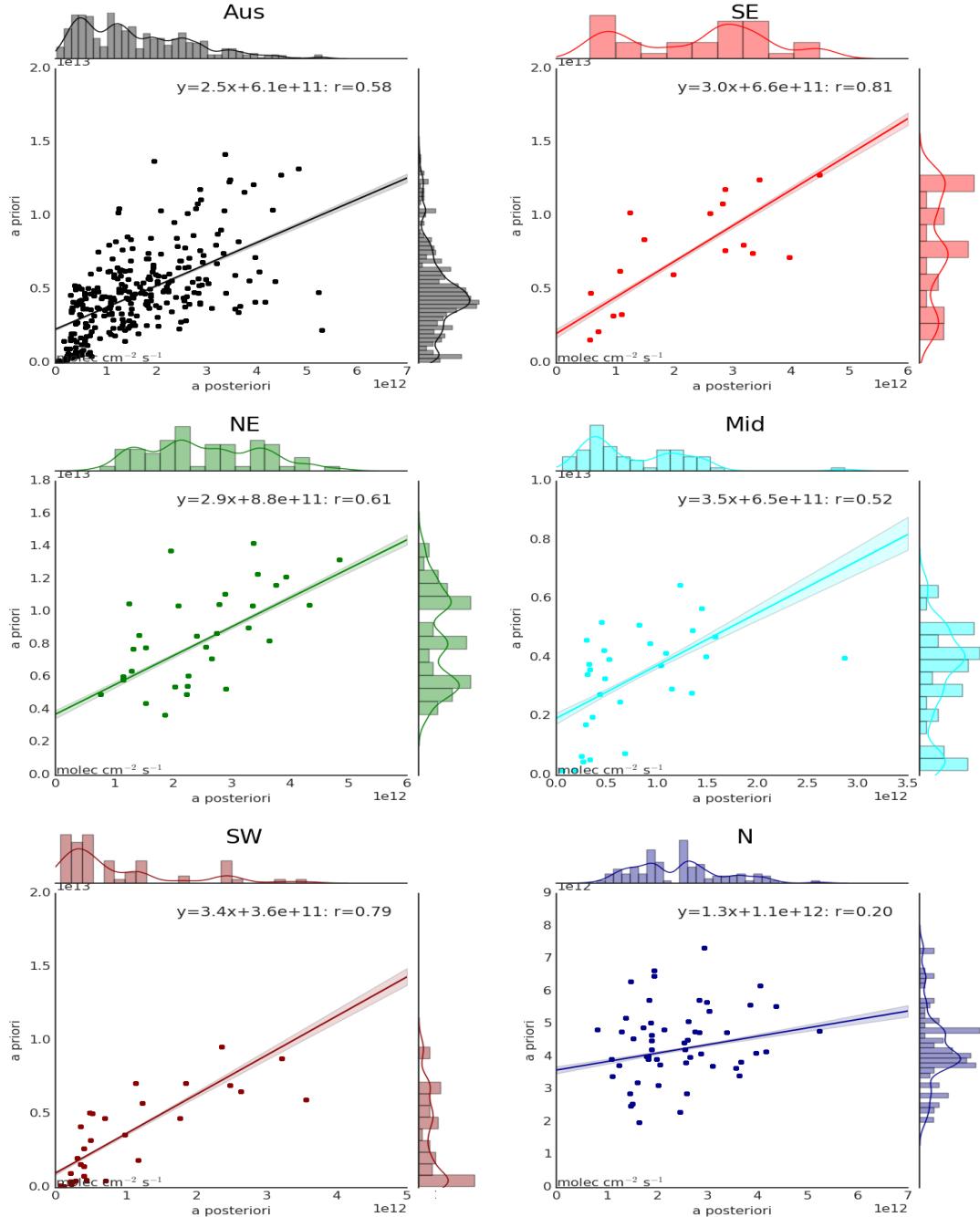


FIGURE 3.13: Scatter plot of a priori emissions against a posteriori using monthly averaged grid squares as regression datapoints. Data points are created using monthly averages (of midday emissions) for each grid box for each month of summer (DJF) within each region shown. Multiple years of data are 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 3.9. The linear best fit regression is inset into each plot along with the line equation and regression coefficient. The normalised distribution of each population is shown at the top and right spine of each subplot, with the right spine (facing the a priori axis) using the a priori axis and scale, and the top spine using the a posteriori axis and scale.

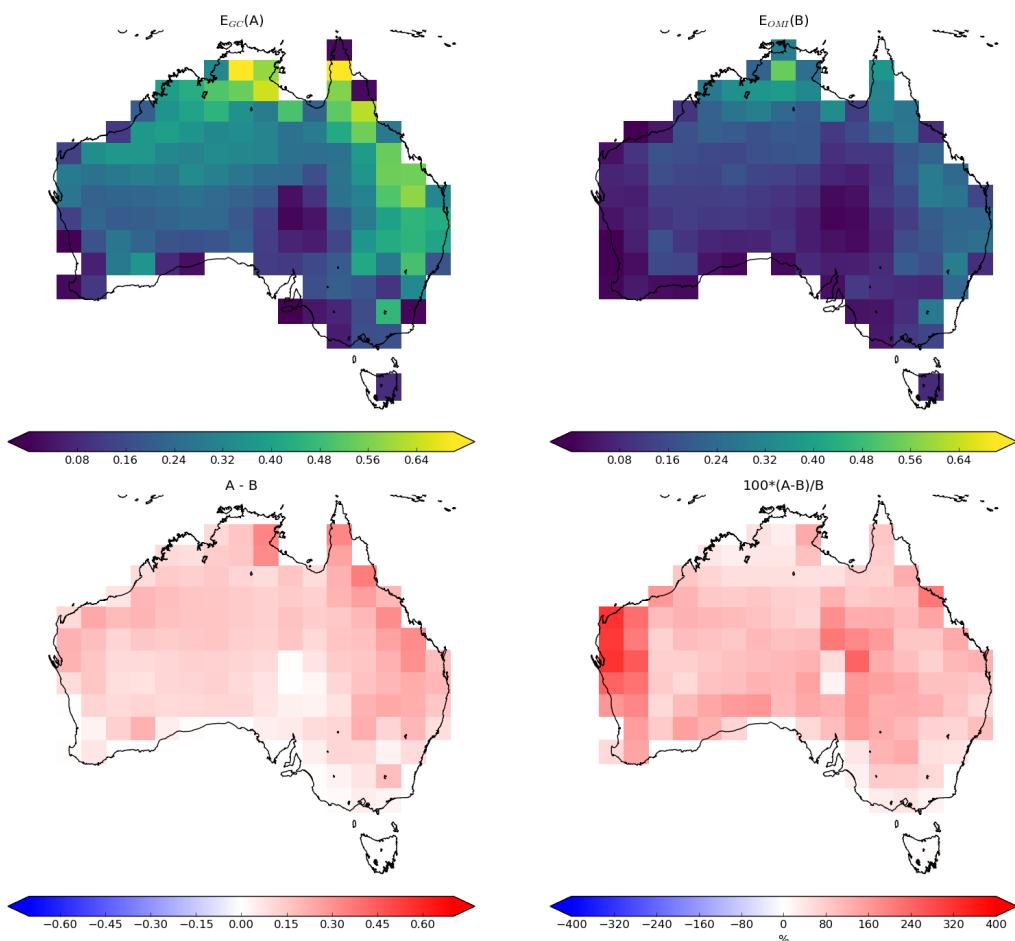


FIGURE 3.14: Top row: multi-year mean a priori emissions in  $Tg\text{ yr}^{-1}$  from  $E_{GC}$  (GEOS-Chem; running MEGAN) and  $E_{OMI}$  (top-down emissions) respectively.  $E_{OMI}$  uses an assumed sinusoidal daily cycle, with daylight hours prescribed for each month: see Section 3.2.10). Bottom left and right show the absolute and relative differences, respectively.

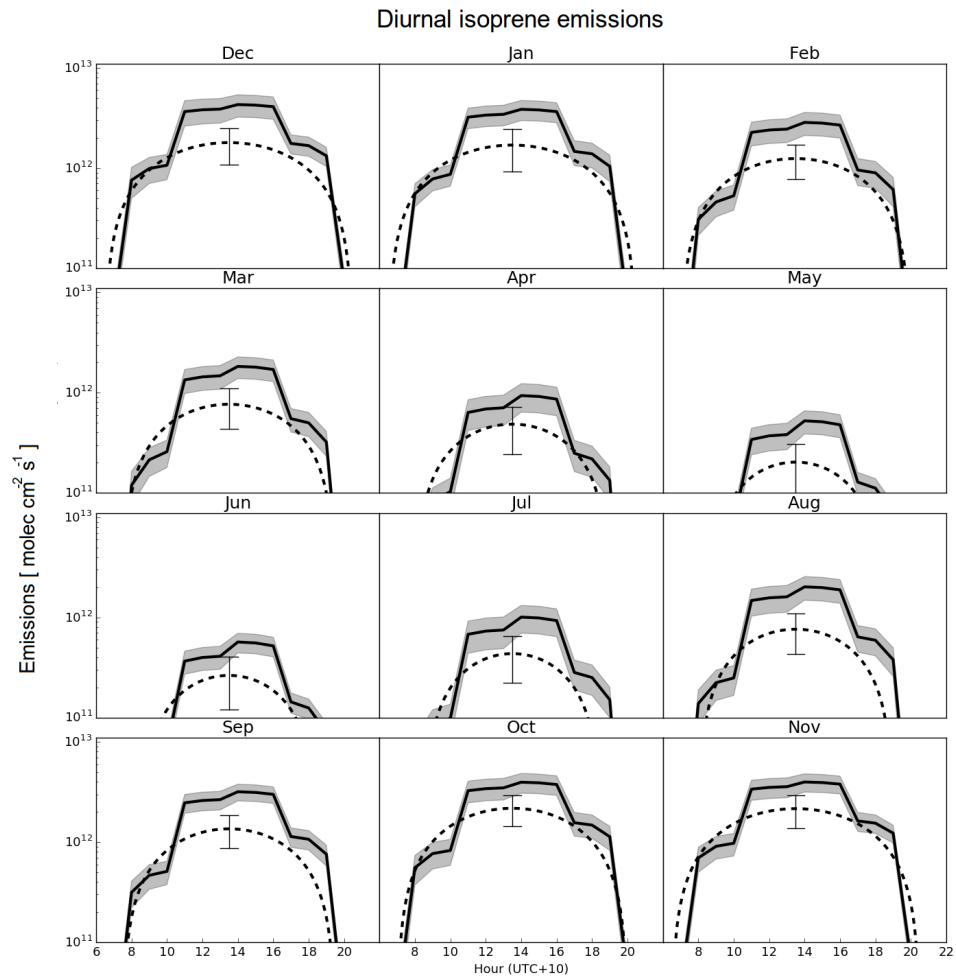


FIGURE 3.15: The diurnal cycle of GEOS-Chem a priori emissions (solid line) averaged by month into hourly bins over from 2005 to 2013 are shown against top-down a posteriori (dashed line) emissions. Standard deviations for the monthly average are shaded for the a priori, and shown with error bars at 13:30 LT for the a posteriori. Top down emissions shown here are based on monthly midday emissions being the peak of a sine wave which drops to zero after and before daylight hours (see Section 3.2.10).

TABLE 3.1: Isoprene emissions (Tg/yr) from Australia

Estimate <sup>a</sup>	Source	Year	Reference
39.2(4.0)	bottom-up	2005-2012	This thesis <sup>b</sup>
20.7(1.6)	top-down	2005-2012	This thesis <sup>c</sup>
~ 80	bottom-up	1980-2010	Sindelarova et al. (2014) <sup>d</sup>
26-94	bottom-up	2005-2013	Bauwens et al. (2016) <sup>e</sup>
36	top-down	2005-2013	Bauwens et al. (2016) <sup>f</sup>

a: Standard deviation shown in parentheses.

b: GEOS-Chem with MEGAN diagnostics based on 3-hourly averages.

c: Based on daily peak emissions integrated over a sinusoidal daily curve.

d: MEGAN run using MERRA meteorology.

e: Range shown here based on three models, two which implement MEGAN.

f: OMI based top-down inversion.

