

## 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 ( $\text{O}_3$ ) and secondary organic aerosol (SOA) production, and methane longevity. Australian forests are strong emitters of isoprene, the primary BVOC emitted globally (Guenther et al. 2006; Messina et al. 2016). However; emissions are poorly understood due to poor measurement coverage. The lack of measurements makes it difficult to estimate the subsequent atmospheric processes. The main BVOC emitted globally is isoprene ( $\text{C}_5\text{H}_8$ ), which 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 of different plant species (phenotypes) in varying conditions, although these are seldom well understood within Australian forests. Guenther et al. (2012) Estimated global biogenic isoprene emissions at roughly  $535 \text{ Tg yr}^{-1}$ , while Sindelarova et al. (2014) estimated around  $411 \text{ Tg yr}^{-1}$ . Reactions following emissions are complex, and are sensitive to other trace gases in the ambient atmosphere. Uncertainties in several important products such as ozone and SOA are increased due to both isoprene measurement difficulties and its complicated subsequent chemical mechanisms. Isoprene emissions may be overestimated in Australia since they are based on measurements taken from a few heavily emitting young eucalyptus trees which are not representative (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 we describe and use a technique using satellite measurements of HCHO to calculate surface isoprene emissions. HCHO is a dominant product of most BVOCs, including isoprene, and is measured by satellites via remote sensing. In situ isoprene concentration measurements are costly and sparse within Australia, while satellite HCHO data are plentiful and freely available, making this technique very attractive. Such techniques have informed isoprene emission inventories in North America (Abbot 2003; Palmer 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 the OMHCHO dataset from the OMI instrument (see Section 2.3) is used as the basis for HCHO amounts.

### 3.1.1 Aims

Recent work suggests that modelled emissions may be overestimated in Australia (Emmerson et al. 2016). This work tries to improve the understanding of isoprene emissions over the whole of Australia, clarifying the spatial distribution of bias and how these biases impact modelled chemistry. We estimate isoprene emissions in Australia using a top-down technique based on OMI HCHO measurements and GEOS-Chem modelled yields. This a posteriori top-down estimate is then checked against sparse available ground-based measurements. The GEOS-Chem model is modified to run with updated isoprene emissions to determine potential impact on modelled chemistry. Wellness of fit between in situ, satellite, and modelled HCHO is determined with and without updated emissions estimates.

In this chapter we outline why current isoprene emissions estimates are inadequate and how they can be improved. We discuss literature which shows how the estimates may be too high, and describe how emissions may be calculated using satellite datasets. Section 3.2 lays out how new isoprene emissions are estimated, with results examined in Section 3.3. Updated satellite HCHO columns (Chapter 2) are compared to available measurements in Section 3.3.4. Uncertainties for each step along the way are quantified in Section 3.4. In Section 3.5 we examine how these changes in emissions would affect ozone concentrations in Australia, along with some other chemical processes.

### 3.1.2 Existing emissions estimates

The Model of Emissions of Gases and Aerosols from Nature (MEGAN) is one of the most popular emissions inventories for biogenic isoprene. Along with other models which rely on measured plant emission rates, it is poorly calibrated for Australian conditions. Emissions of isoprene ( $C_5H_8$ ) appear to be overestimated in some regions within Australia (Sindelarova et al. 2014; Stavrakou et al. 2014; Emmerson et al. 2016). A lack of measurements of isoprene emission rates in Australia, along with missing soil moisture parameterisation are the most likely culprits TODO: cite. Sindelarova et al. (2014) showed how the isoprene emissions could be as much as halved by accounting for lower soil moisture. Stavrakou et al. (2015) saw isoprene emissions overestimated by a factor of 2-3 in January. Emmerson et al. (2016) discuss the suitability of MEGAN's isoprene and monoterpene emission factors over southeast Australia, and suggest isoprene emissions are estimated 2-6 times too high. They also show that no blanket increase or decrease in emission factors is appropriate for the entire southeast of Australia. Additionally, emissions of monoterpenes ( $C_{10}H_{16}$ , two units of isoprene) appear to be underestimated (Emmerson et al. 2016).

Recently Bauwens et al. (2016) estimated isoprene emissions with a top-down technique using the IMAGESv2 global CTM. They calculate emissions which create the

closest match between model and satellite vertical columns, and compare these a posteriori data with their a priori (satellite data) and independent data sets. For Australia they found emissions ranging from 26–94 Tg C yr<sup>-1</sup>. In this thesis we prioritise the analysis of a top-down emissions estimate compared against MEGAN, along with flow on effect of changed emissions on 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 km) altitudes. NMVOCs are generally short lived (< 1 hr), and the most prominent of these is isoprene. Isoprene is emitted and enters the atmosphere in the gas phase, where it begins a complex series of reactions. HCHO is produced with high yield in many reactions beginning with isoprene oxidation, discussed in more detail in Section 1.3.3, and has a lifetime of a few hours (Kefauver, Filella, and Peñuelas 2014).

Top-down estimates look at how much of a chemical is in the atmosphere and try to work out how much of its major precursors were emitted. This generally takes advantage of longer lived products which may reach a measurable equilibrium in the atmosphere. For isoprene this is done by looking at atmospheric HCHO enhancement, which can be largely attributed to isoprene emissions once transport and other factors are accounted for. 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, allowing isoprene emissions estimation by proxy (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 in a review by Streets et al. (2013). In this thesis we use NASA's OMHCHO product, using measurements from OMI on board the AURA satellite (see Section 2.3) as 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, which may be a problem if the chemistry is uncertain.

#### 3.1.3.1 Bayesian

Bayesian inversion optimises model parameters in order to minimise the difference between model output and some separate dataset. Emissions of isoprene (and other precursors to HCHO) will form part of the set of model parameters which are tuned to make the model HCHO output most closely match satellite (for instance) 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 and other variable parameters of interest:

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

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 we know about  $y$ .

This method is used in Shim et al. (2005) to optimise isoprene emissions in areas with high HCHO concentrations. They showed model underestimation of isoprene emissions by 14-46%, which when corrected reduced bias between GOME HCHO measurements and GEOS-Chem model output by 3-25%. An example showing how regional anthropogenic VOC emissions estimates can be improved using OMI HCHO observations and the CHIMERE CTM can be found in Curci et al. (2010).

An advantage of the Bayesian method is that it can account for pyrogenic and anthropogenic emissions rather than filtering them out. However; biases may still arise due to errors in modelled emission estimation (Curci et al. 2010). Also the Bayesian method is computationally expensive due to the requirement that model runs take place using many permutations of changed inputs. In this work we do not use the Bayesian method due to the heavy computational costs involved.

### 3.1.3.2 Linear

This technique is the less complicated of the two, and is performed in this thesis. Using vertical columns of biogenic HCHO and its measured or modelled yield from isoprene, one can infer the local (grid space) isoprene emissions (Palmer 2003; Milliet et al. 2006; Marais et al. 2012; Bauwens et al. 2016). In Australia, effective molar HCHO yield from isoprene has not been extensively studied (in other continents this value varies from 1-3, depending on local  $\text{NO}_x$  concentrations). The linear inversion requires several assumptions, which limits the techniques utility in some regions or seasons. The primary assumption is that HCHO and its precursors (primarily isoprene) in a linear steady state relationship. This allows one to link isoprene emissions to HCHO measurements using production and loss formulae.

The calculation requires reaction rates and yields from isoprene to HCHO, which can be determined most readily using chemical modelling. The methodology for calculating isoprene emissions from HCHO is laid out in Palmer (2003), and takes into account the expected lifetime and reaction rates of the precursor VOCs and HCHO. In their work isoprene emissions fluxes were derived using the Global Ozone Monitoring Experiment (GOME) satellite instrument. Palmer's method improved biogenic isoprene emissions estimates (compared with in situ measurements) over two available inventories: the U.S. EPA Biogenic Emissions Inventory System and the Global Emissions Inventory Activity. An outline of how this method is applied in this thesis is given in Section 3.2.1.

This technique assumes fast HCHO yield and no precursor transport, which is unrealistic in several scenarios; e.g. high winds can lead to transported precursors, or low  $\text{NO}_x$  concentrations can slow down HCHO yield. Filtering out data which does not match assumptions is required but can limit the utility of this technique, and leads to some dependence on environmental factors. As we are estimating biogenic emissions of isoprene, we must filter out areas where HCHO enhancements arise due

to anthropogenic or pyrogenic sources. Uncertainties in the technique are discussed in more detail in Section 3.4.1. On the plus side, the simple nature of the inversion requires very little computational power after acquiring satellite and model datasets, even over large amounts of gridded data.

## 3.2 Methods

We broadly follow the method of Palmer et al. (2001) to create a biogenic isoprene emissions estimate over Australia. In this Thesis the slope between modelled isoprene emissions (a priori or  $E_{GC}$ ) and HCHO total columns ( $\Omega_{GC}$ ) over Australia is calculated monthly using GEOS-Chem. 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 taken from the reduced major axis (RMA) regression between the a priori and  $\Omega_{GC}$  in each model grid box each month. Figure 3.1 shows an example; the modelled regression between emissions and tropospheric columns for January, 2005. Also shown is the time series for these two quantities averaged over Australia, and the squared correlation coefficient along with a sample of scatter plots and regressions from four grid squares.

### 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 3.2.

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) at  $0.25^\circ \times 0.3125^\circ$  horizontal resolution.
2. Level three satellite data are used to make anthropogenic, fire, and smoke influence masks (see Section 2.7). These are applied to remove  $\Omega_{OMI}$  which may be influenced by pyrogenic or anthropogenic sources.
3. Modelled slope ( $S$ ) calculations depend on several assumptions which are not always valid. A mask is created for 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. This smearing value is determined as  $\hat{S} = \Delta\Omega_{GC}/\Delta E_{OMI}$ : the ratio of the differences between model runs of HCHO columns and isoprene emissions. The acceptable range for  $\hat{S}$  over Australia is determined as 800 - 4600 s. A full description of the creation of this smearing filter is given in Section 3.2.7
4. GEOS-Chem modelled biogenic emissions of isoprene ( $E_{GC}$ ) along with total 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 ( $\Omega_{GC} = SE_{GC} + \Omega_{GC,0}$ ). Calculation of this modelled slope is explained in Section 3.2.4.

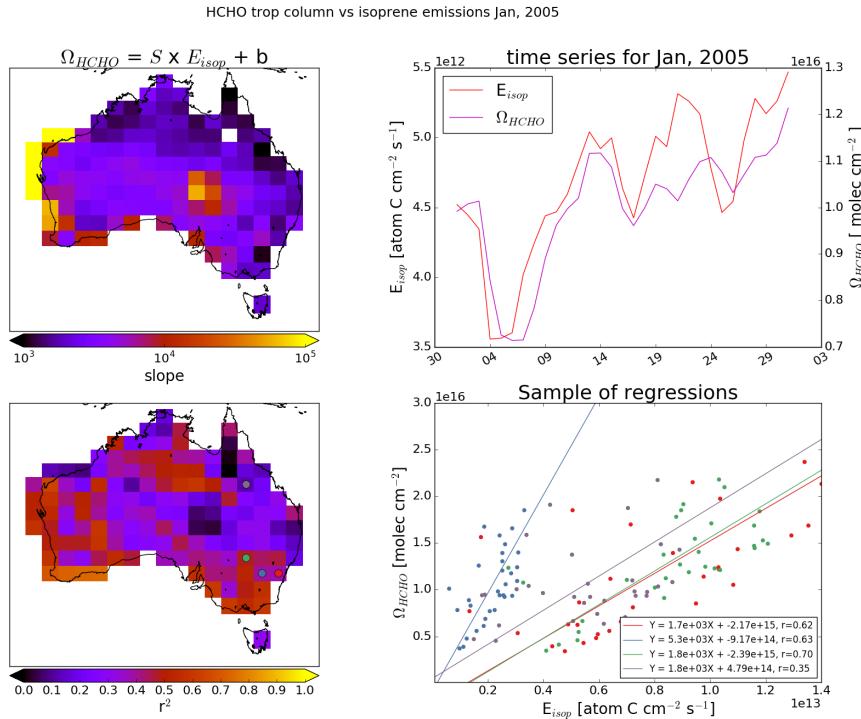


FIGURE 3.1: Top left: RMA slope between modelled tropospheric column HCHO ( $\Omega_{GC}$ ) and isoprene emissions ( $E_{GC}$ ) using midday (13:00-14:00 LT) values over for January 2005, per grid square at  $2^\circ \times 2.5^\circ$  horizontal resolution. Top right: Australia-wide average of midday emissions and tropospheric columns. Bottom left: Squared RMA correlation coefficient for regression in top left. Coloured dots correspond to colour of regressions shown in bottom right panel. Bottom right: Sample of correlations from four grid squares.

5. Satellite measured HCHO columns (Section 2.3) are used to create a top-down estimate of biogenic isoprene emissions ( $E_{OMI}$  atoms C  $\text{cm}^{-2} \text{s}^{-1}$ ). This product is our a posteriori, and calculation details are given in Section 3.2.6.
6. A posteriori top-down emissions  $E_{OMI}$  are compared against a priori emissions, and analysed in conjunction with concentration measurement campaigns (MUMBA, and SPS). Results are examined in Section 3.3.
7. GEOS-Chem is run using the a posteriori emissions (see Section 3.2.8), and HCHO,  $\text{O}_3$ , isoprene, and  $\text{NO}_x$  outputs are compared to campaign and satellite measurements where possible.

### 3.2.2 Masks and reprocessed satellite HCHO

Satellite data pixels are read from the OMHCHO level 2 satellite 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 we use the OMI product as it has better temporal coverage and increased pixel counts over Australia when compared to GOME (on board the ERS-2 satellite).

In order to determine biogenic HCHO enhancements from  $\Omega_{OMI}$ , we require filters for non-biogenic sources. While one source of HCHO production is methane oxidation, the linear regression used to estimate isoprene emissions effectively ignores this source as part of the background, which means a methane filter is not required. Anthropogenic, pyrogenic, and smoke influence masks are created from three satellite products:  $\text{NO}_2$  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 which occur in local or adjacent grid squares at  $0.25^\circ \times 0.3125^\circ$  horizontal resolution.
2. Influence from transported smoke plumes is removed where OMI aerosol absorption optical depth (AAOD, from OMAERUVd) is greater than 0.03.
3. A filter for anthropogenic influence is created daily using OMNO2d  $\text{NO}_2$  tropospheric column amounts; masking any grid squares with greater than  $2 \times 10^{15}$  molec  $\text{cm}^{-2}$  on any particular day, along with grid squares where the yearly average is above  $1.5 \times 10^{15}$  molec  $\text{cm}^{-2}$ .

These masks are binned using daily averaged or summed values within our  $0.25^\circ \times 0.3125^\circ$  horizontal resolution, see Section 2.7 for further details. The recalculated corrected vertical columns are saved to OMHCHORP dataset both before and after applying the filters, to allow filter analysis.