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 3.16). 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 ozone, HCHO, and NO<sub>x</sub> (not shown). Trend results for isoprene are summarised in Table 3.2. The a priori 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 in the a posteriori emissions. 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  $\alpha$  is applied to each grid square seasonally and not changing each year, changing trends are not expected.

### 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 3.2.9. Outputs from the scaled GEOS-Chem run using the a posteriori are denoted by superscript  $\alpha$ . For example, column HCHO from GEOS-Chem before and after scaling are denoted  $\Omega_{GC}$  and  $\Omega_{GC}^\alpha$  respectively.

#### 3.3.2.1 Implications for HCHO

As a preliminary check on the GEOS-Chem output, simulated  $\Omega_{GC}$  and  $\Omega_{GC}^\alpha$  is compared to  $\Omega_{OMI}$  over January and February, 2005 in Figure 3.17. In every region,  $\Omega_{GC}^\alpha$  is

TABLE 3.2: Yearly trend in surface amounts (ppbvC) before and after scaling isoprene emissions.

Region	a priori	a posteriori
<b>Isoprene</b>		
Aus	<b>-.04</b>	<b>-.02</b>
SE	<b>-.15</b>	<b>-.07</b>
NE	<b>-.13</b>	<b>-.04</b>
Mid	<b>-.07</b>	<b>-.02</b>
SW	-.01	-.01
N	<b>-.06</b>	<b>-.02</b>

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

Units are ppbv Carbon rounded to two decimal places.

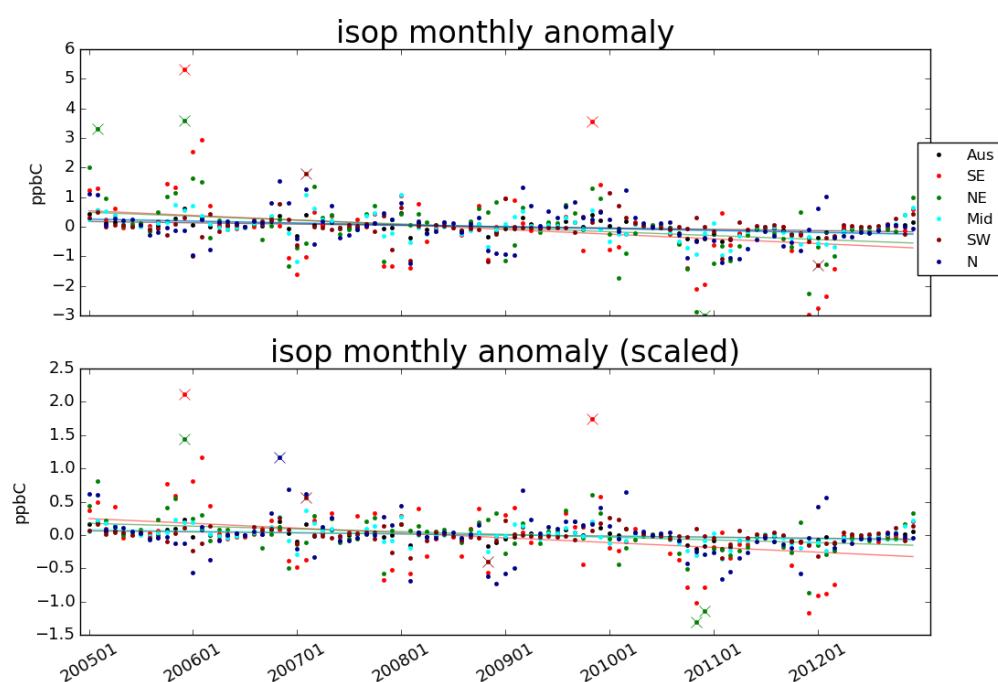


FIGURE 3.16: A priori (row 1) and a posteriori (row 2) emissions anomaly from multiyear monthly mean, split by region (see Figure 3.9).

closer to  $\Omega_{OMI}$  with biases decreasing from  $\sim 100\%$  to  $\sim 50\%$  everywhere except the northern region, which has biases decreasing from  $\sim 50\%$  to  $\sim 25\%$ . Note that this is not an independent validation as  $\Omega_{OMI}$  drive the creation of  $\Omega_{GC}^\alpha$ , 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  $\alpha$ .

Figure 3.18 shows HCHO is most reduced in summer for the majority of Australia. In winter, a reduction is seen along the northern edge of Australia, most likely due to because how emissions in this area are less affected by the seasonal decline. Seasonal means and the standard deviations for each region are summarised in Figure 3.19. Both mean and variance within the vertical column of HCHO are reduced after running GEOS-Chem with scaled isoprene, although the mean is lower still in OMI vertical columns. Model output standard deviations within summer range from  $\sim 20 - 30\%$ , while OMI standard deviations range from  $\sim 32 - 41\%$ . The highest OMI standard deviations occur in winter, ranging over  $\sim 46 - 73\%$ , which is the opposite of model output with standard deviations between  $\sim 12 - 27\%$ . This could be an effect from the increased winter uncertainty in satellite output, which is amplified by the low column amounts in the season. Overall the standard deviation within model output appears to be too low in all regions, ranging from three quarters (south eastern region) to one fifth (middle region) of the standard deviations within the OMI HCHO column output

Decreasing isoprene emissions in the model lead to reduction in HCHO concentrations as one would expect. Figure 3.20 shows an example (over Wollongong) of the modelled HCHO profile before and after scaling isoprene emissions. This reduction of HCHO is greatest at the surface, and is highly correlated ( $r > .8$ ) in all regions to the reduction in isoprene emissions. The high correlation is effectively a sanity check of the top-down method used in this thesis to estimate isoprene emissions.

### 3.3.2.2 Implications for ozone

Isoprene oxidation can eventually lead to ozone formation, especially when isoprene-enriched air masses mix with polluted urban air masses that contain high  $NO_x$ . Figure 3.21 shows surface level (up to  $\sim 150$  m altitude) ozone concentrations over 2005 before and after scaling modelled isoprene emissions. Reducing isoprene emissions lowers surface ozone concentrations by  $\sim 1$  ppbv in all regions in all seasons. The one exception to this reduction is in the northern region in spring, where  $\sim 3$  ppbv reductions are seen. A regression between the change in isoprene emissions, and the change in surface ozone was performed (as in the prior section dealing with HCHO); however, only very weak correlations between reductions are apparent. While the overall decrease in surface ozone is clear, there is no direct correlation between monthly grid square averaged reductions in isoprene emission and surface ozone concentrations. This suggests that changes in isoprene emissions affect ozone in grid squares where absolute emission reductions are not as strong.

*rephrase, makes it sound like it will be an increase*

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 (Millet et al. 2016). 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

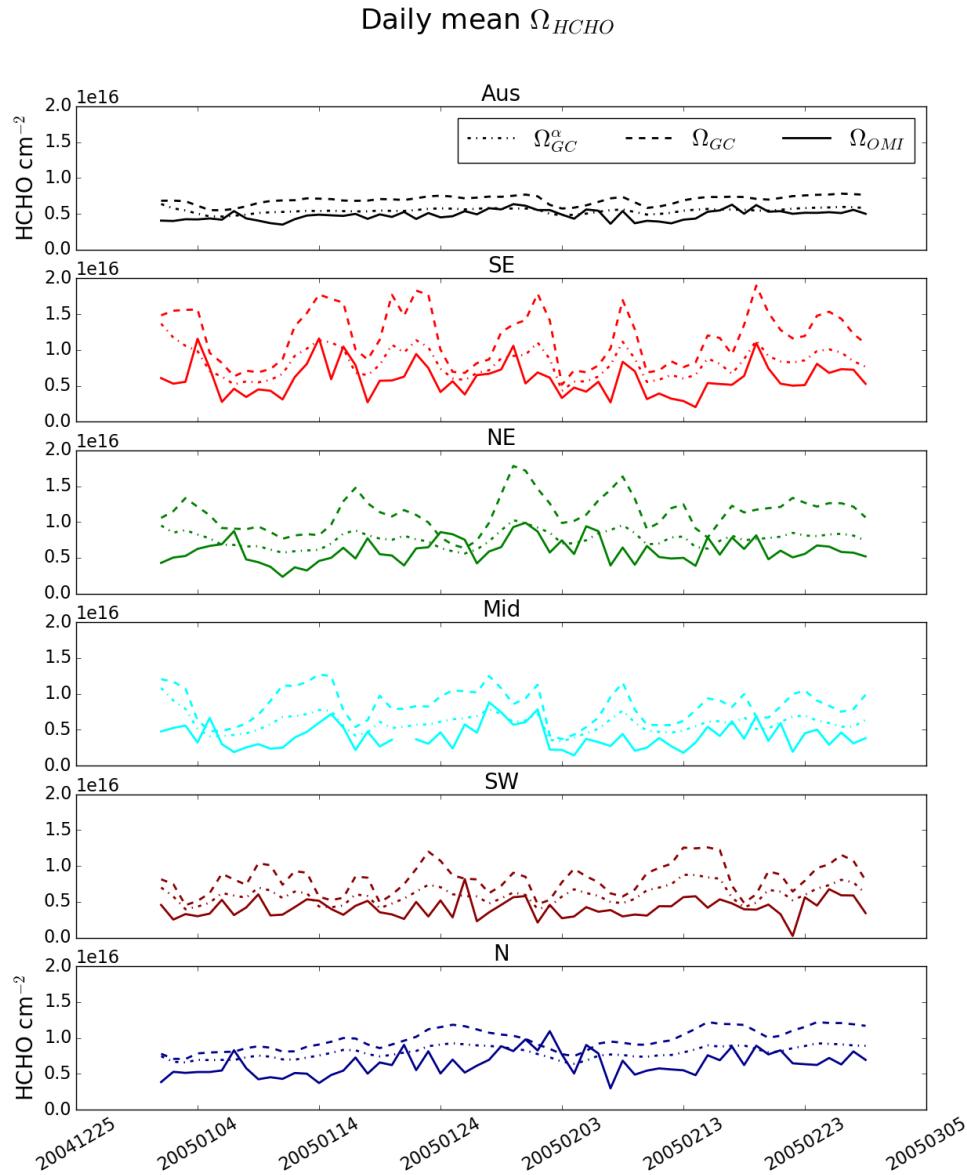


FIGURE 3.17: Daily mean total column HCHO amounts from GEOS-Chem with (new emissions run) and without (tropchem run) a posteriori scaled isoprene emissions, along with the recalculated OMI HCHO columns. Each row shows the average over regions in Figure 3.9.