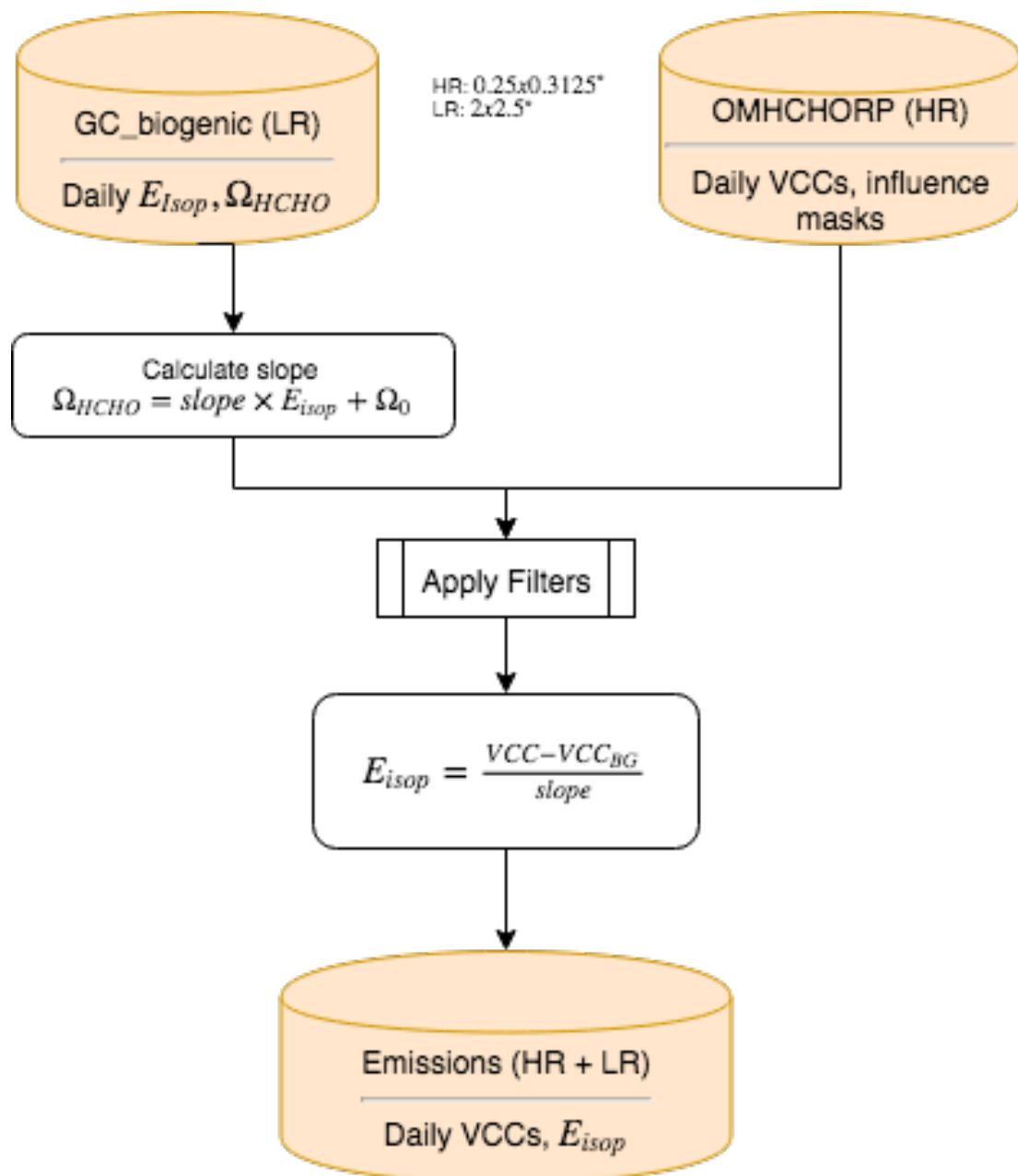


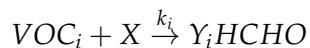
FIGURE 3.2: Creation of isoprene emissions dataset using OMH-CHORP and biogenic GEOS-Chem outputs.

### 3.2.3 GEOS-Chem emissions

Global atmospheric studies often use models and inventories of various chemical emissions, along with a chemical transport model (CTM) to examine transport, emission, deposition, and other chemical processes in the atmosphere. Emissions of Bio-generative Volatile Organic Compounds (BVOCs) including isoprene are often the subject of studies as they are still relatively uncertain, as well as being drivers for important oxidation and pollution events. In this work MEGAN is run as a module within GEOS-Chem, a global CTM which uses emissions inventories and meteorological data to simulate atmospheric gas concentrations and transport. 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 is 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 ( $\sim 1$  yr) VOC. Over continents the variability in HCHO is driven by shorter lived precursor emissions. The intermediate steps are considered negligible as HCHO is produced quickly from short-lived intermediates:



where  $X$  is an oxidant,  $Y_i$  is HCHO yield (per C atom in  $VOC_i$ ), and  $k_i$  is the reaction rate. In specific conditions, HCHO total columns ( $\Omega$ ; molec cm $^{-2}$ ) are linearly related to isoprene emissions.

The isoprene to HCHO relationship is derived using several assumptions that are important to understand. The first assumption is that HCHO is at steady state, which implies productions ( $P_{HCHO}$ ) and losses ( $L_{HCHO}$ ) 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 ( $k_{HCHO}$ ) 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; depending on grid box size and wind speed. This assumption is reasonable for large enough grid boxes as transport becomes negligible in relation 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. Transport can still be an issue however, and is handled in Section 3.2.7.

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

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

By combining Equations 3.2, 3.3, and 3.4, we have the link between  $\Omega$  and precursor emissions:

$$\begin{aligned} k_{\text{HCHO}}\Omega &= \sum_i Y_i E_i \\ \Omega &= \frac{1}{k_{\text{HCHO}}} \sum_i Y_i E_i \end{aligned}$$

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

$$\sum_i Y_i E_i = Y_{\text{isop}} E_{\text{isop}} + \sum_{i \neq \text{isop}} Y_i VOC_i \quad (3.5)$$

This leads to the linear relationship between isoprene emissions and  $\Omega$ , by equating  $P_{\text{HCHO}}$  and  $L_{\text{HCHO}}$  from Equations 3.4 and 3.3, plugging in Equation 3.5, assuming that the lumped terms make up the background:

$$\begin{aligned} k_{\text{HCHO}}\Omega &= Y_{\text{isop}} \times E_{\text{isop}} + \sum_{i \neq \text{isop}} Y_i VOC_i \\ \Omega &= \frac{Y_{\text{isop}}}{k_{\text{HCHO}}} \times E_{\text{isop}} + \Omega_0 \\ &= S \times E_{\text{isop}} + \Omega_0 \end{aligned} \quad (3.6)$$

Here  $S$  is the slope:  $S \equiv \frac{Y_{\text{isop}}}{k_{\text{HCHO}}}$ . This assumption can be false when pyrogenic or anthropogenic emissions occur, however these scenarios are filtered as much as possible (see Section 2.7).

### 3.2.5 Calculation of modelled slope

To determine  $S$ ; the link between biogenic isoprene and midday column HCHO, we use GEOS-Chem. The term  $E_{\text{GC}}$  is used when discussing isoprene emissions estimated within GEOS-Chem, and  $\Omega_{\text{GC}}$  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  $E_{\text{GC}}$  atoms  $\text{C cm}^{-2} \text{s}^{-1}$  at 13:00 LT daily are extracted, along with  $\Omega_{\text{GC}}$  molec  $\text{cm}^{-2}$  output.
2. Grid square days which are likely to be affected by smearing are removed (see Section 3.2.7).
3. A reduced major axis regression slope is determined between  $\Omega_{\text{GC}}$  and  $E_{\text{GC}}$  using a month of modelled output (one value per day) for each grid square (e.g. see Figure 3.3)

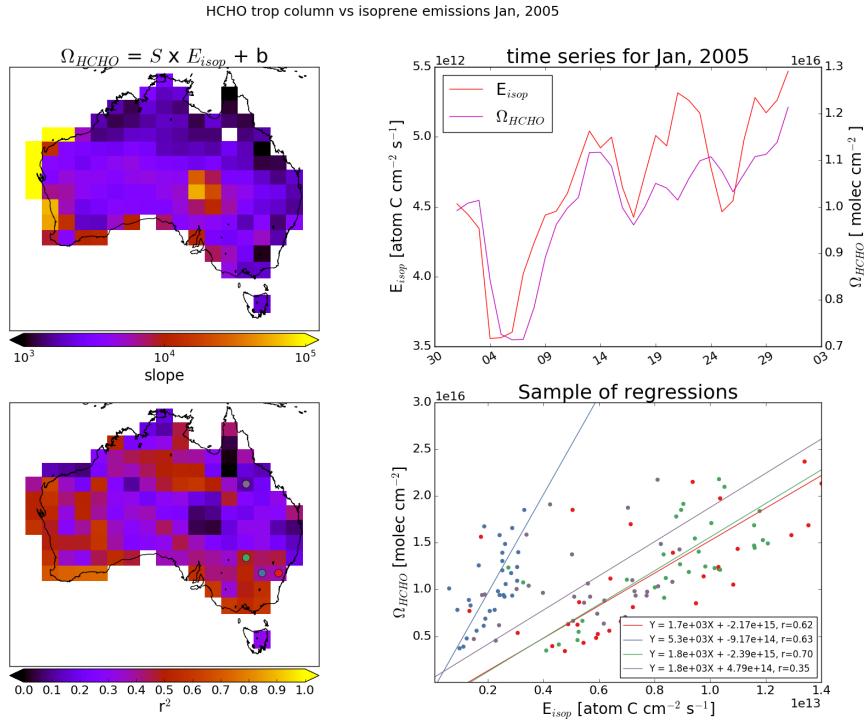


FIGURE 3.3: Top left: the RMA regression slope between modelled HCHO columns and isoprene emissions calculated within each grid box over January 2005, averaged between 1300–1400 LT each day Bottom left: the square of the regression coefficients. Top right: Australia averaged  $E_{GC}$  and  $\Omega_{GC}$  time series. Bottom right: a sample of the regressions coloured by latitude and longitude pairs, matching dots in the bottom left panel.

We look at each  $2^\circ \times 2.5^\circ$  grid box from daily GEOS-Chem (biogenic only) output of  $\Omega_{HCHO} \equiv \Omega_{GC}$  and  $E_{isop}$  within Australia, and calculate the (RMA) slope for each month. Modelled background concentrations can be ignored here as they do not affect slope calculation. This effectively gives us monthly gridded slope ( $S$ ) between biogenic isoprene emissions and HCHO columns, in units of s. Figure 3.3 (top left) shows how  $S$  varies spatially over Australia for an example mid-summer month. Some areas can be seen to be very sensitive to emissions, such as the west coast and Ayer basin, which is likely due to the low precursor and HCHO levels in those areas. The regression coefficients also vary spatially (bottom left), and some areas show little correlation, likely due to weather and transport. The slopes shown in the bottom right panel show a small sample of scatter and regression plots. These can range widely due to differences in emission and yield parameters, which plays a role in the smearing filters applied in Section 3.2.7. Due to the  $2^\circ \times 2.5^\circ$  horizontal resolution of GEOS-Chem, calculations from grid boxes on the coast which are largely oceanic need to be discarded as the change in HCHO is not dominated by emissions of isoprene, as is assumed for equation 3.6.

One issue with slope calculation is potential transport (or smearing), either of isoprene before any HCHO is formed, or of HCHO formed by local emissions. The effects of this are dealt with by forming a smearing filter (see Section 3.2.7 Accounting for smearing). Days where we expect smearing may be affecting local levels of HCHO are removed before they are used to calculate  $S$ , and a quick analysis is performed on how the filter affects monthly slope, correlation, and uncertainty. Figure 3.4 shows the calculated monthly slope for 2005-2012, along with its 95% confidence interval over Sydney. The left column does not apply the smearing filter prior to slope calculation, while the right column does. The bottom two rows show how the slope calculation looks when using multiple years of data for each month. This plot has been repeated for several grid squares over Australia (not shown). When calculating top-down emissions the smear filtered slope ( $S$ ) is used for each grid square month. The multiple year monthly averaged slope is used instead when the regression coefficient ( $r$ ) is less than 0.4, or then number of data points used in the regression ( $n$ ) is less than 10. When  $r$  for the multiple year slope is also lower than 0.4, no estimation is performed, this is also true for negative values of  $S$ .

There are two simple ways to determine the modelled background HCHO, one of which involves running the model with no isoprene emissions, which shows how much they alter vertical column HCHO. This is effective since we have assumed variation in HCHO columns only depends on isoprene emissions, so our background term is theoretically identical to the emission free simulated HCHO. The other way uses HCHO over the remote Pacific at matching latitudes and times, which emulates how the background is determined for the satellite measured HCHO. The difference between these two methods is negligible, so for consistency with satellite data, we determine backgrounds using the remote Pacific. Figure 3.5 shows GEOS-Chem total column HCHO with and without isoprene emissions along with amounts over the remote Pacific at the same latitudes. Background HCHO for any latitude in this thesis are calculated by averaging longitudinally (140°W to 160°W) the matching latitudes over the remote Pacific.

### 3.2.6 Calculation of Emissions

Top-down emissions estimates are calculated using OMHCHO (see Section 2.3) slant columns and an updated AMF calculated using code from Paul Palmer's group (see Section 2.6.5). These emissions are referred to as the a posteriori from here on, or  $E_{OMI}$  in formulae.

As performed elsewhere Palmer (e.g. 2003), Millet et al. (2006), and Bauwens et al. (2016), we assume that HCHO and isoprene columns are in a steady state, with no horizontal transport. We also assume that isoprene is the only compound enhancing the HCHO levels, which requires that we filter out influence from fires, smoke, and anthropogenic emissions. 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:

$$\begin{aligned}\Omega_{OMI} &= S \times E_{OMI} + \Omega_0 \\ E_{OMI} &= \frac{\Omega_{OMI} - \Omega_0}{S}\end{aligned}\quad (3.7)$$

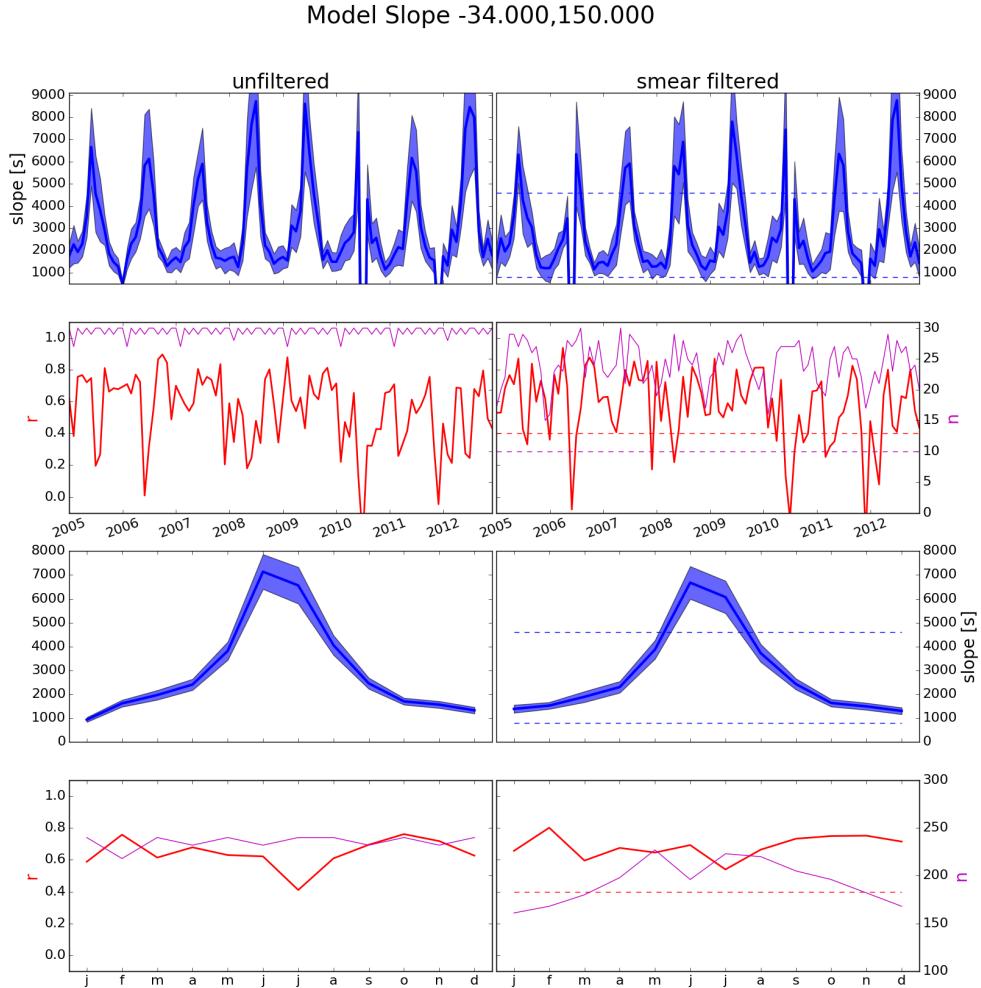


FIGURE 3.4: Row 1: monthly slope along with 95% confidence interval both before (left) and after (right) applying the smear filter for the model grid square containing Sydney over 2005-2012. Limits used in creation of the smear filter are shown with dashed lines. Row 2: regression coefficient and data-point 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.

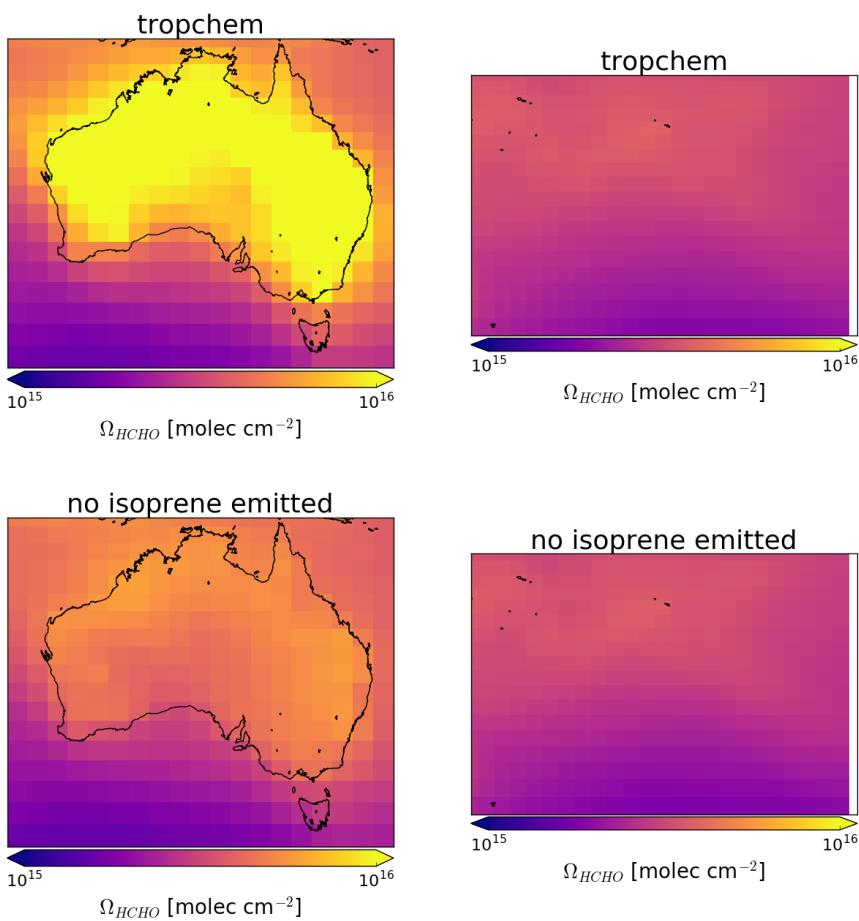


FIGURE 3.5: Total column HCHO over Australia (left) and the remote Pacific region (right) using GEOS-Chem with (top) and without (bottom) isoprene emissions.

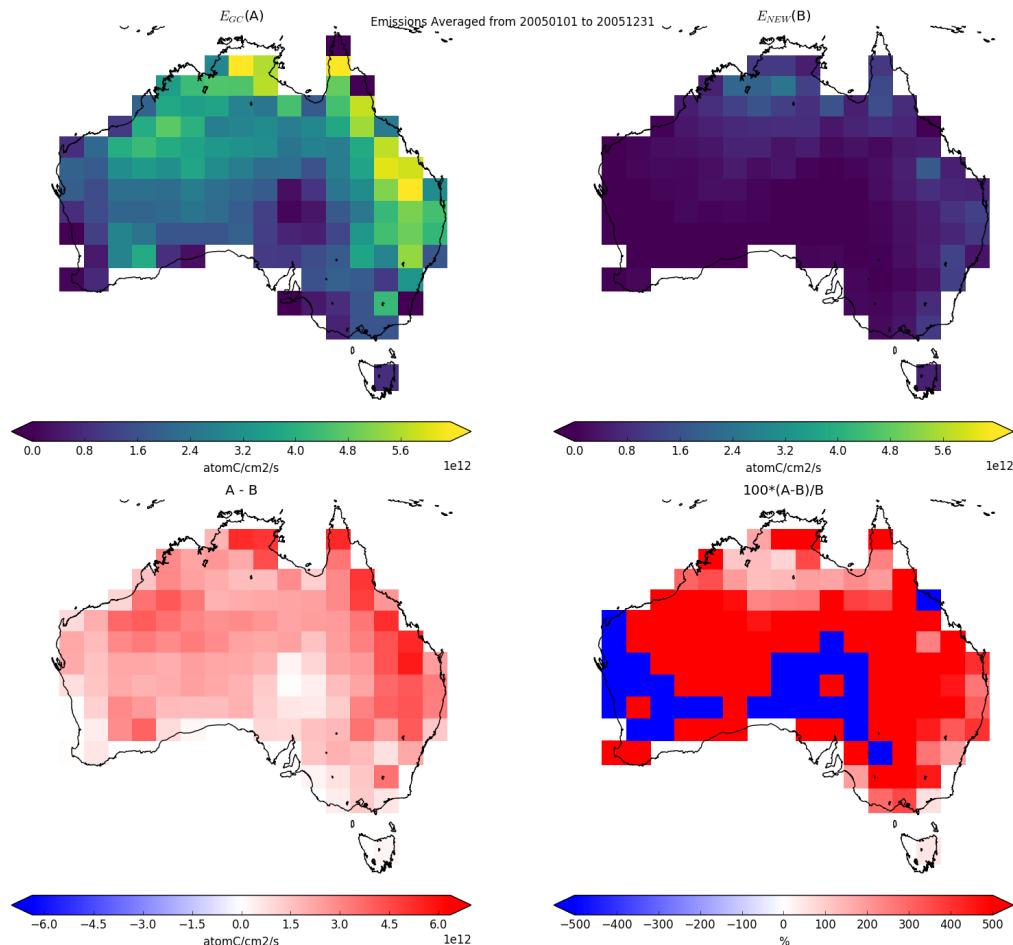


FIGURE 3.6: Top row: isoprene emissions from GEOS-Chem ( $E_{GC}$ , left) simulation and top-down ( $E_{OMI}$ , right) calculations averaged over the month of January, 2005. Bottom row shows the absolute (left) and relative (right) differences between the two.

This is the same as equation 3.6, except now we use the satellite HCHO ( $\Omega_{OMI}$ , and  $\Omega_0$ ).  $\Omega_0$  is calculated using measurements in the remote Pacific averaged monthly and longitudinally, for each  $\Omega_{OMI}$ . This leaves  $E_{OMI}$  as the only unknown once the satellite measurements are processed to match the temporal and horizontal resolution of  $S$ . Figure 3.6 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, analysis of the full record is discussed in Section 3.3. The a priori exceeds 500% of the a posteriori in many regions, however this is mostly in regions of low emissions and represents only minor absolute differences.

One potential issue in this top down estimation technique over Australia is the low number of valid satellite measurements which may occur due to the higher zenith angles in winter and at higher latitudes. When calculating the a posteriori from our modelled slope, negative emissions result wherever measured columns are lower than background amounts (as  $E_{OMI} = \frac{\Omega_{OMI} - \Omega_0}{S}$ ). These are set to zero, which increases

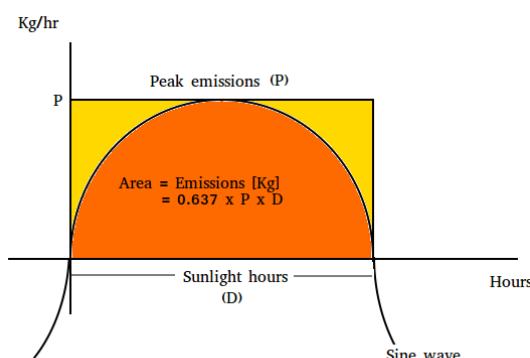


FIGURE 3.7: Total daily isoprene emissions (in kg) is represented by the area under the sine wave.

monthly averages by TODO:  $\sim X - Y \%$  ( $\sim XX - YY \text{ atom C cm}^{-2} \text{ s}^{-1}$ ) over Australia, with the highest increases occurring during (TODO winter?).

Top-down emission rates calculated in this work are in units of  $\text{atom C cm}^{-2} \text{ s}^{-1}$ . In order to calculate the emissions in kg, each grid square is multiplied by its area, and then daily emissions are assumed to follow a sine wave peaking at the midday value. Figure 3.7 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). This approximation is required since OMI observations occur at midday, when isoprene emissions are at their diurnal peak.

### 3.2.7 Accounting for smearing

In high  $\text{NO}_x$  ( $>\sim 1 \text{ ppb}$ ) environments, isoprene has a lifetime on the order of 30 minutes, and HCHO can be used to map isoprene emissions with spatial resolution from 10-100 kms (Palmer 2003). In low  $\text{NO}_x$  conditions, isoprene has a longer lifetime (hours) and may form HCHO further from the source area. Horizontal transport smears the HCHO signal so that source location would need to be calculated using wind speeds and loss rates (Palmer et al. 2001; Palmer 2003). Smearing is a measure of how much HCHO in a given grid box was produced from isoprene emitted in a different (upwind) grid box. Smearing affects emissions estimates as HCHO enhancements may be wrongly attributed as coming from areas downwind of where they occurred. Once smearing sensitive grid squares are filtered out, application of equation 3.6 is used to estimate isoprene emissions across Australia.

Over Australia  $\text{NO}_x$  levels are generally not high enough to ensure quick HCHO formation and we must take care to account for resultant smearing. Smearing limits the horizontal resolution of the top-down inversion process used in this Thesis, as a finer resolution leads to increased issues with transport. The relatively coarse horizontal resolution ( $2^\circ \times 2.5^\circ$ ) used by GEOS-Chem is advantageous in this aspect. Smearing has been analysed in various publications (e.g. Martin et al. 2003; Palmer 2003; Millet et al. 2006; Stavrakou et al. 2009; Marais et al. 2012; Barkley et al. 2013; Zhu et al. 2014; Wolfe et al. 2016; Surl, Palmer, and Abad 2018), and is often calculated using the method designed in Palmer (2003). This involves using two almost identical

model runs, one of which has isoprene emissions scaled globally by a factor (generally from 0.5 to 2). Another method (e.g. Stavrakou et al. 2009) involves the analysis of an adjoint CTM, however this is computationally expensive and is not pursued here.

### 3.2.7.1 Calculation of smearing

In order to understand the smearing calculation the underlying equations and assumptions must first be understood. From Section 3.2.4, Equation 3.6 we have the formulation of a modelled slope ( $S$ ) being the yield of HCHO per C of emitted isoprene divided by HCHO loss rate per second ( $S = \frac{Y_{isop}}{k_{HCHO}}$ ). Using two runs of GEOS-Chem with isoprene emissions being the only difference we have:

$$\begin{aligned} Run_1 : \Omega_{HCHO} &= SE_{isop} + \Omega_0 \\ Run_2 : \Omega'_{HCHO} &= S'E'_{isop} + \Omega'_0 \end{aligned} \quad (3.8)$$

There are several assumptions which need to be understood, as these are what is tested by the smearing calculation. The initial assumption is that the system is in a steady state, with no transport of isoprene affecting HCHO columns, this is the basis for equations 3.8. It is assumed that background values ( $\Omega_0$ ) are from oxidation of methane and other long lived VOCs, so that  $\Omega_0 = \Omega'_0$ . Between these two runs we are only changing the  $E$  term, we do not change any chemistry and so we can expect that the yield and loss rate is not changing between the two runs  $S = S' = \frac{Y_{isop}}{k_{HCHO}}$ . So that Equations 3.8 may be combined as follows:

$$\begin{aligned} Run_1 - Run_2 : \Omega_{HCHO} - \Omega'_{HCHO} &= SE_{isop} - S'E'_{isop} + \Omega_0 - \Omega'_0 \\ \Omega_{HCHO} - \Omega'_{HCHO} &= S(E_{isop} - E'_{isop}) \\ \Delta\Omega_{HCHO} &= S\Delta E_{isop} \\ \hat{S} &\equiv \frac{\Delta\Omega_{HCHO}}{\Delta E_{isop}} \approx \frac{Y_{isop}}{k_{HCHO}} \end{aligned} \quad (3.9)$$

This allows the combination of outputs from the two runs to be used as a check against  $\hat{S}$  being excessively different from expected values for  $S$ .

In order to filter potential smearing, a daily modelled value for  $\hat{S} \approx Y_{isop}/k_{HCHO}$  is determined. By assuming midday HCHO lifetime typically falls within 1.5 to 4 hrs (as seen in the USA), and isoprene to HCHO yield (HCHO per isoprene carbon emitted) lies within the range 0.2 to 0.4 (scenarios estimated in Palmer (2003)): one can set a simple bound on  $\hat{S}$  of  $[0.2 \times 1.5, 0.4 \times 4]$  hrs or 1080 to 5760 seconds. As  $\text{NO}_x$  levels across Australia are relatively low, the yield is likely lower than seen in Palmer (2003): and here we reduce the bounds by 20% and round to the nearest hundred to get a bounding range of 800 to 4600 for  $\hat{S}$ . This range strikes a balance between unlikely modelled yields and how much data is lost to filtering. A better approximation of lifetimes for HCHO is required to properly account for seasonality and regional  $\text{NO}_x$  concentrations. Table 3.1 shows the smearing filters or typical slopes seen in other works.

Similarly to smearing sensitivity calculations in Marais et al. (2012), we run GEOS-Chem with isoprene emissions halved, then calculate  $\hat{S} = \frac{\Delta\Omega_{HCHO}}{\Delta E_{isop}}$  from Equation 3.9.

TABLE 3.1: Smearing filters or slopes ( $S$ ) seen in literature.

Publication	min. (s)	max. (s)	Notes
Palmer (2003)	1270	2090	Slope ranges seen in north America summer
Marais et al. (2012)		4000	Smearing limits for Africa
Barkley et al. (2013) <sup>a</sup>	1300	1800	Smearing limits for South America
Surl, Palmer, and Abad (2018)	2200	4900	Slope ranges seen in India
In this Thesis	800	4600	Smearing limits for Australia

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

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

Here  $\Delta$  represents the departure (daily over 1300-1400 LT) from default run values. If  $\hat{S}$  sits outside an acceptable range then we remove that grid square day from both  $S$  and subsequent a posteriori calculations. A relatively large change in  $\Omega_{HCHO}$  compared to local emissions ( $\hat{S} > 4600$ ) suggests that HCHO is being formed from non-local isoprene emissions. Alternatively, a relatively low value of  $\hat{S}$  ( $\hat{S} < 800$ ) suggests emissions from the local grid square are being exported before they form HCHO, meaning local HCHO levels are not due to local emissions.

Smearing is sensitive to the time for which it is determined, Figure 3.8 shows smearing over two seasons. Isoprene emissions ( $E_{isop}$ ) are defined in two ways in this figure: left column from daily averaged flux, and right column from and midday (1300-1400 LT) flux. Essentially the midday isoprene emissions are at the peak of their daily cycle (shown later in Figure 3.18) which means the effect of smearing is relatively smaller during these hours.

### 3.2.7.2 Sensitivity to smearing

Figure TODO: shows averaged isoprene emissions with added markers showing when the threshold of 800-4600 affects at least one day within the season (cyan or pink diamonds) and where it removes all data for that grid box (blue or red x). Smearing can be dependent on local or regional weather patterns, as greater wind speeds will reduce the time any emitted compound stays within the local grid box. As such smearing can vary greatly both spatially and temporally. The filter removes more data on coastal grid squares as they are more affected by winds and transport, particularly when only a portion of the grid square is land. During summer data-loss from smearing is only minor (todo: Y%), however data-loss peaks in winter (TODO X %), especially in higher southern latitudes.