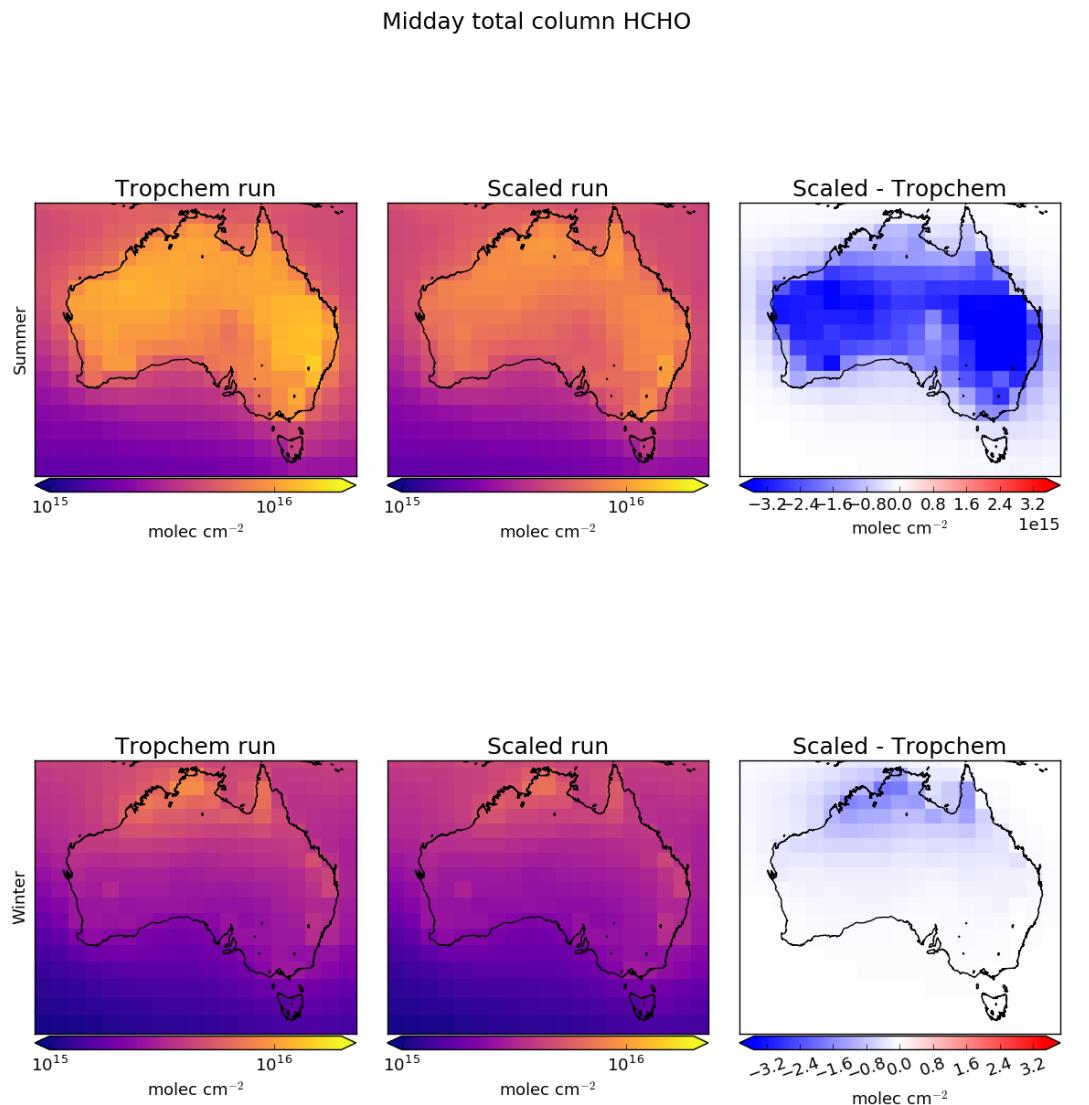


FIGURE 3.18: Total column HCHO before (left) and after (middle) scaling isoprene emissions, and their absolute differences (right). Top row shows summer (DJF) averaged total columns, while bottom row shows the winter (JJA).

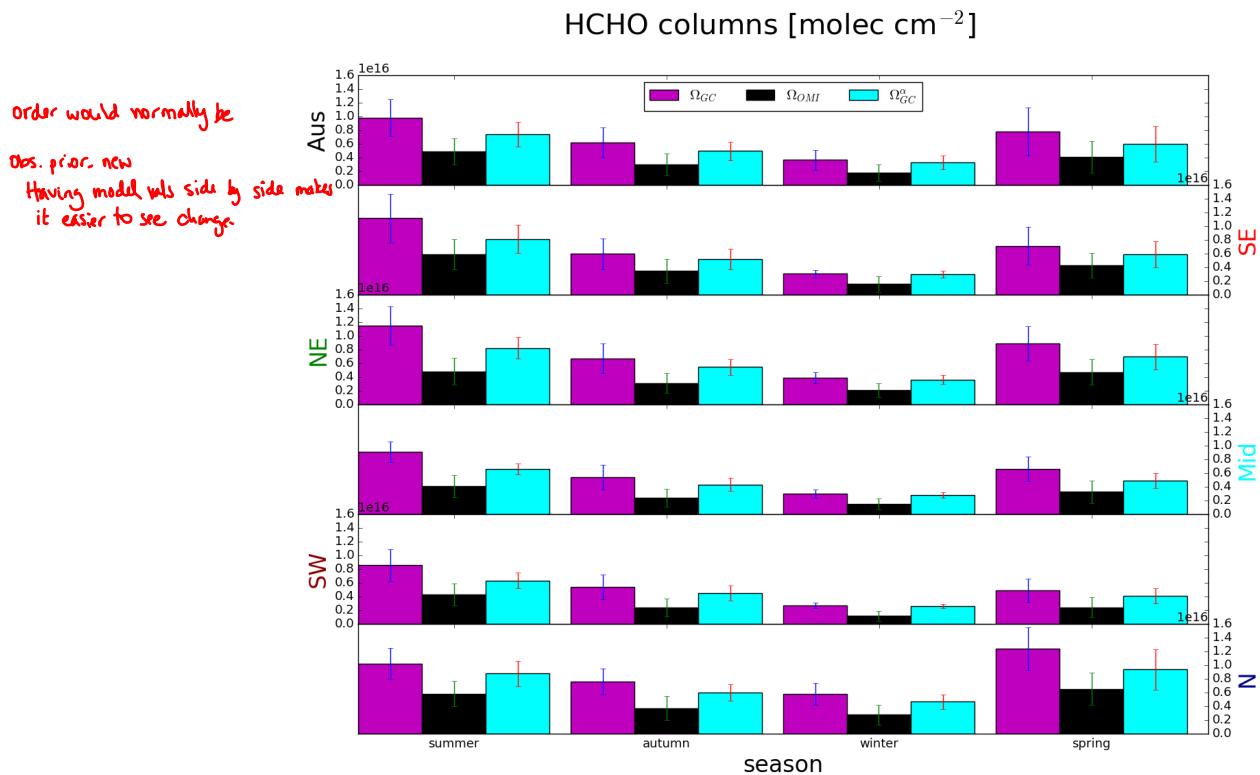


FIGURE 3.19: Regionally and seasonally averaged HCHO total columns from GEOS-Chem and recalculated OMI measurements side by side. Each row represents one region within Australia, while each column represents from left to right: summer, autumn, winter, spring.

Standard deviations are shown with error bars.

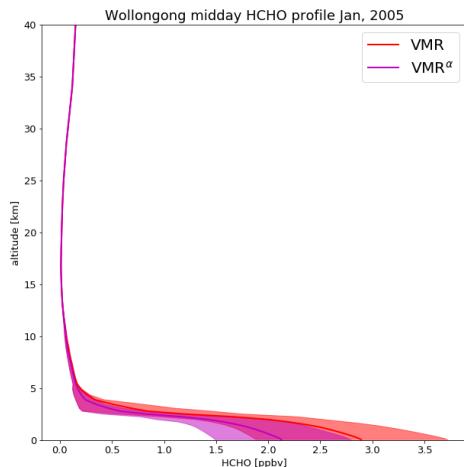


FIGURE 3.20: Monthly averaged HCHO profile over Wollongong modelled by GEOS-Chem before (VMR) and after ( $\text{VMR}^\alpha$ ) scaling isoprene emissions. Shaded areas represent the inter-quartile range over the month.

VMR

are mostly non-urban. This may play a part in explaining why ozone and fine particulate concentrations in Australian cities have not reduced over the last 10 years, unlike other atmospheric pollutants such as CO, NO<sub>2</sub>, and SO<sub>2</sub> (Keywood, Emmerson, and Hibberd 2016). Figure 3.22 shows how ozone is reduced in summer and winter after scaling isoprene emissions. In summer, reductions are strongest in Sydney and Melbourne, but strong reductions can also be seen in the west coast and central Australia. Winter reductions are more uncertain (relatively); however, they are strongest in the coastal and oceanic regions around Darwin.

Some ordering issues in this paragraph. Currently reads as if O<sub>3</sub> hasn't decreased because of an average! Also might be a good place to note that future work will downscale your emissions to evaluate impacts on O<sub>3</sub> when spatial segregation of NO<sub>x</sub> and isoprene are properly accounted for. See Yu et al. ACP 2016 about resolution dependence.

### 3.3.3 Comparison with in situ measurements

Comparison between ground-based measurements and large ( $2^\circ \times 2.5^\circ$ ) averaged model output suffers from representation error. Figure 3.23 shows the SPS and MUMBA measurement sites, along with the extent of the  $2^\circ \times 2.5^\circ$  relevant GEOS-Chem grid box, <sup>a</sup> A rectangle with edge lengths of roughly 200 km<sup>2</sup>. 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 which averages modelled values within the grid box. Due to high uncertainty in components of the top-down emissions estimate, temporal resolution is also limited. MUMBA (Section 2.2.3.1), SPS1 and SPS2 (Section 2.2.3.2) each provide on the order of one month of hourly or daily data, which are compared in this section against surface level concentrations from GEOS-Chem before and after scaling the biogenic emissions.

Figure 3.24 shows GEOS-Chem output in the grid square containing Sydney and Wollongong campaign measurements. The measurements are subset to those taken between 13:00 and 14:00, which are averaged on days when more than one exists.

↳ at what temporal resolution?

data point in this time window

technically not quite right – as written sounds like there are multiple model values in that box but really inputs are averaged and model computes one value for the box

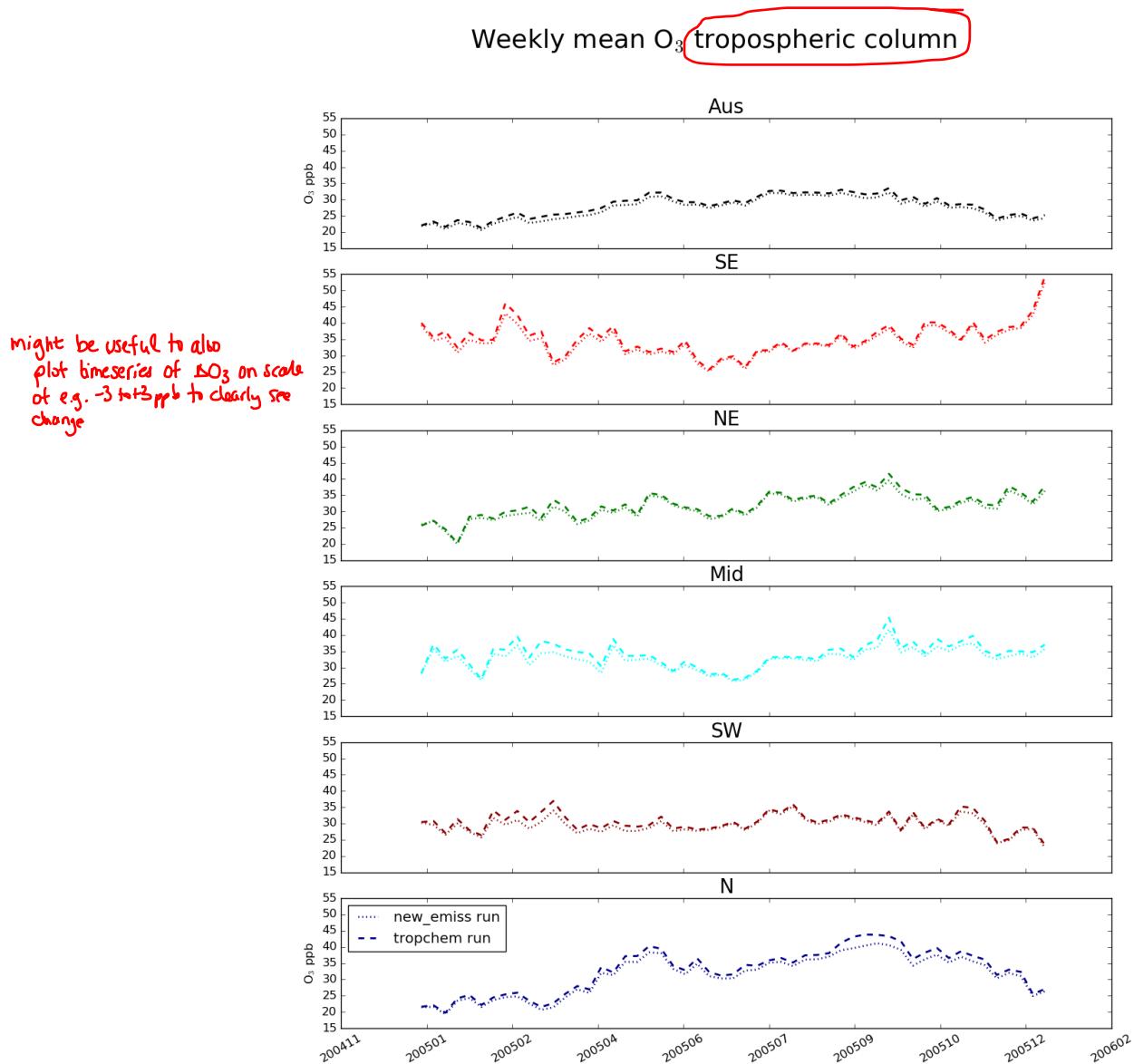


FIGURE 3.21: Surface ozone concentrations (ppb) per region over 2005.  
*Column or surface? Make sure both match!*

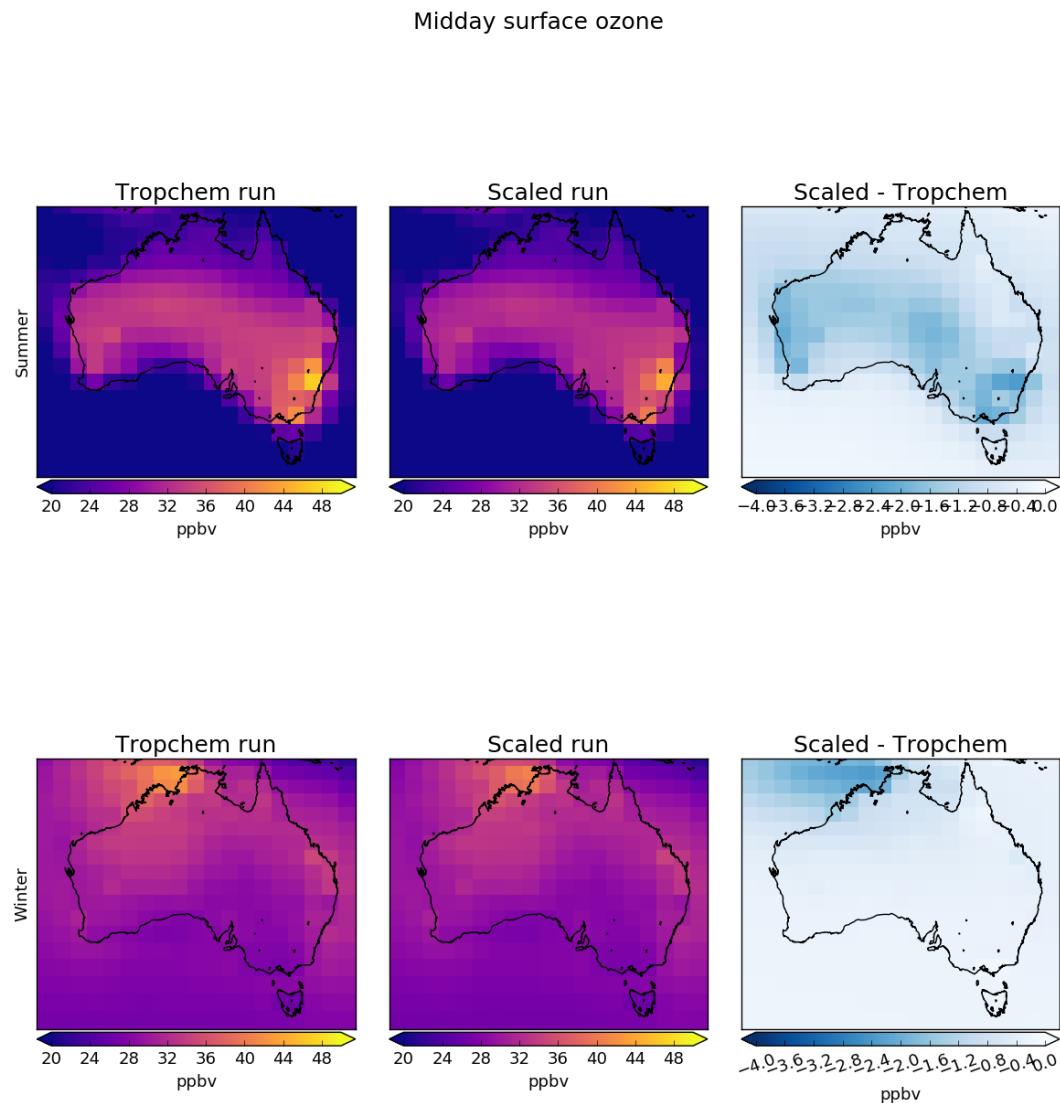


FIGURE 3.22: Surface (up to  $\sim 150$  m) ozone before (left) and after (middle) scaling isoprene emissions, and their absolute differences (right). Top row shows summer (DJF) averaged total columns, while bottom row shows the winter (JJA).

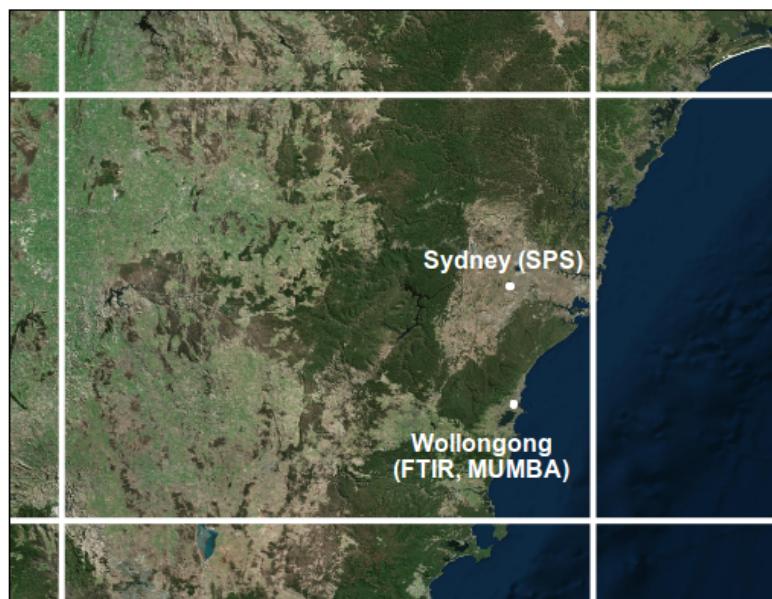


FIGURE 3.23: GEOS-Chem grid box ( $2^\circ \times 2.5^\circ$ ) containing Wollongong FTIR, SPS, and MUMBA campaign data.

TABLE 3.3: Mean total column HCHO amounts in  $10^{15}$  molec cm $^{-2}$ .

Season	$\Omega_{GC}$	$\Omega_{FTIR}$	$\Omega_{GC}^a$
summer	17.7	15.4	13.6
winter	4.41	3.68	4.49

lower than what?

Wollongong?

MUMBA measurements (summer, Sydney) show lower isoprene and HCHO due to the lack of nearby forest, and lower ozone due to titration. SPS measurements are closer to the GEOS-Chem outputs than MUMBA, suggesting the model grid square is more representative of an environment similar to Wollongong, with some forest and limited city influences. Variation in SPS data is higher than modelled, which is likely due to a dependence on local meteorology, as plumes of HCHO or isoprene enriched air can be detected as they pass by while the modelled grid square averages these out.

A spectrometer (FTIR) on the roof of the Chemistry building at Wollongong University measures HCHO in the atmosphere during clear sky conditions. This is the only non-satellite long-term measurement record of total column HCHO available in Australia. In order to compare modelled profiles against retrievals from the FTIR, modelled profiles are first convolved with the instrument averaging kernel and a priori as shown in Section 2.2.3.4. FTIR output is resampled to only include measurements taken at midday (13:00–14:00) and GEOS-Chem overpass outputs are interpolated onto matching vertical levels for days where FTIR output exists. Figure 3.25 shows total column HCHO from FTIR, and GEOS-Chem before and after scaling isoprene. While the representational error between FTIR measurements and GEOS-Chem output prevents quantitative analysis, one can see that the summer overestimation of HCHO from GEOS-Chem is removed by isoprene scaling. The mean summer underestimation of HCHO shown after scaling makes more sense due to the relatively dense forested areas around Wollongong, which would raise local HCHO concentrations above the average for the large grid square represented by GEOS-Chem. Another feature is the January and December dip in all three data sets. FTIR has a lower data count and higher variance in these months due to holidays, increasing uncertainty for the measurements. Analysis of this dip could be performed in future work; however, it would require an in depth examination of local and synoptic HCHO patterns, and higher resolution model data.

Using FTIR measurements as a measure of potential model bias over Wollongong by looking at the mean difference with (co-located, convolved, and resampled) model output is summarised in Table 3.3. The modelled summer vertical column is higher than the FTIR measurements by  $\sim 15\%$ , which decreases with isoprene scaling to  $\sim -12\%$ . In winter the column is modelled high by  $\sim 20\%$ , which increases to  $\sim 22\%$  after scaling. This is only qualitative as Wollongong measurements are not representative of the entire grid box which GEOS Chem averages, and more measurements would be required to estimate model bias over any large grid square.

I am really confused here.  
MUMBA=Wollongong  
SPS=Sydney

You imply opposite so I'm not sure what you are trying to say!

Also need some comment on how the updated emiss. perform relative to the original  
Some quantification would be good e.g. mean bias, correlation

Comparison suffers from

again would like to see bias quantified and maybe correlation—is there a reason you don't show the timestack of monthly means?

clarify meas. involves manual process.

remove italics

Seasonal mean HCHO from the

You've already said this ad nauseum; just integrate these two paragraphs or add bias statistics here.