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 3.4). Where smearing is prevalent the relationship between a priori emissions and HCHO columns may already be weak due to low actual emissions or unsuitable geological and meteorological conditions. Loss of data due to filtering is handled by

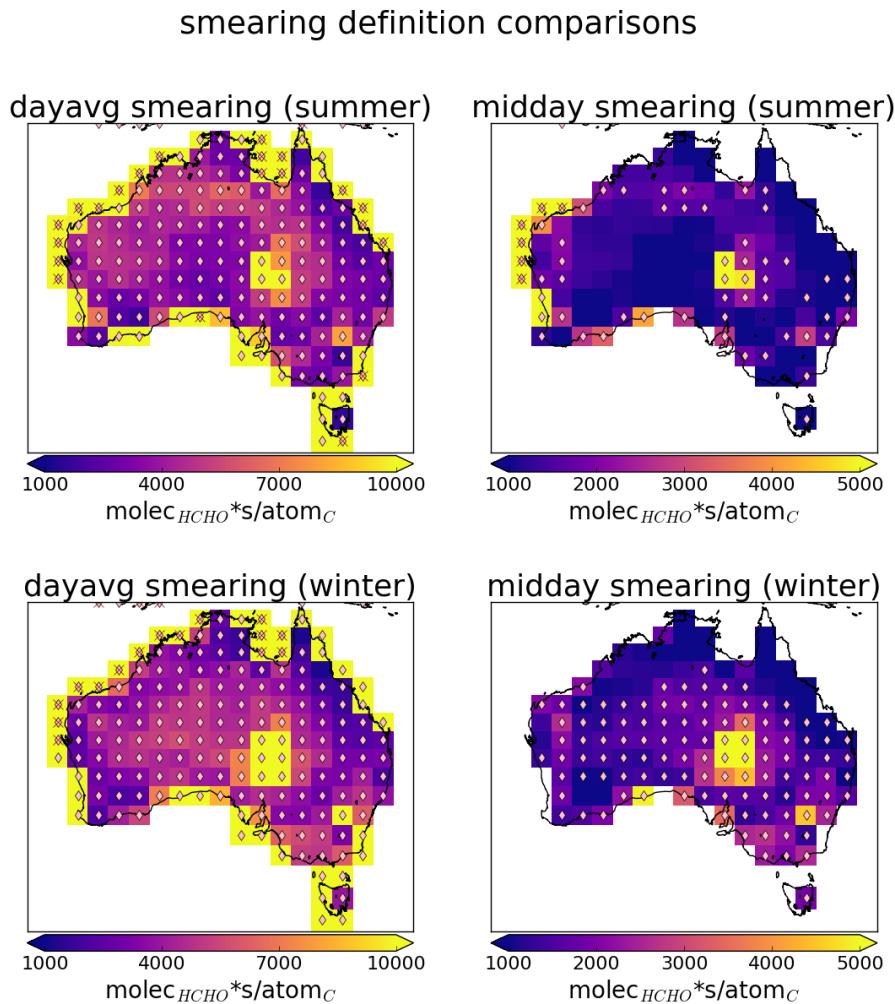


FIGURE 3.8: Smearing ( $\hat{S}$ , see text) in summer (DJF, top row) and winter (JJA, bottom row) using averaged isoprene emissions daily (left column) and 1300-1400 LT (right column). The scale changes between left and right columns.

using multiple years of data for any affected grid square month as follows.  $S$  and associated regression coefficients ( $r$ ) are calculated monthly, if  $r < 0.4$  then the regression is calculated using multiple years of data for that month. This multiple year regression is discarded completely if it also has a coefficient of  $r < 0.4$ , leaving no slope for the grid square for the month.

### 3.2.7.3 Smearing length scale

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

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

where  $U$  is wind-speed. Palmer (2003) further define a smearing length scale:  $L_{s,i}$  as the distance downwind where most of the precursor is completely transformed into HCHO. Here we use the simplification of  $L_{s,isop} \approx \frac{U}{k_{HCHO}}$ , as isoprene loss rates ( $2 \text{ h}^{-1}$ ) exceed those of HCHO ( $0.25\text{--}0.67 \text{ h}^{-1}$ ) by a factor of 3–8. Throughout most of the year, monthly averaged wind speeds do not exceed  $20 \text{ km h}^{-1}$  ([http://www.bom.gov.au/jsp/ncc/climate\\_averages/wind-velocity/](http://www.bom.gov.au/jsp/ncc/climate_averages/wind-velocity/) [accessed Feb., 2019]), although daily wind speeds will be highly variable. This means that a reasonable smearing length scale is less than  $25/0.25 = 100 \text{ km}$  over Australia. Grid boxes used in top-down estimation of isoprene emissions describe rectangles with both side lengths approximately 200 km, so in this average case we do not expect to be overly impacted by smearing.

The half-life of HCHO is around 1 hr depending on environmental conditions (Kaden DA, Mandin C, Nielsen GD 2010). This would make the expected lifetime ( $\tau = \text{half-life} / \ln 2$ ) around 1.4 hours. Over the majority of Australia conditions are relatively clean (low  $\text{NO}_x$ ) which extends the expected lifetime. The estimated loss rate of HCHO in GEOS-Chem is up to three times higher in summer and along the north and eastern regions associated with denser forest regions, when compared against other regions. This is largely due to loss rates being proportional to concentrations.

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

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

The expected lifetime of HCHO is determined by assuming loss is linear (first order) and dividing grid box daily averaged concentrations of GEOS-Chem HCHO ( $[HCHO]$  in molecules  $\text{cm}^{-3}$ ) by their modelled losses ( $L_{HCHO}$  in molecules  $\text{cm}^{-3} \text{ s}^{-1}$ ). For each grid square over Australia this daily averaged surface lifetime in summer (Jan., Feb.) and winter (Jul., Aug.) is shown in Figure 3.9. Additionally lifetimes coloured by land grid squares (dots in top right panel) are shown over time in the bottom panel. Note that this figure shows the daily averaged lifetime, but further analysis and calculations are based on mid-day (13:00–14:00 LT) values when the lifetime will be significantly shorter. This figure highlights the seasonal nature of HCHO lifetime, although midday

numbers are expected to have less seasonality than shown here, since midday lifetime will be less affected by how long the daylight lasts. Another highlighted issue is the potential latitudinal dependence of HCHO lifetimes, since there is less total insolation leading to lower HCHO loss rates at higher latitudes. The overall takeaway is that the accuracy and utility of any top down HCHO precursor estimation technique will be limited by both season and potentially latitude. These limitations are due to both data availability (as satellite HCHO suffers from high zenith angles) and spatial smearing (from HCHO lifetimes which increase with lower insolation and temperature).

### 3.2.7.4 Nitrous oxide dependence

Isoprene production of HCHO depends on several factors, one of the most important of these is  $\text{NO}_x$ , which directly affects the fate of VOCs in the atmosphere. Isoprene first reacts with the hydroxy radical, producing an organic peroxy radical ( $\text{RO}_2$ ). At higher NO mixing ratios (at least a few hundred pptv),  $\text{RO}_2$  react mostly with NO. At low NO (less than 50 pptv),  $\text{RO}_2$  is more likely to either isomerise, or react with  $\text{HO}_2$ , or another  $\text{RO}_2$ . In low  $\text{NO}_x$  environments, reported HCHO yields from isoprene are around 0.2 - 0.3 C per C (or 100-150 molar %), while in high  $\text{NO}_x$  environments this value becomes two to three times higher (Palmer 2003; Wolfe et al. 2016). Some values for HCHO yield from prior literature are shown in Table 3.2.

$\text{NO}_2$  from the OMNO2d product provides daily mid-day measurements which we compare to output from GEOS-Chem  $\text{NO}_2$  (see Section 2.4.6). The effect of  $\text{NO}_2$  on smearing can be seen in Figure 3.10. This plot shows how smearing over Australia compares to satellite  $\text{NO}_2$ , with smearing distributions binned by  $\text{NO}_2$  both with and without filtering for smearing. At lower  $\text{NO}_2$  levels the smearing is often 2-4 orders of magnitude above the upper threshold. This abruptly decreases at around  $5 \times 10^{14}$  molec  $\text{cm}^{-2}$   $\text{NO}_2$ . There is also a higher number of data points below the lower threshold before that same  $\text{NO}_2$  level, showing that transport is a bigger issue at  $\text{NO}_2 < 5 \times 10^{14}$  molec  $\text{cm}^{-2}$ . This is an issue in Australia since  $\text{NO}_2$  levels sit at roughly this threshold.

### 3.2.8 Running GEOS-Chem using a postriori emissions

After creating our a postiori isoprene emissions estimate, we run GEOS-Chem again with biogenic emissions scaled to match the new estimate. First all 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. A monthly factor ( $\alpha$  at  $2^\circ \times 2.5^\circ$ ) that scales MEGAN towards our top-down estimate is then read by the emissions module in GEOS-Chem, and applied whenever isoprene emissions are calculated. This scaling allows the shorter time-scale variability and dependencies to be retained, while moving the monthly mean value towards the multi-year monthly averaged a posteriori. Short time-scale variability is largely due to meteorological parameters such as temperature and light, etc.. TODO: Figure X shows an example of this process for January: with the multi-year average emissions mapped over Australia (rows one and two) being used to create the  $\alpha$  in the third row.

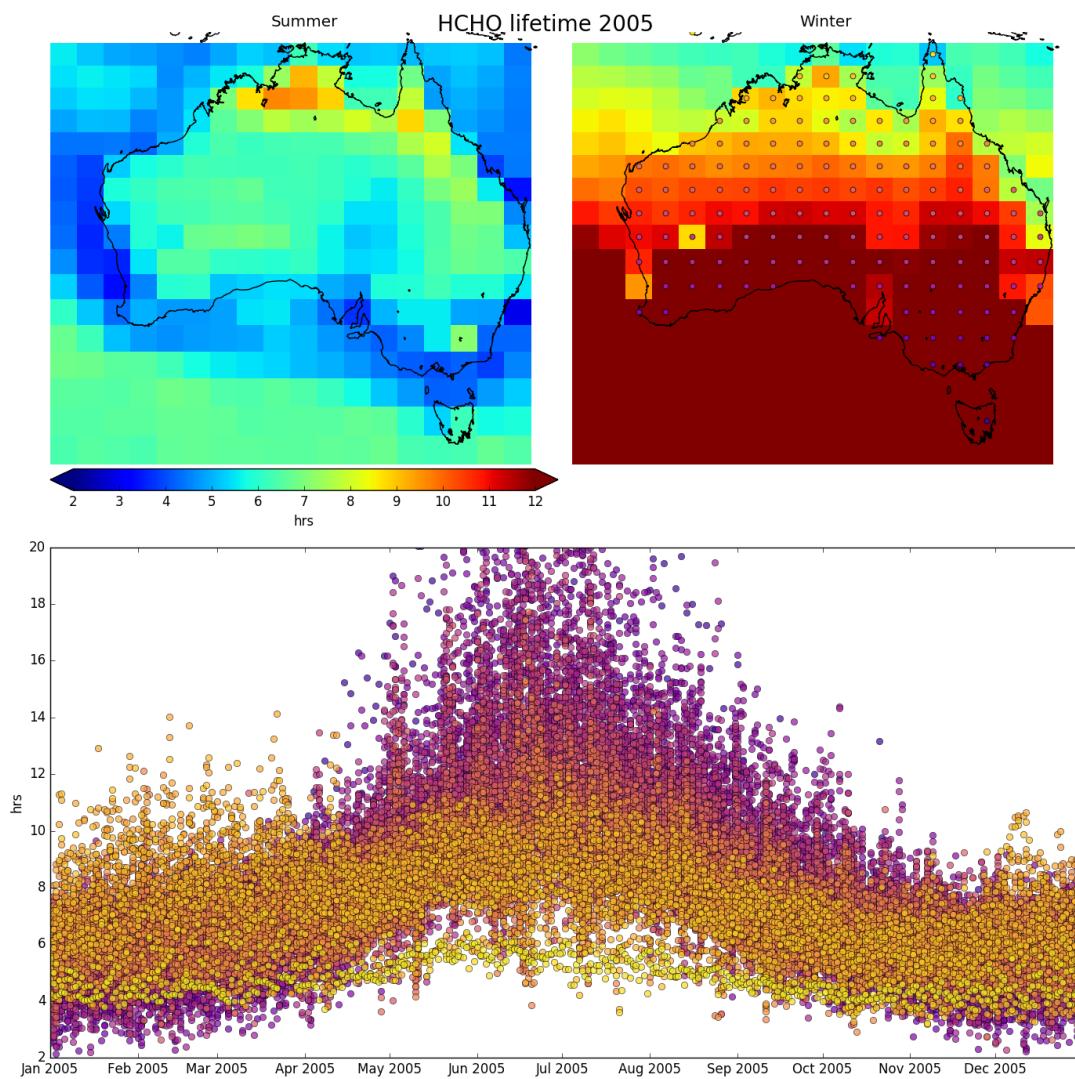


FIGURE 3.9: Top left, right: summer (Jan., Feb.) and winter (JJA) averaged daily surface HCHO lifetime ( $\tau$ ). Bottom panel:  $\tau$  over the year, coloured by grid square (see dots in top right panel).

TABLE 3.2: Isoprene to HCHO yields and lifetime.

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

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

b Palmer (2003): lifetimes assume [OH] is 1e15 mol cm<sup>-3</sup>.

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

d Wolfe et al. (2016): “prompt yield”: change in HCHO per change in isop<sub>0</sub>.  $[isop]_0 = [isop] \exp(k_1[OH]t)$ ; where  $k_1$  is first order loss rate. Effectively relates HCHO abundance with isoprene emission strength.

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

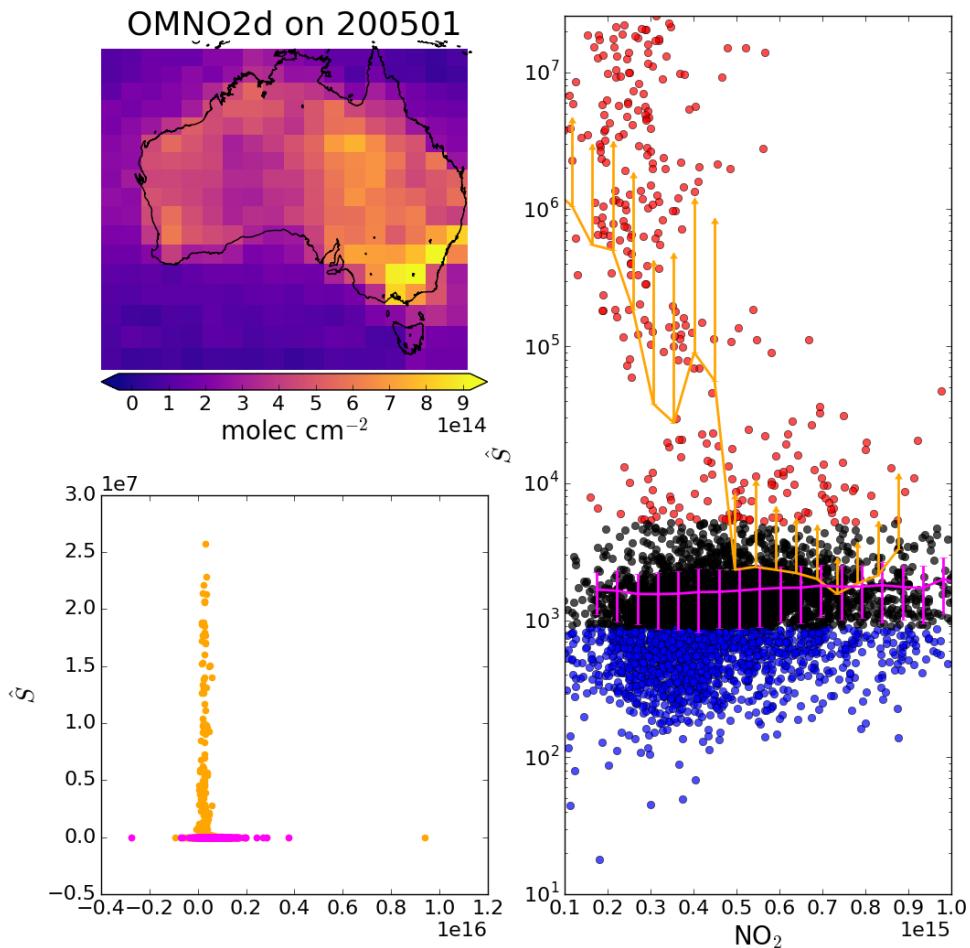


FIGURE 3.10: Top left: OMNO2d tropospheric NO<sub>2</sub> columns (NO<sub>2</sub>: molec cm<sup>-2</sup>) averaged into  $2^\circ \times 2.5^\circ$  horizontal bins for Jan, 2005. Right: Scatter plot of NO<sub>2</sub> against smearing calculations from GEOS-Chem ( $\hat{S}$ ), with points above and below the smearing threshold range of 900–5200 s coloured red and blue respectively. Points are binned by NO<sub>2</sub> with and without the smearing filter applied (orange and magenta respectively). Overplotted is the mean and standard deviation (error bars) within each bin. Due to the logarithmic Y scale we only show the positive direction of standard deviations for unfiltered data. Bottom left: Daily NO<sub>2</sub> scattered against smearing with (magenta) and without (orange) applying the smearing filter. This plot is a zoomed out version of the right panel.