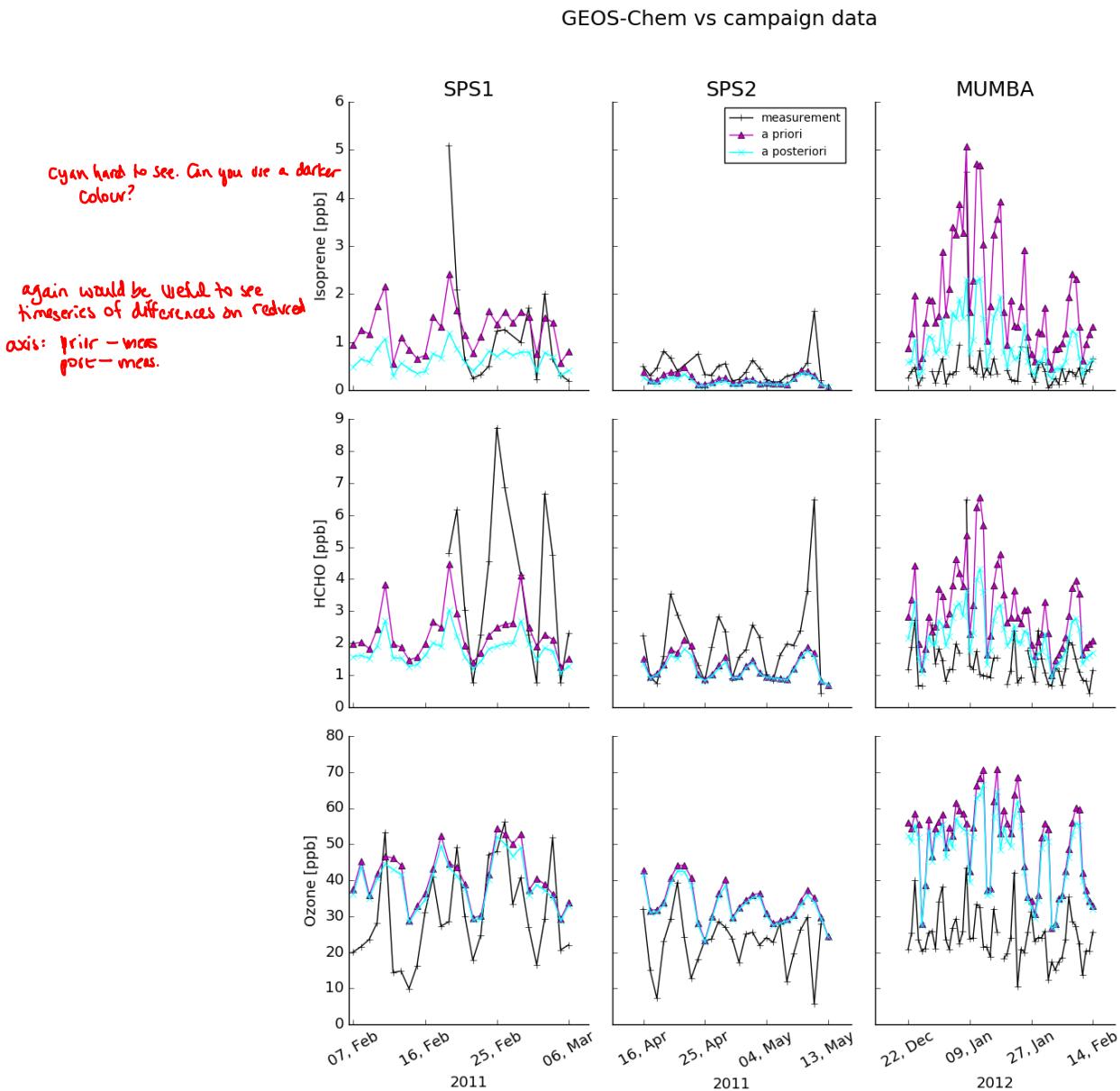
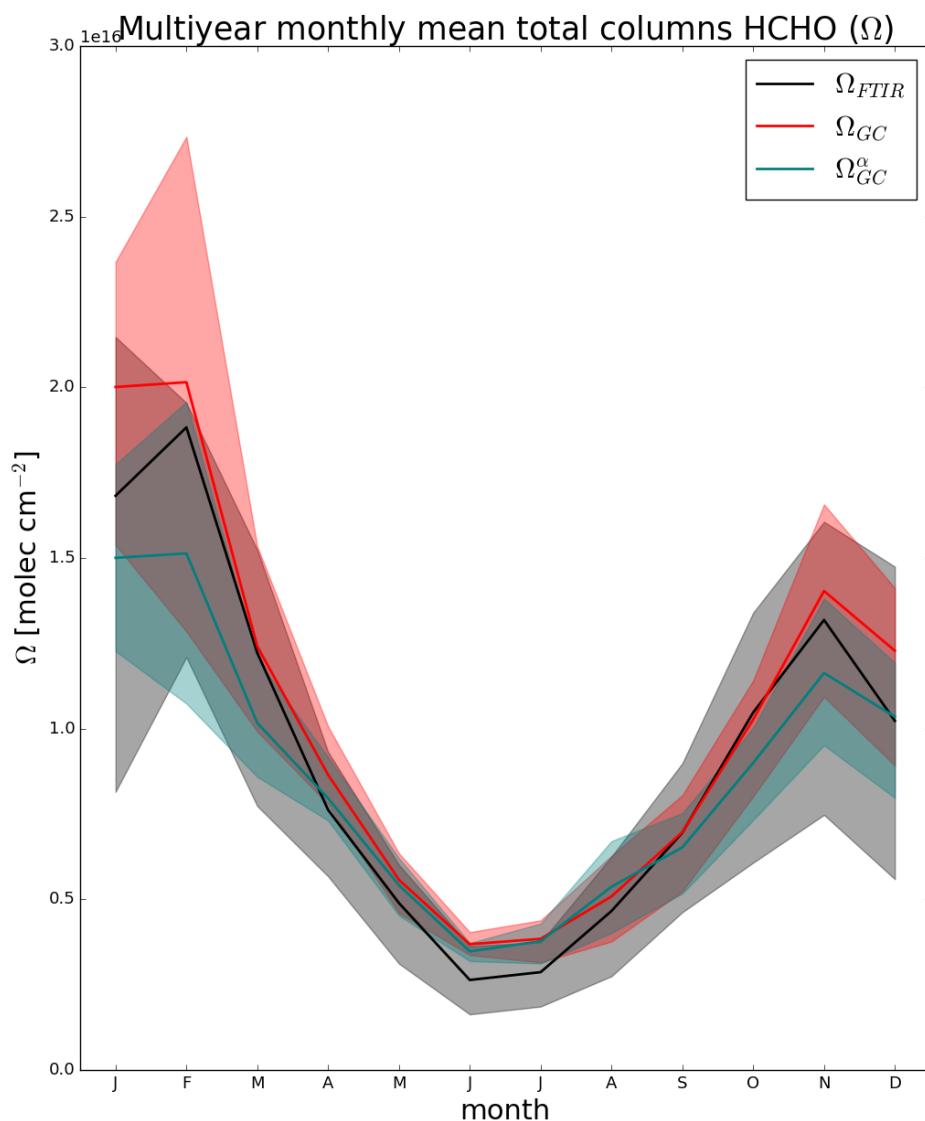


FIGURE 3.24: SPS1, SPS2, and MUMBA (left to right columns respectively) midday (13:00–14:00 local time) measurements of isoprene, HCHO, and ozone (top to bottom respectively). Shown in magenta and cyan respectively are the a priori and a posteriori GEOS-Chem surface outputs for the matching grid square at midday for months containing measurements.

days?



which years? same for obs & model?

FIGURE 3.25: Multi-year monthly mean total column ( $\Omega$ ) HCHO from the FTIR instrument, and the colocated convolved GEOS-Chem equivalent before ( $\Omega_{GC}$ ) and after ( $\Omega_{GC}^\alpha$ ) scaling isoprene emissions. Shaded areas show inter quartile range.

This section doesn't really fit here.  
Seems it would either go in  
methods or Uncertainties.

### 3.3.4 Modelled HCHO lifetimes

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:

$$\begin{aligned} S &= \frac{Y_{isop}}{k_{HCHO}} \\ \tau &\equiv \frac{1}{k_{HCHO}} \\ \tau &= \frac{S}{Y_{isop}} \end{aligned} \quad (3.10)$$

$\tau$  is heavily influenced by assumed yield, and improved methods of estimating yield over Australia are required to improve this estimate. Figure 3.26 shows the GEOS-Chem HCHO lifetime estimated throughout the year. There is a clear seasonal cycle with longer lifetimes in winter months. In June (and sometimes March, July and August), HCHO lifetimes increase, 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; however, outliers along with low data availability (in some months) are issues which lead to wide estimate ranges and high uncertainty. 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.

## 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 against measurements. Even as the top-down inversion performed in this chapter attempts to work around 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 become dominant (see Section 3.4.3). Monthly calculated slope uncertainty mostly lies within 30% to 50%, and since output is averaged to form monthly values, the slope uncertainty is not reduced by the 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 with latitudes and during winter. Table 3.4 shows the estimated uncertainty calculated in this work in summer and winter over each region described by Figure 3.9. The assumptions and calculations made to determine uncertainties in the top-down estimate ( $\Delta E_{OMI}$ ), the satellite column ( $\Delta \Omega_{OMI}$ ) and the slope ( $\Delta S$ ) are described in the following subsections.

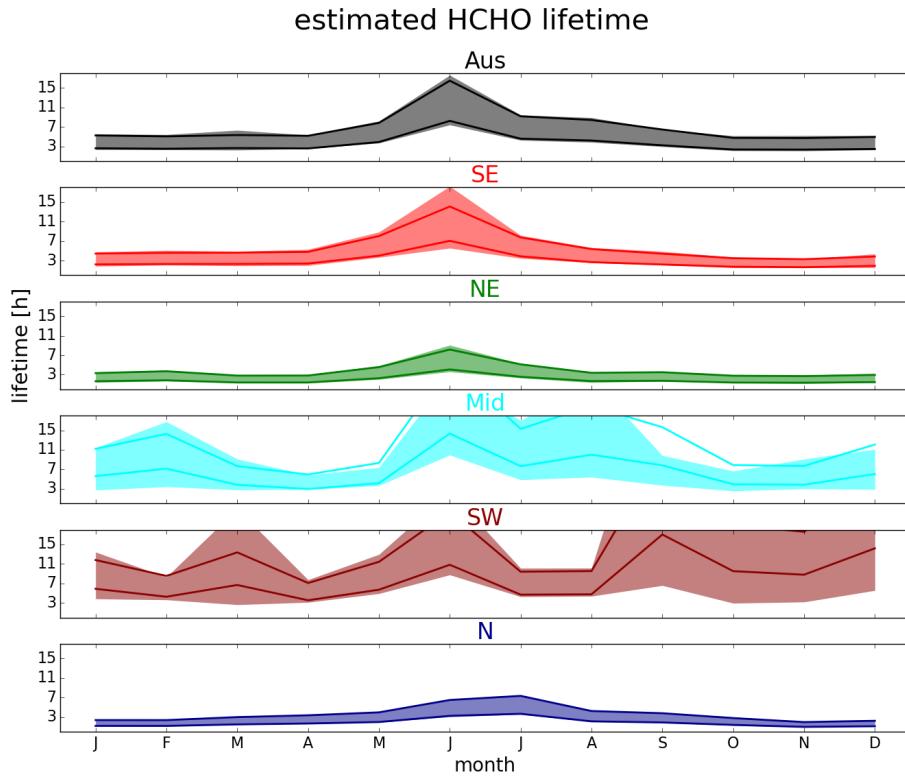


FIGURE 3.26: Monthly area averaged HCHO lifetime ( $\tau$  in hours), with IQR shaded. Solid lines show lifetime assuming yield is 0.2, and 0.4 (higher and lower lines respectively). Coloured by regions shown in Figure 3.9.

TABLE 3.4: Relative uncertainty estimates.

Region	$a \frac{\Delta E_{OMI}}{E_{OMI}}$	Summer			Winter		
		$\frac{\Delta \Omega_{OMI}}{\Omega_{OMI}}$	$\frac{\Delta S}{S}$	$a \frac{\Delta E_{OMI}}{E_{OMI}}$	$\frac{\Delta \Omega_{OMI}}{\Omega_{OMI}}$	$\frac{\Delta S}{S}$	
Aus	46%	21%	36%	51%	258%	35%	
SE	54%	14%	37%	66%	141%	38%	
NE	53%	18%	35%	51%	40%	37%	
Mid	45%	38%	38%	47%	41%	35%	
SW	43%	27%	32%	61%	96%	36%	
N	39%	16%	33%	38%	22%	30%	

a: Grid squares with monthly uncertainty over 200% are removed when calculating the mean uncertainty, which has small  $< 1\%$  impacts on  $E_{OMI}$  overall, but removes many data points in winter.

### 3.4.1 Top down emissions

Important factors in the calculation of isoprene emissions using OMI HCHO include the modelled relationship between HCHO and isoprene, and 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.

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  $E_{OMI}$ :

$$z = x + y : \Delta z = \sqrt{(\Delta x)^2 + (\Delta y)^2} \quad (3.11)$$

$$z = x/y : \Delta z = z \sqrt{\left(\frac{\Delta x}{x}\right)^2 + \left(\frac{\Delta y}{y}\right)^2} \quad (3.12)$$

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

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

$$\Delta\Phi = \sqrt{(\Delta\Omega_{OMI})^2 + (\Delta\Omega_{OMI,0})^2} \quad (3.13)$$

$$\Delta E_{OMI} = E_{OMI} \times \sqrt{\left(\frac{\Delta\Phi}{\Phi}\right)^2 + \left(\frac{\Delta S}{S}\right)^2} \quad (3.14)$$

$\Delta E_{OMI}$  is calculated using the uncertainty in underlying terms:  $\Delta S$ ,  $\Delta\Omega_{OMI}$ , and  $\Delta\Omega_{OMI,0}$ . For  $\Delta S$  ( $\Omega_{GC} = S \times E_{GC} + \Omega_{OMI,0}$  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 3.27 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 3.28 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) can be seen, through the increase in uncertainty with increasing latitude.

### 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 ( $\Omega_{GC}$ ) and emissions ( $E_{GC}$ ) is used to estimate  $\Delta S$  in

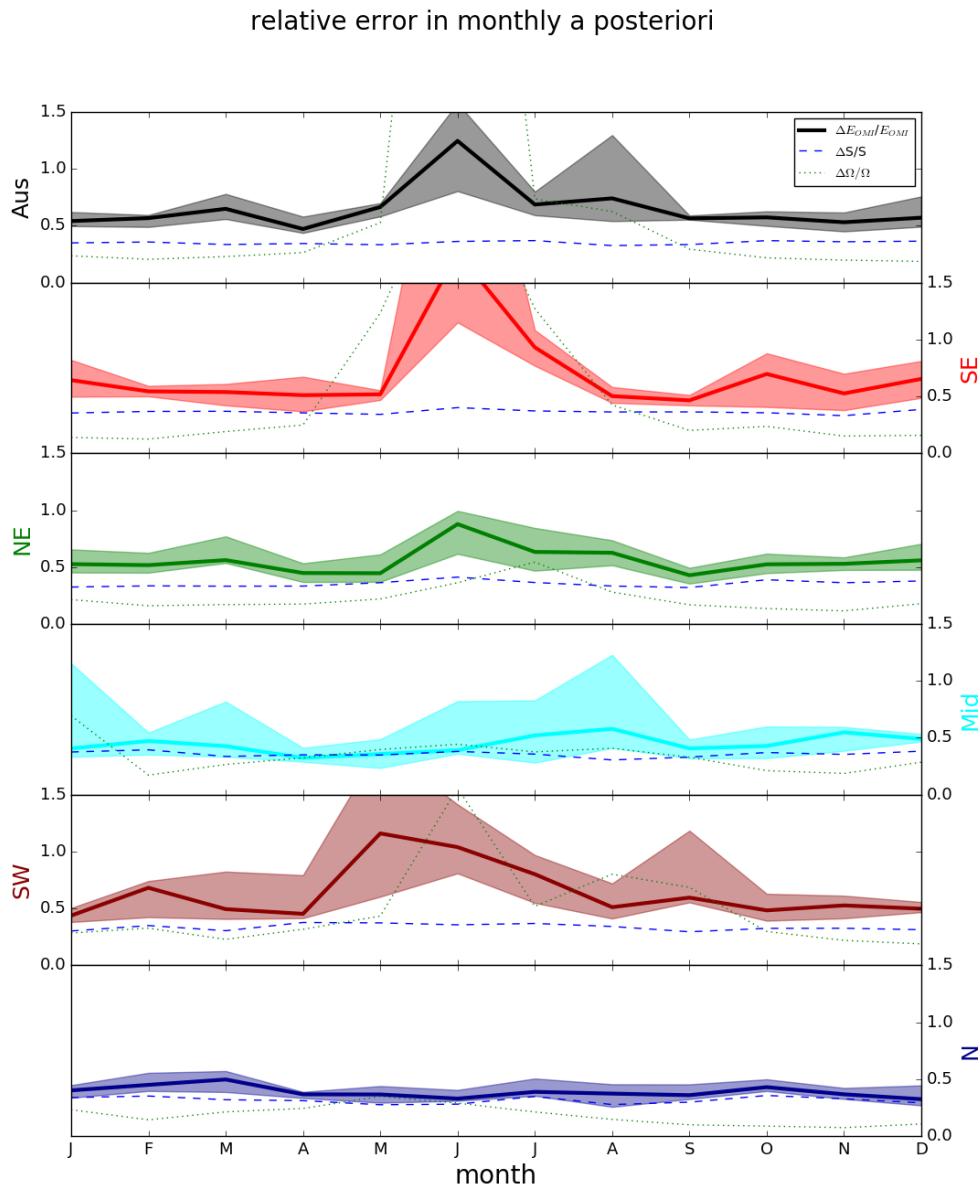


FIGURE 3.27: Median and inter-quartile range of multi-year monthly relative uncertainty in the a posteriori. Median relative uncertainty in  $S$  and  $\Omega$  are added as dashed and dotted lines respectively.

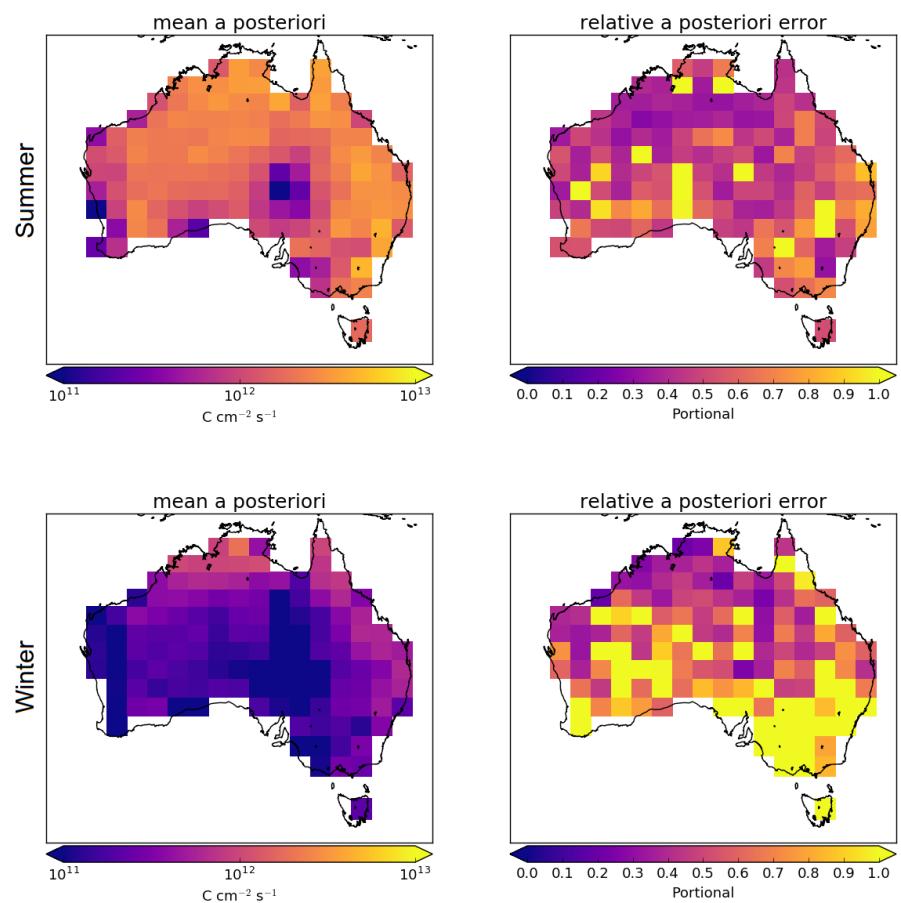


FIGURE 3.28: Summer (DJF, top row) and winter (JJA, bottom row) a posteriori emissions (left column) and relative error (right column).

Equation 3.14. Here I use the ratio of the upper bound of the 95% confidence interval ( $CI_{UB}$ ) over  $S$  to represent the relative uncertainty.

$$\frac{\Delta S}{S} = \frac{CI_{UB}}{S} - 1 \quad (3.15)$$

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

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, which would be difficult to quantify. Figure 3.29 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 comparing with another chemical transport 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 when compared to measurements.

Filtering for spatial smearing (see Section 3.2.8) reduces the number of data points making up the regression slope  $S$ . The process generally improves the linear relationship, which suggests it is working as intended to remove days when local biogenic emissions are not driving HCHO enhancement. Where  $S$  regressions have a correlation coefficient ( $r$ ) of less than 0.4, 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 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  $\text{NO}_x$  seasonality and regional concentrations, is required. The filtering process is accounted for in the calculation of  $\Delta S$ , and the effect is a small reduction of the overall uncertainty from  $S$ .

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 quantify uncertainty at a national scale.

### 3.4.3 Satellite Uncertainty

Corrected vertical columns of HCHO from the OMI product are calculated using Equation 2.26:  $\Omega = \frac{SC - RSC}{AMF}$ . 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.

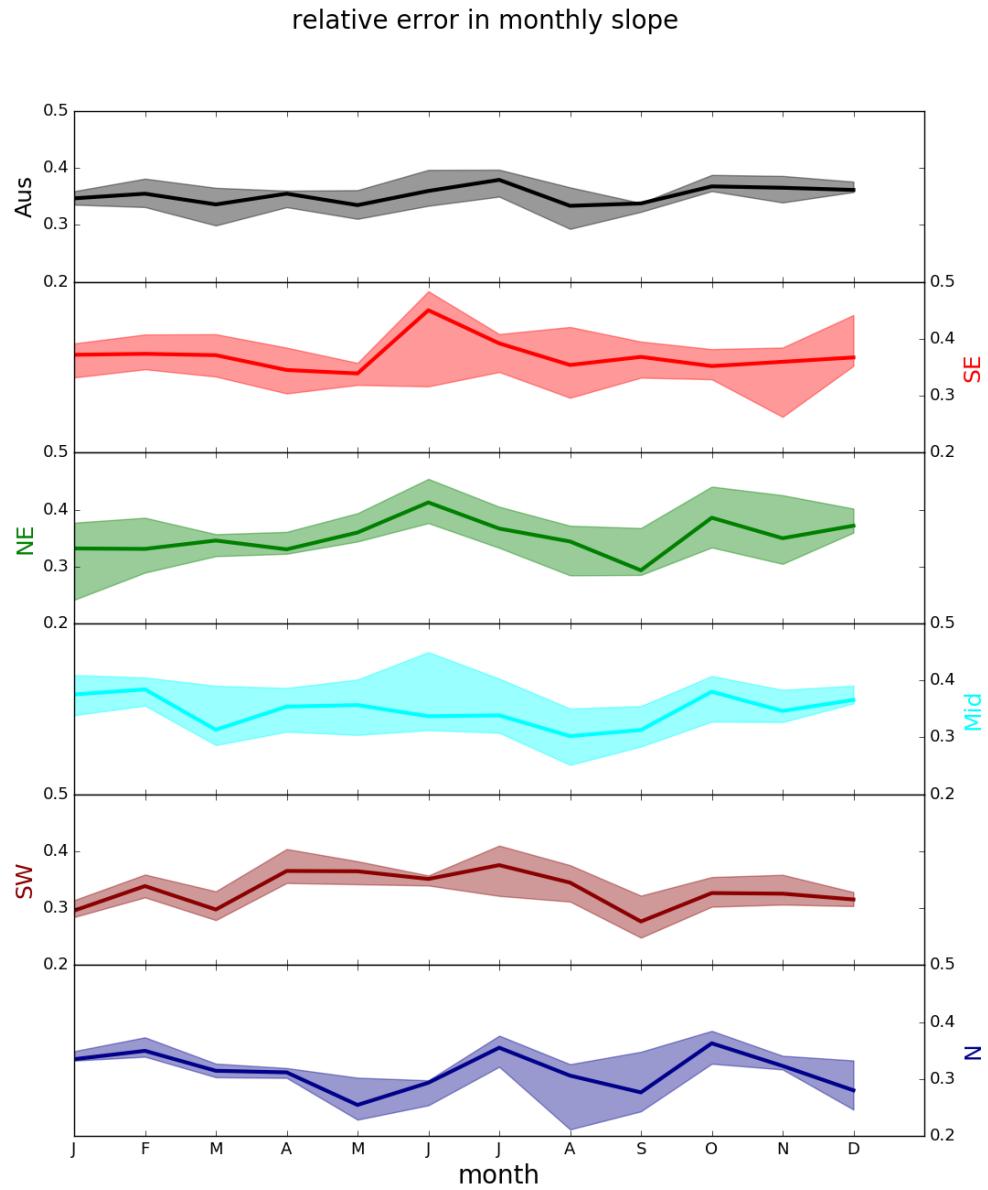


FIGURE 3.29: Median and inter-quartile range of monthly binned uncertainty in  $S$ .

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  $\Delta SC$  in Equation 3.16. 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).

### Uncertainty in AMF calculations

Air mass factors model the satellite instruments vertical sensitivity to slant path measurements, and uncertainties arise predominantly from uncertain cloud parameters (Palmer et al. 2006). Vertical columns from OMI are recalculated using AMFs derived from GEOS-Chem (Section 2.6.3). 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. For comparison, Wollongong FTIR is over-predicted by GEOS-Chem modelled vertical columns by  $\sim 15 - 20\%$ . Here the assumed error of 30% is ascribed to the *AMF* term.