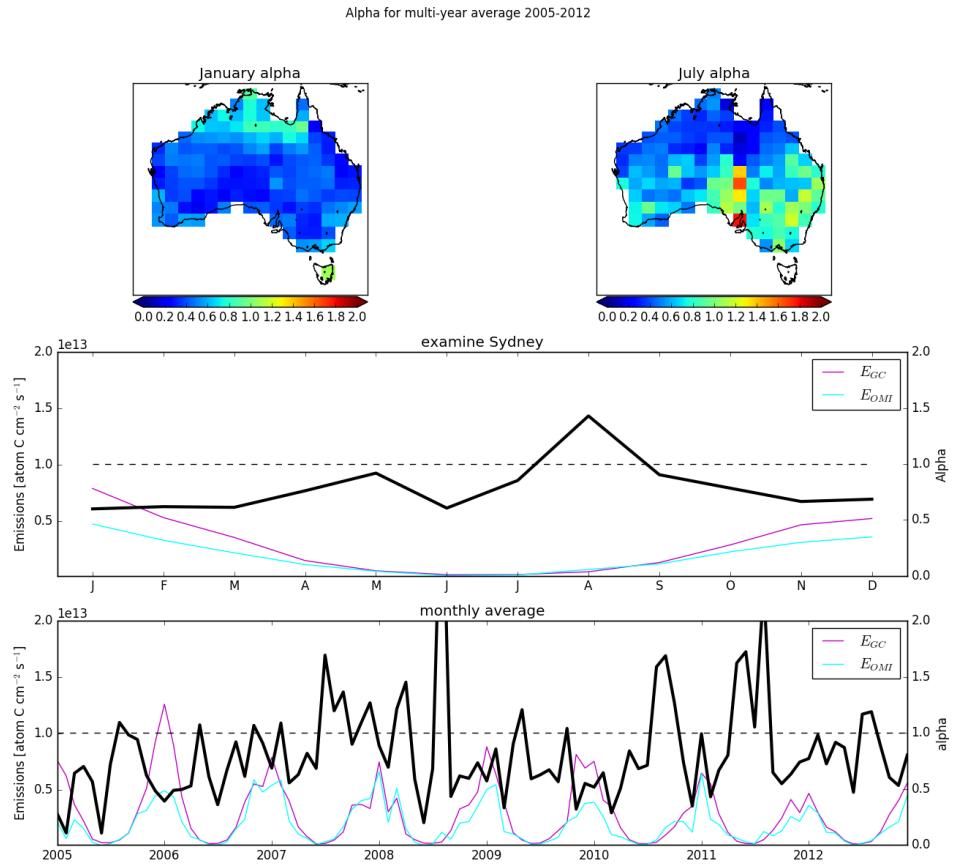


FIGURE 3.11: Row 1:  $\alpha$  for the average January (left) and June (right) over 2005–2012. Row 2:  $E_{GC}$  (xxx, left axis),  $E_{OMI}$  (xxx, left axis), and  $\alpha$  (black, right axis) multi-year monthly averages calculated for Sydney. Row 3: Monthly averages of the same terms in Row 2.

The scaling factor  $\alpha$  is the ratio between the multi-year averaged monthly emissions from GEOS-Chem  $E_{GC}$  and the a posteriori  $E_{OMI}$ :

$$\alpha = \frac{E_{OMI}}{E_{GC}} \quad (3.10)$$

A small modification of GEOS-chem source code multiplies biogenic emissions by  $\alpha$  for each grid-square based on the default emission factors and meteorology. Initially  $\alpha$  is uniformly set to 1 globally. Where we have top down emissions and  $E_{GC}$  is non-zero, we can set  $\alpha$  using Equation 3.10. Missing values for  $\alpha$  when  $E_{GC}$  are zero are an issue, however this is negligible since the dominant discrepancies between estimates occurs during summer when high emission rates are overestimated. Figure 3.11 shows  $\alpha$  for the average January and June over 2005–2012, along with a time series of  $E_{GC}$  and  $E_{OMI}$  and  $\alpha$  calculated for Sydney, and their multi-year seasonal average.

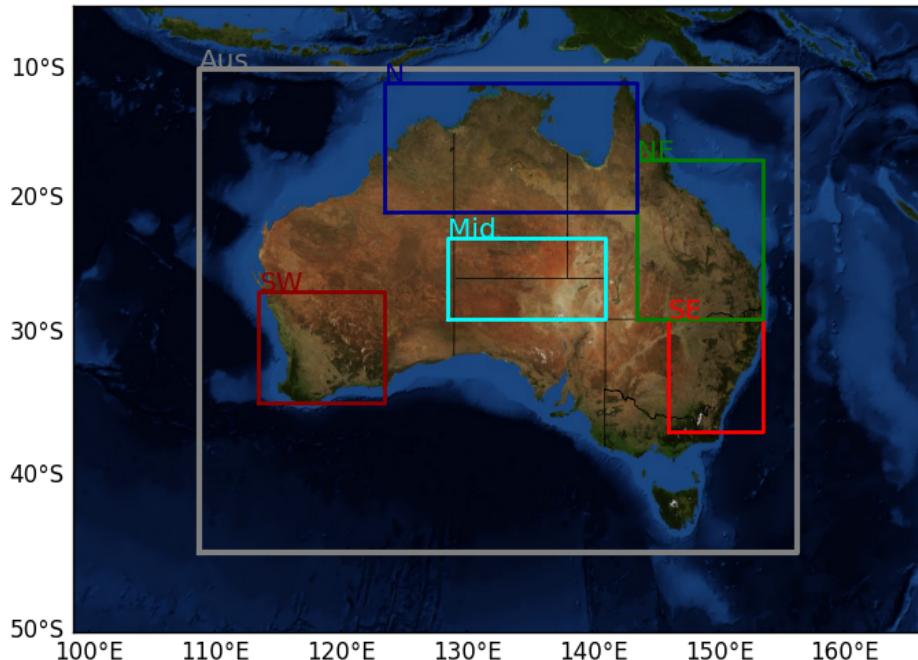


FIGURE 3.12: Sub-regions used in subsequent figures. Australia-wide averages will be black or grey, while averages from within the coloured rectangles will match the colour shown here.

### 3.3 Results

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

#### 3.3.1 Updated emissions

Figure 3.13 shows a priori emissions over Australia and the updated emissions calculated using OMI. The figure shows the time series (daily midday averages) their absolute differences (a priori - a posteriori) over 2005. The seasonal cycle and the difference between the two metrics is clear, with much higher spring and summer averages in the a priori. The northern region also shows a departure from the Australian mean in the a priori data from mid March through to November, which is not seen in the a posteriori. This overestimate by the a priori may be caused by overly high emission factors pushed up by high temperatures during the dry season. January and February also show a low feature in the a posteriori data, which may be due to filtering (TODO check data count), or may have been caused by meteorological factors (TODO Check if drought or fires occurred). This figure is repeated using monthly means (of the daily

midday estimates) in Figure 3.14 over 8 years from 2005 to 2012. Over the entirety of Australia the seasonal cycle of emissions is shown to be overestimated by the a priori. The difference between a priori and a posteriori estimates differs between regions, and some noise can be seen in the monthly time series due to data loss (TODO: check data counts). While most regions generally adhere to this behaviour of overestimation, the northern region of Australia follows a different cycle of bias. Northern Australia appears to be overestimated throughout the year, however occasional low points occur in summer in which the a priori is near to or lower than the a posteriori. One potential reason is the onset of the wet season (November-April) which disrupts the ecosystems and limits satellite measurement through increased cloud coverage.

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

Figure 3.15 shows the multi-year monthly mean and IQR of daily midday isoprene emissions estimates, averaged regionally. Months outside of May to August show the a posteriori lower than the a priori, except in the south eastern region. Here the difference is only prominent during February, TODO check if this may be because of data point counts. TODO: Which regions show best correlation? Worst? Why might that be, and what does it mean?

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

### 3.3.2 A posteriori GEOS-Chem output

This section examines how outputs from GEOS-Chem change after applying a scaling factor  $\alpha$  (see Section 3.2.8) to biogenic isoprene emissions. Here we need to be careful of notation, outputs from the scaled GEOS-Chem run will be denoted by a superscript  $\alpha$ . For example column HCHO from GEOS-Chem before and after scaling are  $\Omega_{GC}$  and  $\Omega_{GC}^\alpha$  respectively.

Figure 3.18 shows the multi-year monthly mean daily cycle of isoprene emissions from GEOS-Chem, along with the top-down midday estimate, for an example grid box located 2.5°West of Sydney. TODO: Analysis of this and lead into more analysis of differences

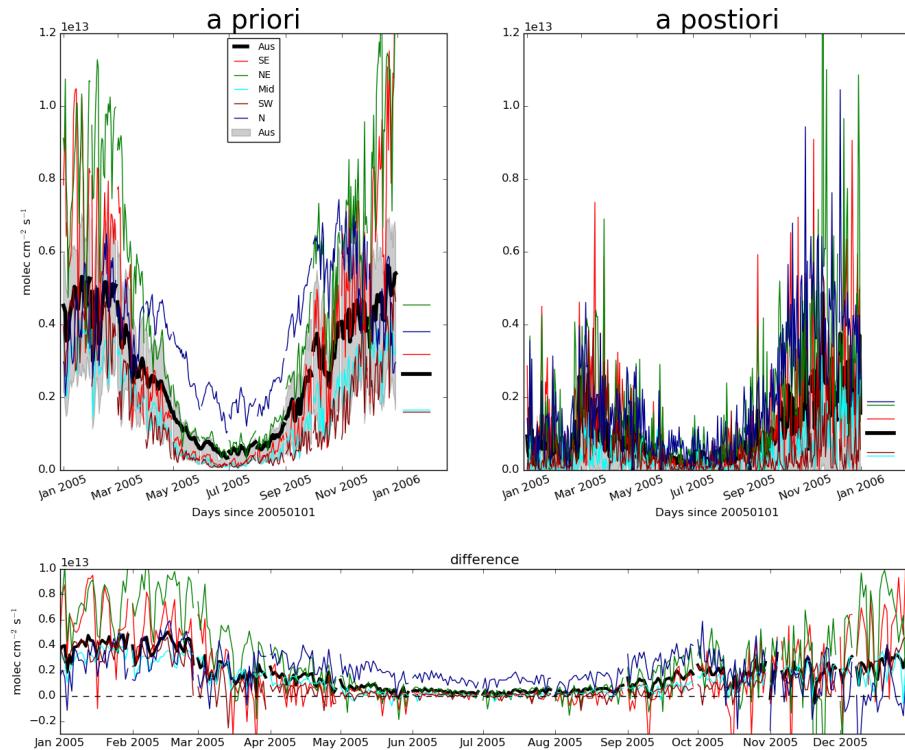


FIGURE 3.13: Biogenic emissions of isoprene from GEOS-Chem (running MEGAN), which is our a priori in column 1, and the emissions calculated using the OMI top down inversion, which is our a posteriori in column 2. Row 1: time series of daily midday averages for all of Australia along with several subregions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

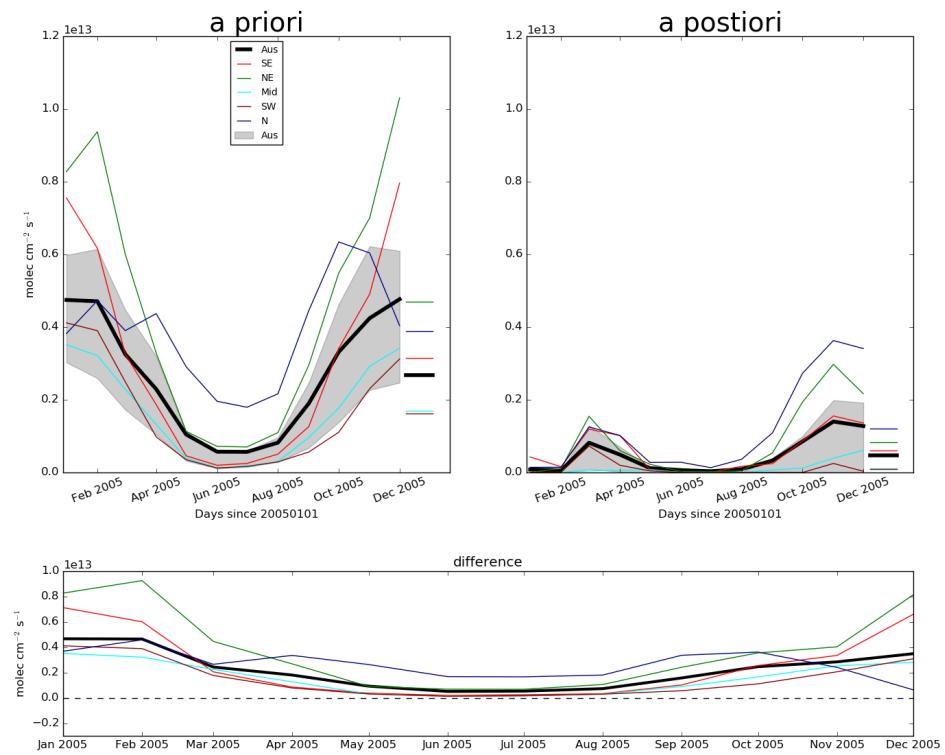


FIGURE 3.14: As Figure 3.13 using monthly medians of the daily mid-day emissions estimates.