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 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 ( $1\sigma$ ) of  $25 - 27\%$  in HCHO vertical columns with calculated AMFs where cloud fraction  $< 0.2$ .

### 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.

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  $\Delta RSC$  is set to the standard deviation of the *RSC* over Australian latitudes ( $45^{\circ}\text{S}$  to  $10^{\circ}\text{S}$ ) in all tracks. These terms are used in Equation 3.16. Background error calculated in this way is on the order of 5%-10% after monthly averaging, however this error increases in the higher southern latitudes during autumn and winter to  $\sim 15\%$ . For comparison, the background error is assumed to be 15% in Curci et al. (2010) following (Dufour et al. 2008).

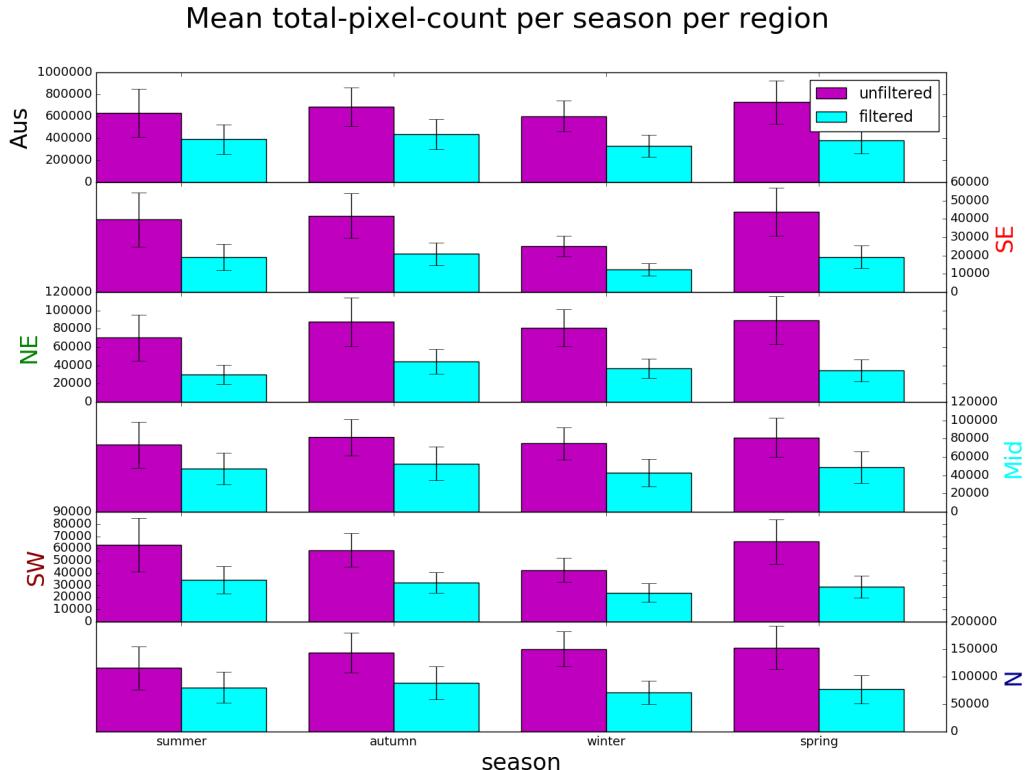


FIGURE 3.30: Mean and standard deviation (vertical error bars) of total pixel counts per region per season, before (magenta) and after (cyan) applying smearing, pyrogenic, and anthropogenic filters.

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 each  $2^\circ \times 2.5^\circ$  grid square for each day or month. For example, daily averaging reduces pixel uncertainty by a factor of 2-4. Figure 3.30 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 May, June, and July. The lowest pixel counts occur in the southeast in winter, likely due to a mix of relatively high solar zenith angles and filtering of 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 3.11 and 3.12. Error in the slant column ( $\Delta SC$ ) is combined with our assumed relative AMF error ( $\frac{\Delta AMF}{AMF}$ ) of 30%, and background

TABLE 3.5: Uncertainties in satellite total column HCHO.

uncertainty	location	notes
40%	North America	<sup>a</sup> GOME, mostly due to cloud interference <sup>a → footnotes @ end rather than start.</sup>
26%	North America	<sup>b</sup> GOME, OMI, with cloud fraction less than 20%
30%-40%	global	<sup>c</sup> GOME-2
> 60%	Mid-latitude	<sup>c</sup> GOME-2 in winters
1%-10%	Australia	<sup>d</sup> OMI, monthly uncertainty at $2^\circ \times 2.5^\circ$
50%-100+%	Australia	<sup>d</sup> OMI, 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)

d: This work

error to calculate  $\Delta\Omega_{OMI}$  (and  $\Delta\Omega_{OMI,0}$ ) as follows:

$$\begin{aligned} \Delta(SC - RSC) &= \sqrt{(\Delta SC)^2 + (\Delta RSC)^2} \\ \Delta\Omega &= \Omega \sqrt{\left( \frac{\Delta(SC - RSC)}{(SC - RSC)} \right)^2 + \left( \frac{\Delta AMF}{AMF} \right)^2} \\ \frac{\Delta\Omega}{\Omega} &= \sqrt{\frac{(\Delta SC)^2 + (\Delta RSC)^2}{(SC - RSC)^2} + \left( \frac{\Delta AMF}{AMF} \right)^2} \end{aligned} \quad (3.16)$$

The  $RSC$  term is described in Chapter 2 Section 2.6.6  $\Delta RSC$  for each latitude is based on the standard deviation of corrections over the remote Pacific. Negative columns can occur where 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 impacts the uncertainty calculations in winter for the non-northern regions, where occasional highly negative absolute uncertainty was seen when  $\Omega$  approached 0.

Figure 3.31 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 3.5. 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 found an excess of 60% uncertainty. OMI measurements will have similar 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  $\sim 40\%$ . For example OMI underestimated aircraft measurements by 37% in Guyana (Barkley et al. 2013). OMI underestimates range from 20-37% when compared against aircraft

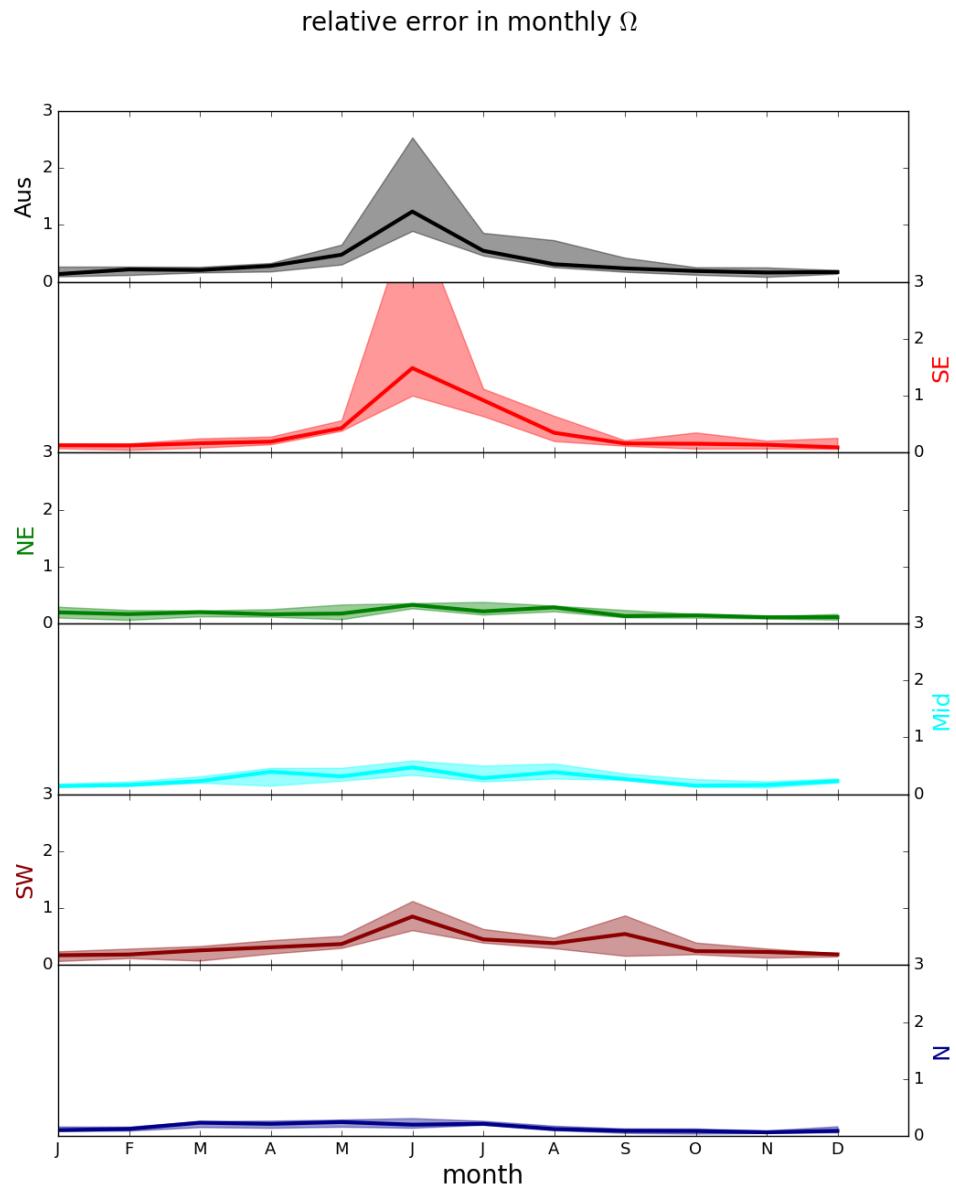


FIGURE 3.31: Median and inter-quartile range for monthly binned relative uncertainty in satellite vertical columns.

data over the southeast United States (Zhu et al. 2016)). OMI validation against 7 sites (the most southern site being Reunion Island at 20.9°S) using MAX-DOAS or FTIR retrievals showed up to 50% underestimation by satellite HCHO retrievals (De Smedt et al. 2015). The highest underestimate occurred during periods of high concentration; however, there was less underestimation when compared with the FTIR instrument which was at the southernmost site (Vigouroux et al. 2009; 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. The conclusion drawn here is that the isoprene emissions product in this work may be affected by satellite HCHO underestimation of up to 40%, and also by monthly HCHO overestimation of 13%, which gives a potential bias of 1/0.6 to 1/1.13. 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 cannot be quantified and so are not included in this thesis. Lacking suitable measurements to estimate satellite bias over Australia, the potential bias range is applied to the mean emissions in each month when calculating the uncertainty range shown in the results (see Figure 3.11).

#### 3.4.4 Sensitivity to AMF recalculation

This section examines 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 3.32 shows AMFs and emissions recalculated differently over 2005. This figure only depicts grid squares with non-zero emissions, since we are interested in regions where emissions are stronger. The fully recalculated  $AMF_{PP}$  is higher in summer, but lower in other seasons compared against the other recalculations. The direct effect of a lower AMF is an increased vertical column, which should lead to higher emission estimates, and vice versa. This can be seen in all seasons except for summer, although even here the emissions estimates are higher than if they are calculated using the original  $AMF_{OMI}$ . Emission estimates vary widely over Australia, but the sensitivity to AMF recalculation technique is proportional and potentially non-linear. Changes in AMF of 5 – 30% cause changes in emission estimates of 5 – 50%. Further analysis is warranted and may be a focus in future work.

#### 3.4.5 Sensitivity to filtering

Figure 3.33 shows emissions estimates with and without filtering for smearing, anthropogenic, and pyrogenic influences. The overall effect of filtering is to slightly raise emissions in all non-summer months, with relatively little change to the mean in winter months. This is true in all regions except for northern Australia, which shows a slight decrease in summer. Eastern regions are most frequently filtered, with more

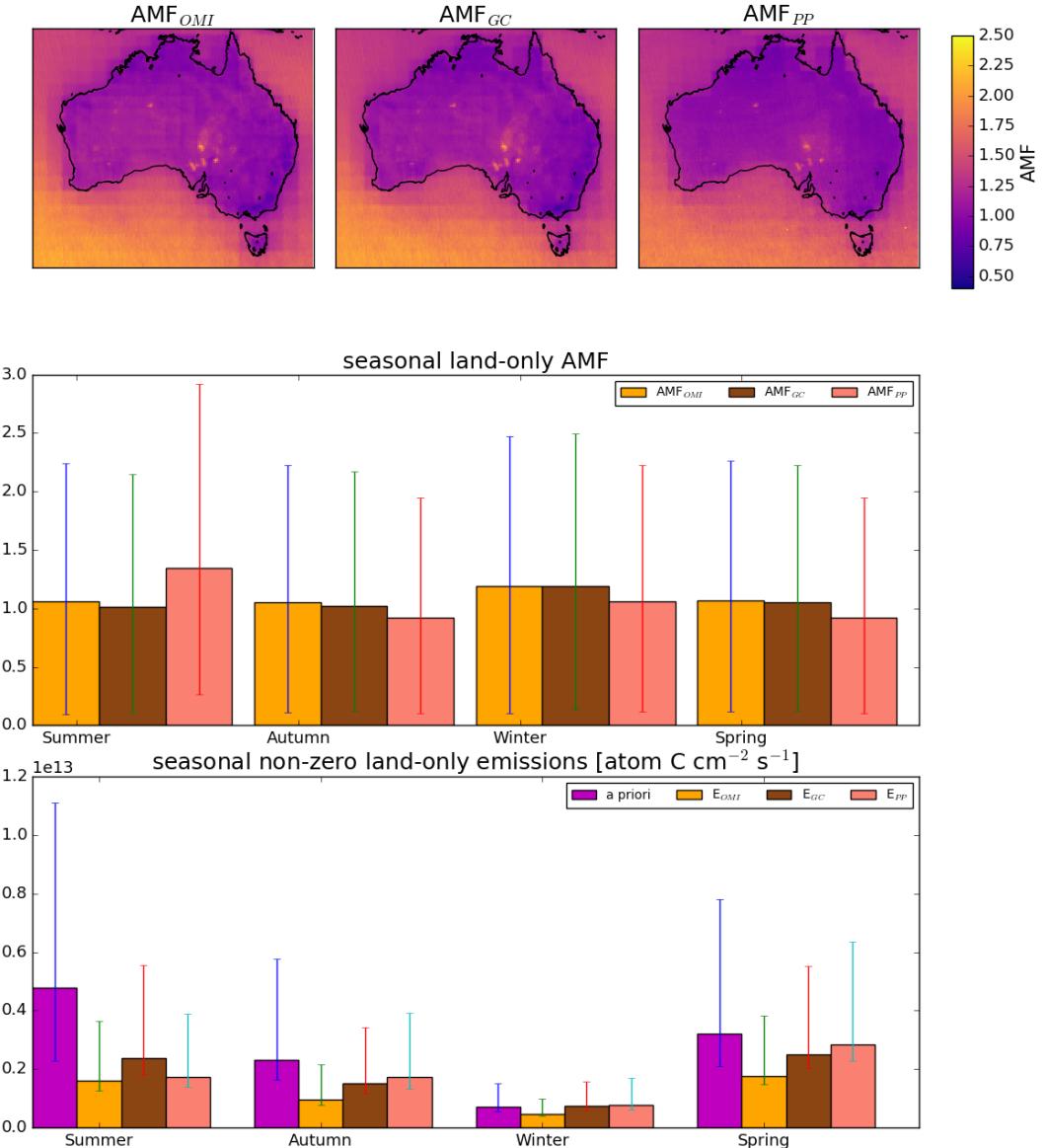


FIGURE 3.32: Top row: averaged OMI Satellite AMF for 2005, from the OMHCCHO data set (left,  $AMF_{OMI}$ ), recalculated using GEOS-Chem shape factors (middle,  $AMF_{GC}$ ), and recalculated using GEOS-Chem shape factors and scattering weights (right,  $AMF_{PP}$ ). Middle row: mean and inter-quartile range over 2005 for each season. Bottom row: mean and inter-quartile range of non-zero emissions based on the three recalculations (with matching subscripts) next to the a priori emissions from GEOS-Chem.

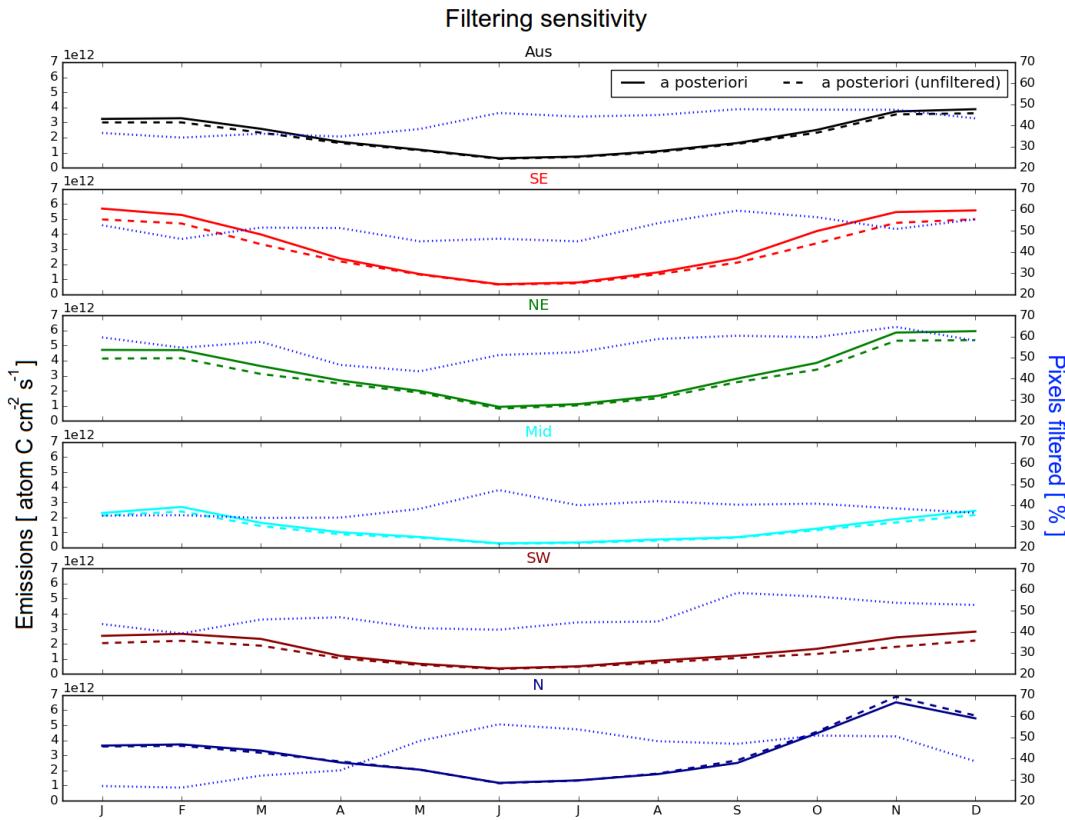


FIGURE 3.33: Multi-year monthly mean values for a posteriori emission estimates calculated with (solid) and without (dashed) applying filters for anthropogenic, pyrogenic, and smearing influences. The portion of pixels within each region which are filtered is shown on the right axis with a blue dotted line.

than 50% of the available good pixels filtered throughout the year, while other regions vary between 20% and 60% with lower filtering rates in summer months.

Each row shows a regionally averaged time series for emissions with (solid) and without (dashed) applying the anthropogenic and pyrogenic filters. Portion of good pixels filtered is also shown (dotted, grey) using the right axis.

### 3.5 Conclusions and implications

Very few ground based measurements of BVOC concentrations and emissions are available in Australia. Emissions from models use largely unverified extrapolations for emission factors in Australia, leading to overestimated isoprene emissions (Emmerson et al. 2016). This leads to uncertainty and error when modelling atmospheric ozone and other trace gases.

In this chapter I created and tested an isoprene emissions estimate (a posteriori) based on OMI satellite measurements of HCHO, a high-yield product of isoprene oxidation. The a posteriori emissions' effects on ozone and HCHO were tested by running

*calculated from the a posteriori*

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 emissions calculation ~~were~~ are calculated where possible, and potential biases ~~are noted~~ identified.

The a posteriori isoprene emission estimate showed that GEOS-Chem (running the MEGAN emissions model) overestimates emissions in summer by a factor of 2-5. Total yearly Australian emissions are reduced from  $39 \text{ Tg yr}^{-1}$  in GEOS-Chem to  $21 \text{ Tg yr}^{-1}$  (decrease of  $\sim 46\%$ ) in the a posteriori. The overestimate is spatially and temporally diverse and leads to model overestimates of both HCHO and ozone. In a sample of two months, running GEOS-Chem using scaled emissions based on OMI HCHO columns reduced the model HCHO overestimate (when compared to OMI HCHO) by half (from 50-120% down to 25-50% depending on the region). Model output vertical column HCHO variance is somewhat ( $\sim 10\% - 50\%$ ) lower than that seen by the OMI satellite, and this difference is increased after scaling isoprene emissions. Scaling GEOS-Chem emissions also lowered simulated surface ozone concentrations by  $\sim 5\%$ . A posteriori uncertainty was shown to be on the order of 50% (monthly, per grid square), with large satellite-based uncertainty in winter and a potential bias coming from satellite data of  $-13\%$  to  $40\%$ . 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 prohibitively high ( $> 100\%$ ).

The full reason for the a priori overestimate remains unclear, and both global and Australian emissions estimates for isoprene range widely. It is in part due to the MEGAN emission model's sensitivities to inexact parameters including leaf area indices, plant functional type emission factors, and meteorological factors. In Australia, a mixture of poorly defined emission factors (e.g., Emmerson et al. 2016), unaccounted for soil moisture (e.g., Sindelarova et al. 2014; Emmerson et al. 2019) and poorly understood forest responses to meteorological stresses is likely driving uncertainty and model biases (Jiang et al. 2018; Emmerson et al. 2019). A potential improvement to the MEGAN soil parameterisation has been described (Jiang et al. 2018); however, a detailed map of soil properties is not readily available in Australia, and this would be required to tune the parameterisation.

In the US, bias between OMI and in-situ measurements is as high as 40%, but bias across Australia cannot be determined as there are not enough independent observations. This leads to a wide range for potential emissions (see Figure 3.11) and limits potential top-down refinements to isoprene emissions. Ground-based VOC, NO<sub>x</sub>, and ozone measurements over large areas at relatively fine temporal resolution would help quantify unknown satellite biases while additionally providing constraints for bottom-up models. In the northern region forests are affected by monsoonal seasons, but and increased cloud coverage may limit representative satellite coverage. This makes characterisation of forest emissions and their response to sunlight, temperature, and moisture ~~are~~ even more important in these areas. In addition to ground measurements, further analysis ~~determining the sensitivity of~~ depicting how sensitive modelled emissions are to model resolution and changing soil moisture parameters (and parameterisations) would provide the foundation for how to update GEOS-Chem and MEGAN to improve modelled emissions along with child products like ozone.

HCHO and

*why only two months? I thought you ran the whole period! And which months? Summer?*

*don't think you can say that since you don't show improved agreement w/ O3.*

*I would think the parameters are better constrained than the MEGAN parameterisation of isoprene response to those parameters.*

*What did Emmerson 2019 do?  
Could future work be to apply her results here?*

In the future, other satellites (e.g., GOME-2) could be used to improve emission estimates ~~even~~ further, with differing overpass times potentially allowing a measure of diurnal emission patterns. Additionally, an adjoint version of GEOS-Chem could provide details of isoprene transport and ozone production sources. This version of GEOS-Chem could also provide an evaluation of how resolution-limited the linear top-down emission estimates are over Australia. The emission estimate created in this chapter could be refined to higher temporal resolution, with further analysis of uncertainty. Oversampling techniques could be applied near populated areas in order to improve the understanding of isoprene, HCHO, and ozone relationships over cities (e.g., Surl, Palmer, and Abad 2018). Furthermore, publishing the updated underlying emission factors for Australia and implementing the changes in GOES-Chem would improve the understanding of the natural atmosphere over this relatively remote portion of the planet ~~for modelling scientists world wide~~.