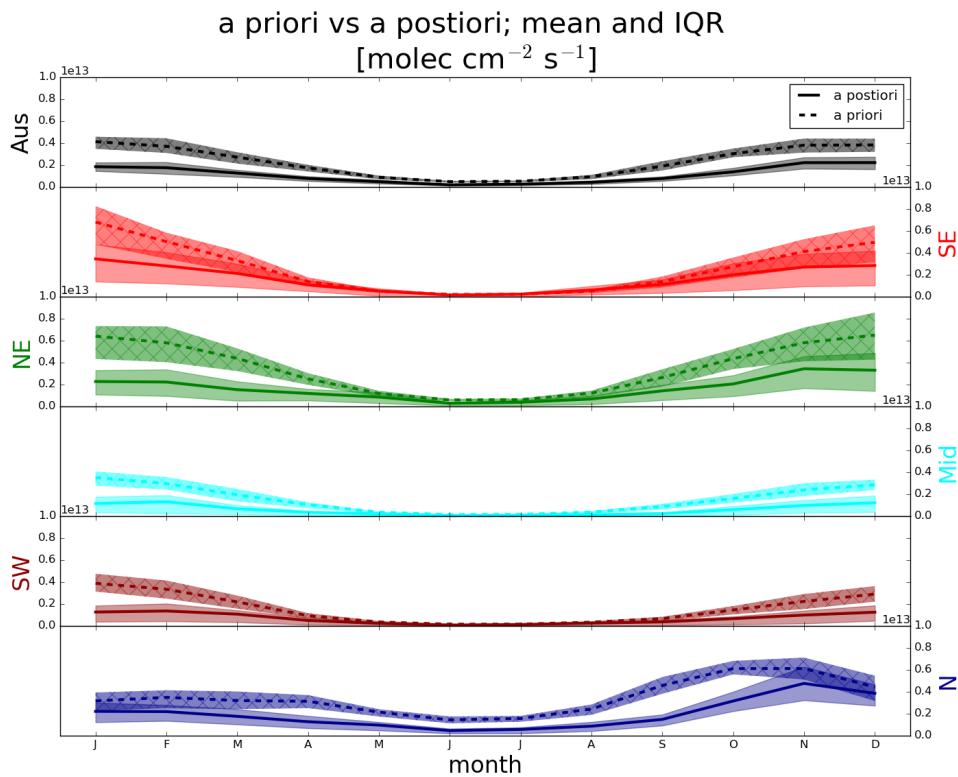


FIGURE 3.15: The multi-year monthly mean (lines) and IQR (shaded) of midday (13:00-14:00 LT) isoprene emissions estimates. A priori emissions are shown by the dashed lines and hatched shaded areas show the IQR. A posteriori emissions are shown using the solid lines, with IQR shown by unhatched shaded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

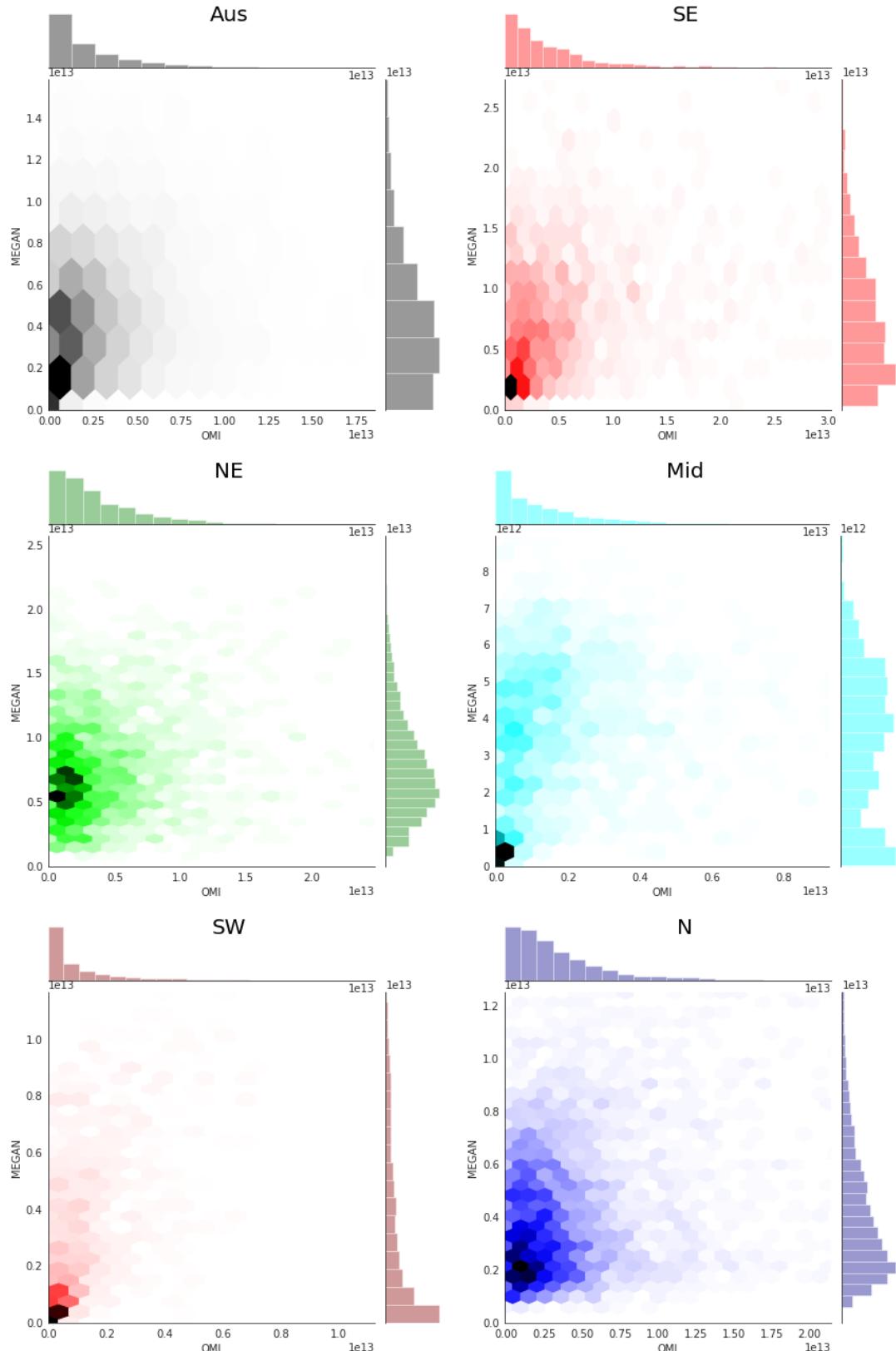


FIGURE 3.16: Scatter plot (binned hexagonally to show data-point frequency) along with the distributions of MEGAN (y axis) and the top down estimate (x axis). This figure is based on summer (DJF) midday values over multiple years. Coloured by regions shown in Figure 3.12.

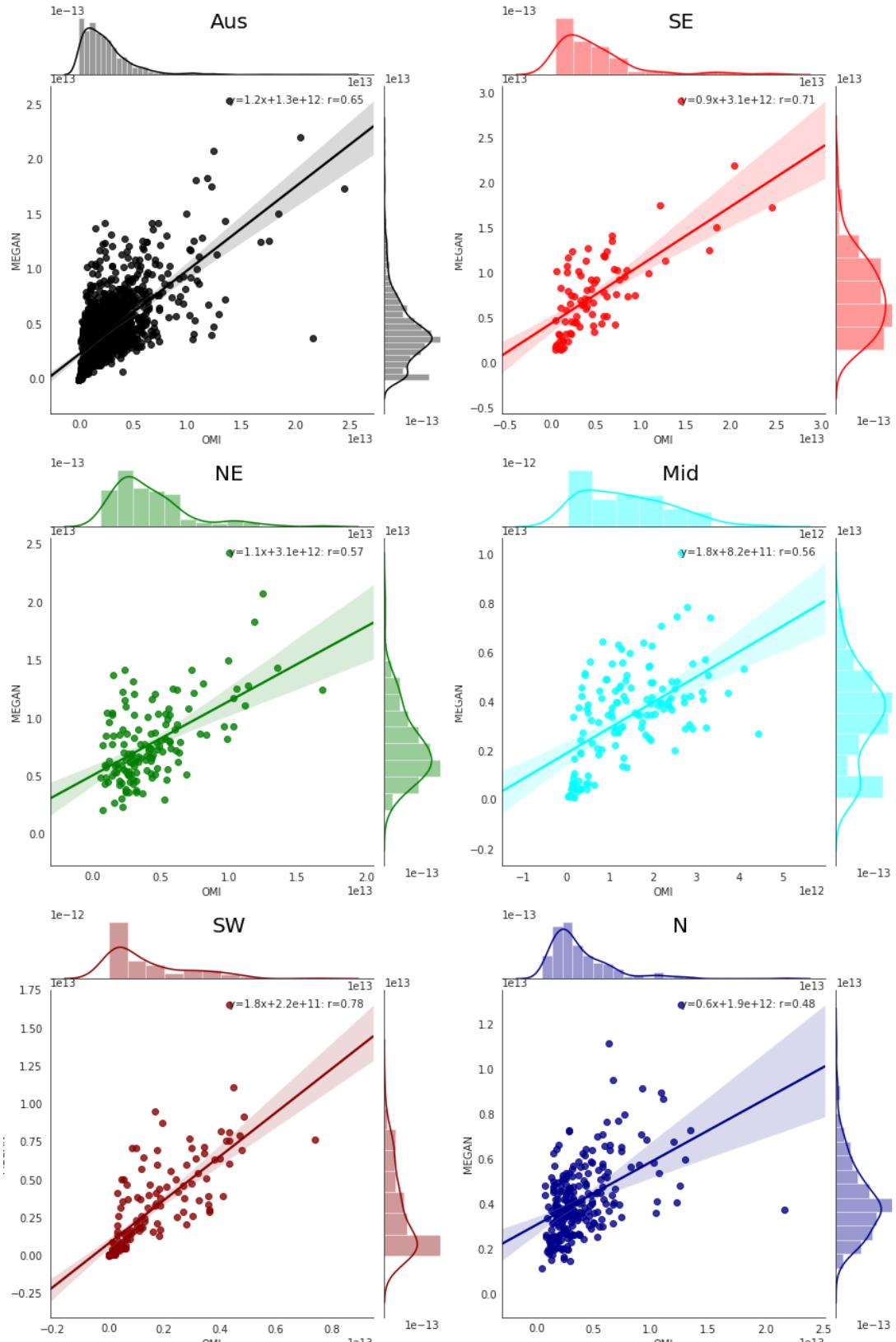


FIGURE 3.17: Scatter plot of a priori emissions against a posterior using monthly averaged grid squares as regression datapoints. Data points are created using monthly averages (of midday emissions) for each grid box for each month of summer (DJF) within each region shown. Multiple years of data are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

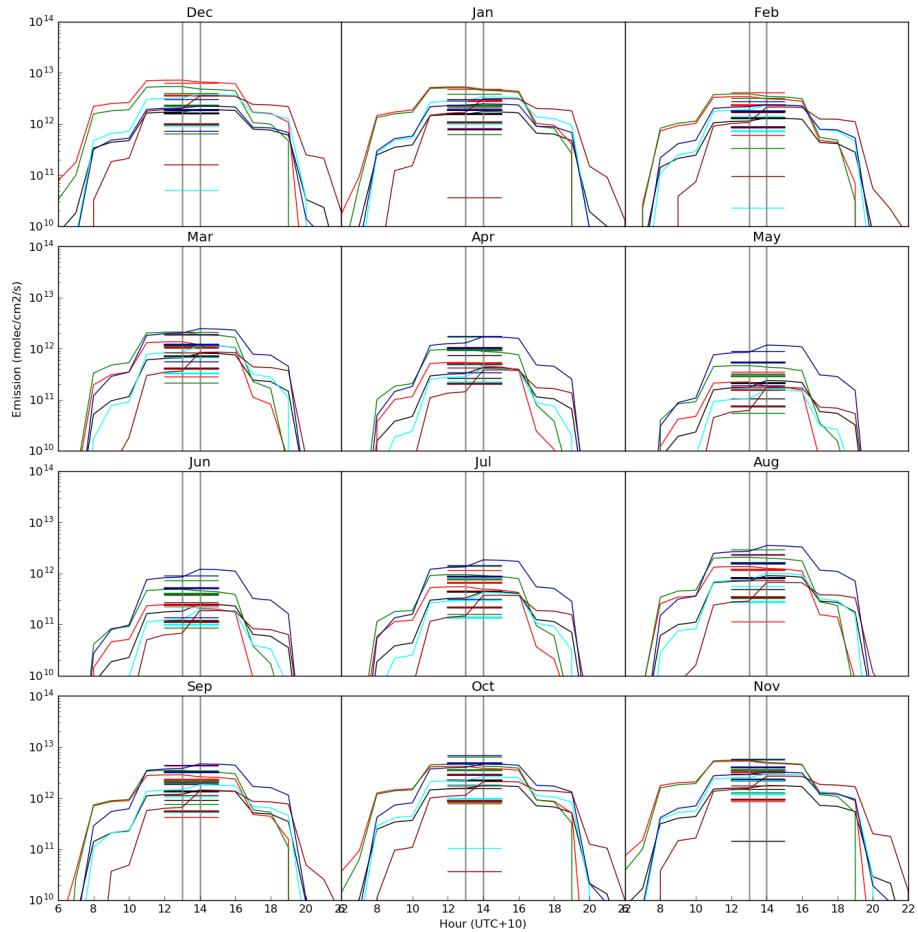


FIGURE 3.18: The diurnal cycle of MEGAN emissions averaged by month over 1, Jan, 2005 to TODO, 2013 are shown with lines, while top-down emissions estimates are shown with plus symbols. MEGAN emissions are estimated hourly per  $2^\circ \times 2.5^\circ$  horizontal grid box, shown here are the area averages (denoted by colour, see Figure 3.12). Top-down estimates are similarly grouped by colour, and shown at the 13:00 LT mark for each month.

### 3.3.2.1 HCHO levels

After running GEOS-Chem with isoprene emission scaled to match the a posteriori, we first check  $\Omega_{GC}$  and  $\Omega_{GC}^\alpha$  outputs against  $\Omega_{OMI}$  over January in 2005 in Figure 3.19. In every region  $\Omega_{GC}^\alpha$  is closer to the satellite measured amounts with biases lowering from 50-120% down to 30-50%. Multiple years are combined to form a similar plot showing the same metrics seasonally in Figure TODO.

We also examine the effects of scaling isoprene emissions on the correlation between modelled and satellite based HCHO columns. Figure TODO: shows the regressions between GEOS-Chem tropospheric column amounts of HCHO and satellite columns for two runs of GEOS-Chem: a) using standard MEGAN emissions, b) using our updated emissions. The a posteriori emissions interpolated over the day are lower than a priori GEOS-Chem emissions by X% per year (from XX to YY Tg yr<sup>-1</sup>) over Australia. GEOS-Chem HCHO at midday is reduced by X% when isoprene emissions are scaled to down.

Wollongong FTIR measurements (see Section 2.2.3.4) provide vertical profiles of HCHO which can be converted to total column amounts using modelled air densities. This is the only non-satellite long-term record of vertical profile HCHO available in Australia and we use it here to examine trends and seasonality. The time series for HCHO is shown in figure TODO, along with GEOS-Chem output before and after updating isoprene emissions. TODO: discuss plot here

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

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

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

### 3.3.2.2 Ozone levels

Isoprene breaks down in the atmosphere and can lead to increased ozone concentrations, especially when mixing with polluted urban air masses which contain high NO<sub>x</sub> concentrations. Figure 5.3 shows surface level (up to ~ 150 m altitude) ozone concentrations over 2005 before and after scaling modelled isoprene emissions. Reducing isoprene emissions lowers surface ozone concentrations by TODO: XX to YY % in summer, and XX to YY % in winter. The direct correlation between reduced emissions and surface ozone (Figure TODO) shows that reduced emissions directly relate to reduced ozone concentrations, and using multiple years enhances/reduces the relationship. A look at the regression using only summer data reveals a stronger/weaker relationship, analyse per region (TODO).

TODO Figure showing isoprene vs ozone time series and correlation.

The relatively clean atmosphere over the majority of Australia is likely to be NO<sub>x</sub> limited outside of major population centres. This means that cities may have higher sensitivity to changes in modelled isoprene emissions than shown here, since we are examining large area averages which are mostly non-urban. TODO: Notes about ozone exposure from State of the Environment report 2016.

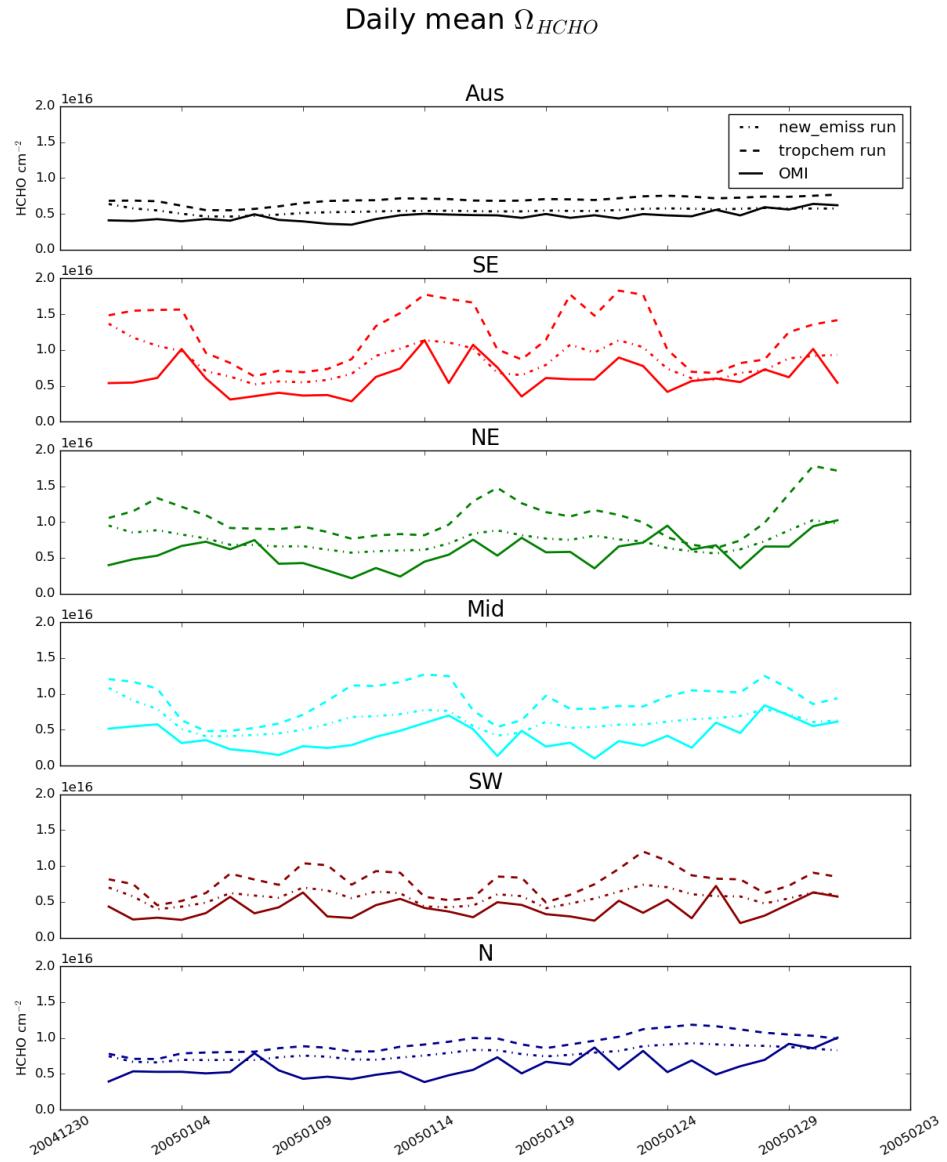


FIGURE 3.19: Daily mean total column HCHO amounts from GEOS-Chem with (new emissions run) and without (tropchem run) a posteriori scaled isoprene emissions, along with the original OMI HCHO columns. Each row shows the average over regions in Figure 3.12.

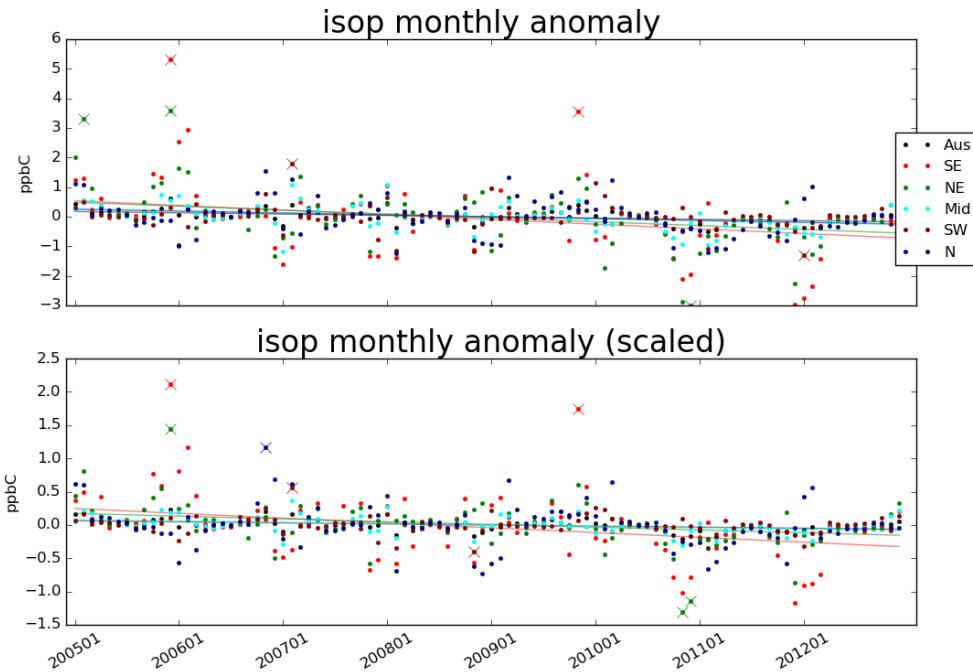


FIGURE 3.20: A priori (row 1) and a posteriori (row 2) emissions anomaly from multiyear monthly mean, split by region (see Figure 3.12).

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

### 3.3.2.3 Trends

Figure 3.20 shows monthly deseasonalised a priori and a posteriori midday emission anomalies for each region. First the emissions are spatially averaged within each region to form a daily time series of midday emission rates, this is averaged into monthly data, and then the multiyear monthly mean is subtracted to form the anomaly series. Any anomaly which is greater than 3 standard deviations from the mean is removed (these points can be seen as crosses in Figure 3.20). An ordinary least squares linear regression is then performed to look for any significant trend. A trend is considered significant if the p-value from a Wald test is less than 0.05. The midday emissions show no significant trends over the 8 year period from 2005-2012. The same process is repeated for surface ozone, surface HCHO, and surface NO<sub>2</sub>, with anomaly series shown in Figures 3.21 to 3.23. None of these metrics shows any significant changes in trend due to scaling isoprene emissions. Some trends are no longer significant in surface HCHO, however none of the trends change sign.

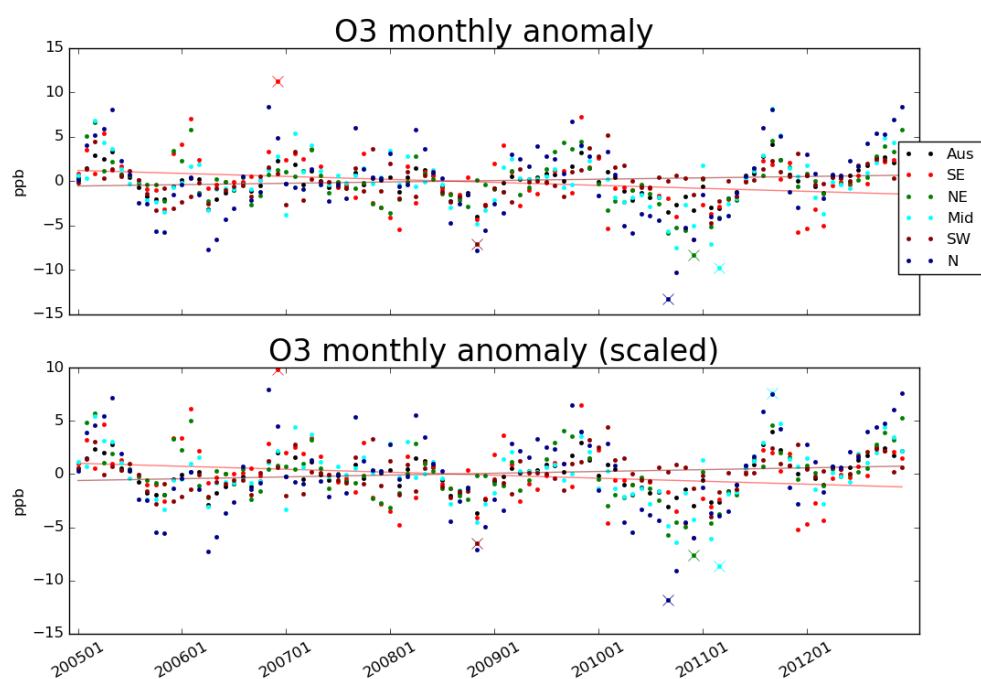


FIGURE 3.21: A priori (row 1) and a posteriori (row 2) surface ozone anomaly from multiyear monthly mean, split by region (see Figure 3.12).

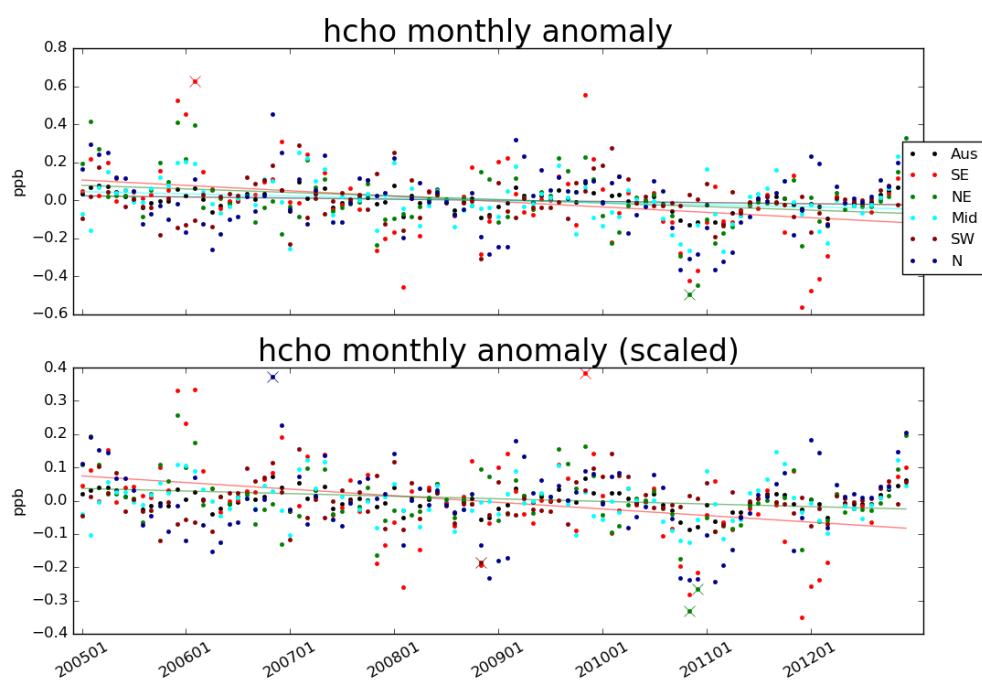


FIGURE 3.22: A priori (row 1) and a posteriori (row 2) surface HCHO anomaly from multiyear monthly mean, split by region (see Figure 3.12).

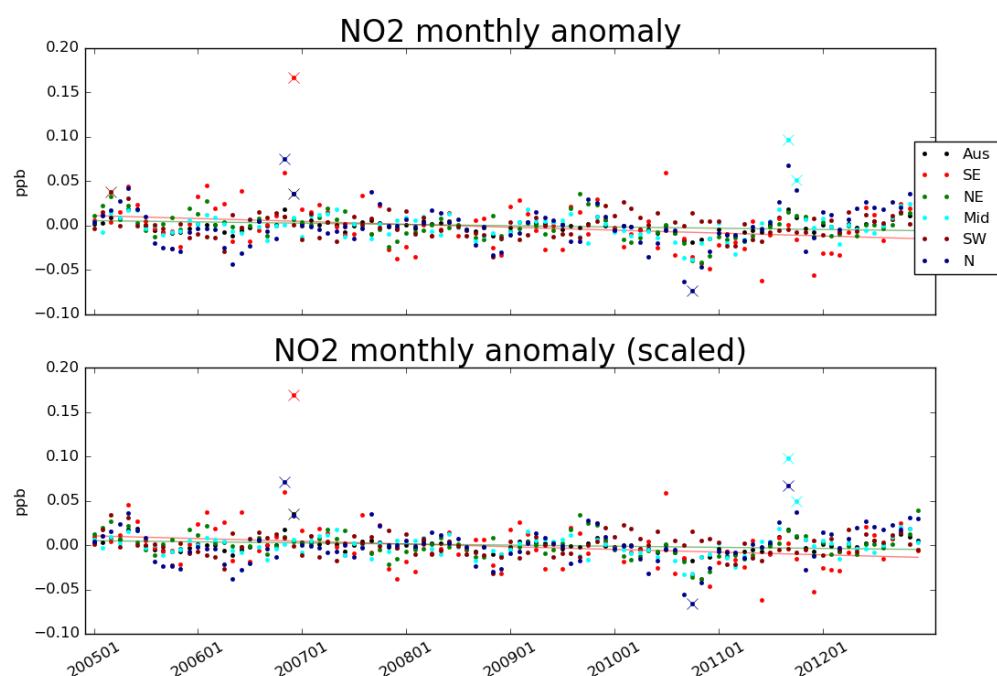


FIGURE 3.23: A priori (row 1) and a posteriori (row 2) surface  $\text{NO}_2$  anomaly from multiyear monthly mean, split by region (see Figure 3.12).

TABLE 3.3: Isoprene emissions (Tg/yr)

Australia	notes
43(2)	a) GEOS-Chem: 2005-2010
19(2)	b) Top-down: 2005-2010
~ 80	c) Sindelarova et al. (2014): 1980-2010
26-94	d) Bauwens et al. (2016): 2005-2013

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

### 3.3.3 Literature comparisons

When comparing GEOS-Chem (which runs MEGAN) emissions to those calculated using our top-down inversion, we see a decrease of around  $29 \text{ Tg yr}^{-1}$ (66%). Sindelarova et al. (2014) see a  $41 \text{ Tg yr}^{-1}$ decrease in Australia when introducing soil moisture parameterisation. Global and Australian emissions estimates for isoprene still range widely due to the emission models sensitivities to inexact parameters including leaf area indices, plant functional type emission factors, and meteorological factors. Table 3.3 shows yearly isoprene emissions from this work and some other works for Australia. 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 top-down estimate performed in Bauwens et al. (2016) of  $26 \text{ Tg yr}^{-1}$ . Figure 3.24 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.

### 3.3.4 Comparison with available measurements

TODO: Analyse comparison of grid box with campaigns of measurements

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

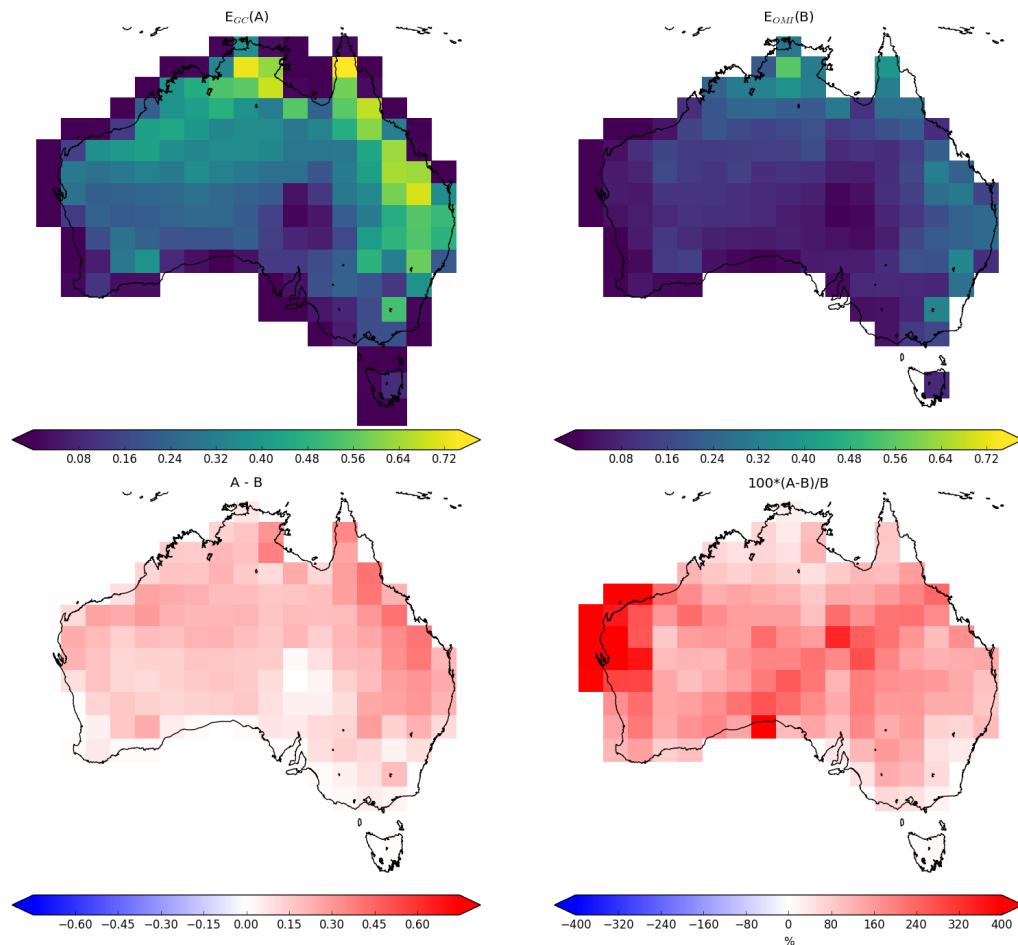


FIGURE 3.24: Top row: multi-year mean emissions in  $\text{Tg yr}^{-1}$  from  $E_{GC}$  (GEOS-Chem; running MEGAN) and  $E_{OMI}$  (top-down emissions) respectively.  $E_{OMI}$  uses an assumed sinusoidal daily cycle, with day-light hours prescribed for each month: see Section 3.2.6). Bottom left and right shows the absolute and relative differences respectively.

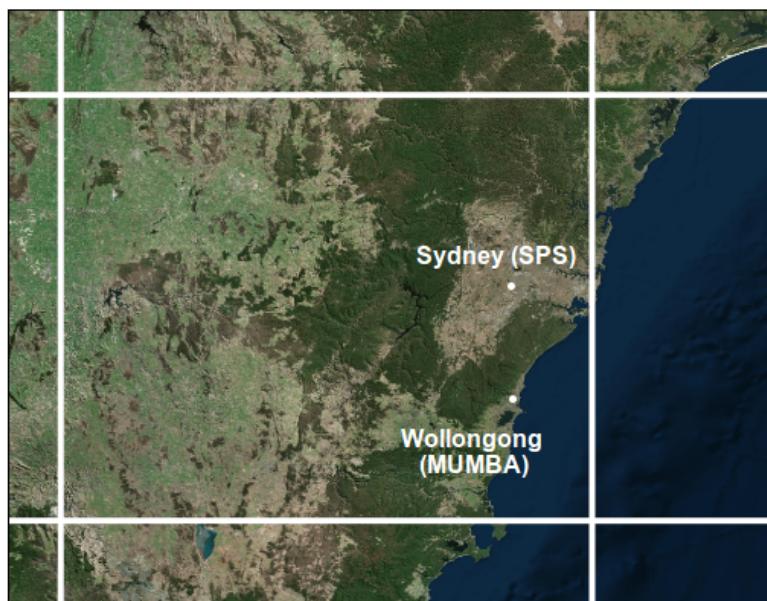


FIGURE 3.25: grid box at  $2^\circ \times 2.5^\circ$  containing SPS, and MUMBA campaign data.

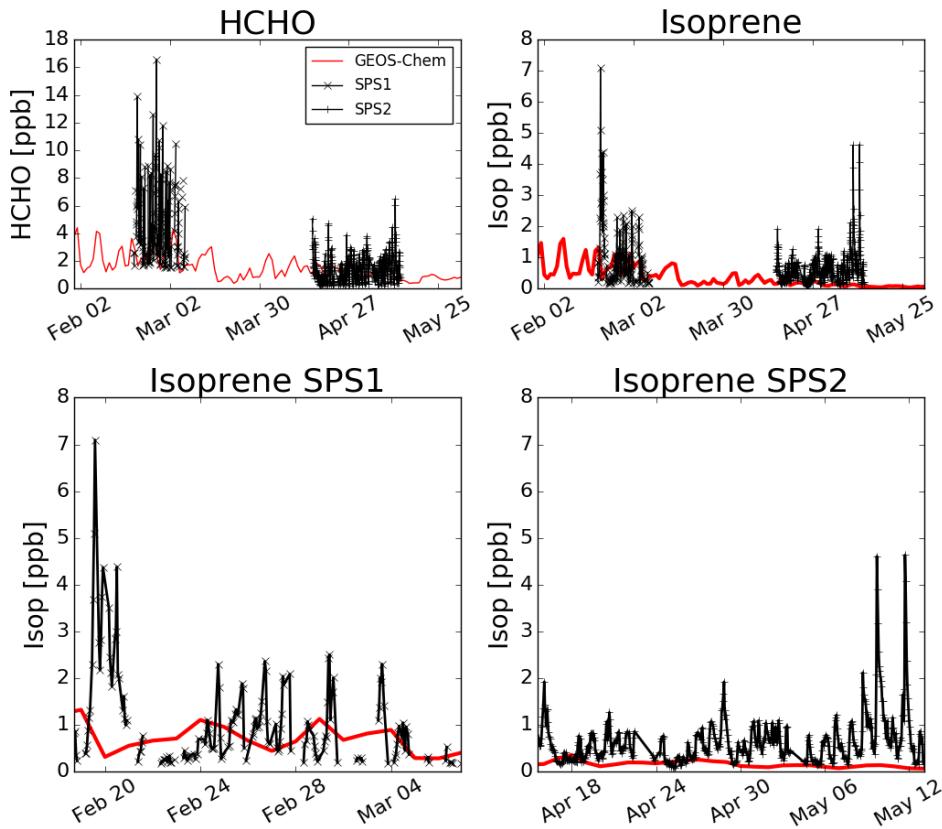


FIGURE 3.26: Comparison between GEOS-Chem HCHO concentrations in the grid square containing Sydney for the duration of the SPS 1 and 2 campaigns

Figure 3.26 shows GEOS-Chem output in the grid square containing Sydney overlaid on SPS measurement data. Superficially the comparison is not too bad between these two datasets. This figure shows at a glance how the measurements compare to modelled data, and also the limited nature of the temporal and spatial scale of available measurements. TODO: if possible use midday isoprene multi-year averages, and error bars? then update discussion. The SPS data is point-source and taken during the daytime when isoprene is higher, so it is very likely that GEOS-Chem HCHO and isoprene output is in fact too high since the daily average should not match the peak of the measurements.

Figure TODO shows how isoprene compares against measurements from the MUMBA and SPS campaigns (described in Section 2.2) both before and after scaling isoprene emissions to match the top-down estimation. TODO: discuss results and differences, does isoprene improve?

MUMBA data captured concentrations at the surface in Wollongong, and is compared in figure TODO against simulated HCHO concentrations in the first level ( $\sim 0$  - 150 metres) over corresponding dates. GEOS-Chem midday concentrations are shown,

along with Wollongong bottom-level FTIR measurements to compare against the MUMBA data. TODO Discussion of comparison.

### 3.3.5 HCHO Products and yield

Isoprene reaction chains are diverse, with many branches forming HCHO. HCHO production yields are often classed into two categories: first generation HCHO yield and total (or molar) yield. First generation yield refers to the amount of HCHO produced per unit isoprene consumed by initial oxidation. Total yield refers to time dependent yield of HCHO over multiple oxidation stages (Wolfe et al. 2016). In this work yield ( $Y_{isop}$ ) is not calculated as the slope ( $S = Y_{isop}/k_{HCHO}$ ) can be used instead. This section shows a brief analysis of HCHO lifetimes ( $\tau$ ) at midday.

By assuming yield  $Y_{isop}$  lies between 0.2 and 0.4, we find a range for midday lifetimes of HCHO using equation 3.6:

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

$\tau$  is heavily influenced by assumed yield, and improved methods of estimating yield over Australia are required to improve this estimate. Figure 3.27: shows the GEOS-Chem HCHO lifetime estimated throughout the year. There is a clear seasonal cycle with longer lifetimes in winter months. A clear June (and sometimes March, July and August) increase in HCHO lifetimes is shown, which is caused by the reduced winter HCHO concentrations, temperature, and insolation. These factors may limit the utility of any top-down emissions estimation technique using HCHO in the winter months. The figure is produced using filtered slope information from 2005, and outliers along with low data availability in some months is an issue. Noise in the southwest and middle regions may be indicative of heavy filtering, potentially driven by westerly winds which can bring transported pollution and also lead to smearing.

## 3.4 Uncertainty

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

This section identifies and quantifies the overall uncertainties of calculating isoprene emissions using OMHCCHO and GEOS-Chem in the top-down method. The major source of uncertainty lies in TODO, which is due to TODO. Uncertainty in satellite HCHO is seasonally dependent, with better signal during the summer. Good OMI measurements are less frequent at high solar zenith angles, which depends on latitude

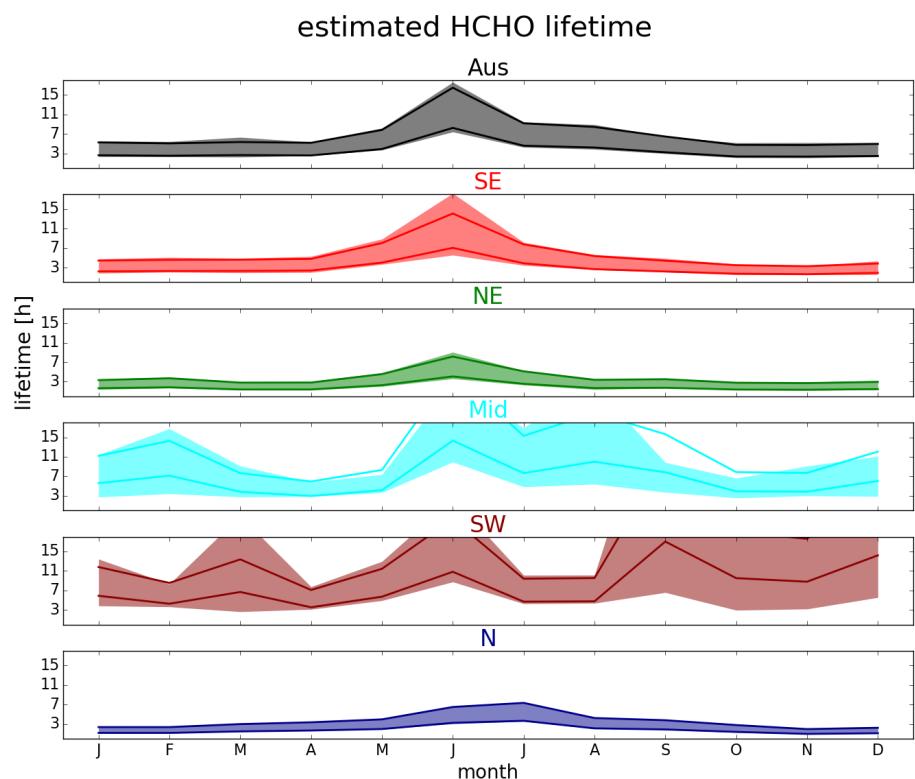


FIGURE 3.27: Monthly area averaged HCHO lifetime ( $\tau$  in hours), with IQR shaded. Solid lines show lifetime assuming yield is 0.2, and 0.4 (higher and lower lines respectively). Coloured by regions shown in Figure 3.12.

TABLE 3.4: Uncertainty estimations.

Region	$\Delta E_{OMI}$ Summer	$\Delta \Omega_{OMI}$	$\Delta S$	$\Delta E_{OMI}$ Winter	$\Delta \Omega_{OMI}$	$\Delta S$
Aus						
SE						
NE						
...						

as well as season, which exacerbates the winter uncertainty. Table 3.4 shows the estimated uncertainty calculated in this work in summer and winter over each region described by Figure 3.12. How each of these terms is calculated is laid out in the following subsections.

### 3.4.1 Top down emissions

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

The final determination of top-down emissions comes from equation 3.7:

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

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

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

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

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

To quantify  $\Delta E_{OMI}$  we determine 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 3.6) we examine related GEOS-Chem output in Section 3.4.2, since  $S$  comes directly from the monthly linear regression of modelled isoprene emissions and column HCHO. Uncertainty in terms

$\Omega_{OMI}$ , and  $\Omega_{OMI,0}$  come from both the OMHCHO slant columns ( $SC$ ) and the  $AMF$  calculated to transform them into vertical columns ( $\Omega = \frac{SC}{AMF}$  from equation 2.7). Section 3.4.3 describes these calculations.

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

### 3.4.2 Model Uncertainty

$E_{OMI}$  depends partly on the product it is trying to improve, as modelled yield is based on MEGAN run by GEOS-Chem. The uncertainty in the RMA regression slope between model HCHO and emissions ( $\Omega_{OMI}$  and  $E_{GC}$  respectively) is used to estimate the  $\Delta S$  within Equation 3.14. This is a simple method of approximating the uncertainty of this term, only accounting for monthly uncertainty of the slope calculation. It does not take into account uncertainty in the underlying model, nor uncertainties arising from temporal or spatial resolution issues. Figure TODO shows the multi-year mean summer and winter  $\Delta S$  over Australia. Analysis todo For comparison Palmer et al. (2006) found  $\frac{\Delta S}{S}$  to be 30%, after comparison with another chemical model and in situ measurements.

To improve understanding of uncertainty in  $S$  would require further analysis of GEOS-Chem yield and sensitivities over Australia, ideally along side representative measurements. Here we present a brief analysis of how GEOS-Chem HCHO and isoprene concentrations and variances compare to in situ measurements from MUMBA, SPS1, and SPS2. Comparison caveats are that in situ point measurements are quite different to modelled (large) area averages, and that these measurements only take place in one grid box within Australia. We see model bias of TODO: difference for isoprene conc before and after, HCHO conc before and after, in the  $\sim 4$  months of available measurements. Figure TODO shows the measured and modelled metrics for the comparable times.

Filtering for spatial smearing (see Section 3.2.7) reduces the number of data points making up our regression slope  $S$ . The process generally improves the linear relationship, and where it does not a multi-year average (or in the worst cases no value at all) is used in lieu of monthly  $S$  data. The filtering process is accounted for in the calculation of  $\Delta S$ , and the overall effect on uncertainty is limited, so we do not analyse it further.

Model biases are not analysed in this thesis, except to note that they would impact both preliminary OMI calculations as well as affecting the modelled slope.

### 3.4.3 Satellite Uncertainty

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

1. Fitting error from the OMI retrieval.
2. Uncertainty in AMF calculations.
3. Uncertainty of HCHO background.

TABLE 3.5: Uncertainties in literature and here.

uncertainty	location	notes
40%	North America	(a) mostly due to cloud interference
26%	North America	(b) with cloud fraction less than 20%
30-40%	global	(c) GOME-2 instrument
> 60%	Mid-latitude	(c) GOME-2 instrument in winters
X%	Australia	accumulated uncertainty in calculation
X%	Australia	range found when scaling satellite HCHO

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)

Fitting error is provided by the OMI product and reduced through spatial and temporal averaging. AMF uncertainty can be determined through comparison of GEOS-Chem output to measured HCHO columns. Here we use 30% as a rough estimate of error in this term, since measurements over Australia are lacking. Since we expect oceanic background HCHO to be invariant, variance in remote ocean HCHO can be used as a rough estimate of background uncertainty. We explicitly calculate background error in the OMI HCHO columns ( $\Delta\Omega_{OMI,0}$ ) in the same way as we calculate  $\Delta\Omega_{OMI}$ . Uncertainty in satellite HCHO ( $\Delta\Omega$ ) from literature and calculated here is listed in Table 3.5.

Provided with the OMI product is the measurement of uncertainty in each pixel, calculated by SAO from the backscattered solar radiation fit (Gonzalez Abad et al. 2015; Abad et al. 2016). Uncertainty introduced through AMF calculation needs to be additionally determined to give a representation of the confidence in vertical column amounts. We use the uncertainty in each pixel averaged over our  $2^\circ \times 2.5^\circ$  grid squares for each day and divided by the square root of the number of pixels. This is considered to be the error in the slant column ( $\Delta SC$ ), which is combined with our assumed relative AMF error ( $\frac{\Delta AMF}{AMF}$ ) of 30%.  $\Delta\Omega_{OMI}$  and  $\Delta\Omega_{OMI,0}$  are calculated for Equation 3.14 through applying the quadrature rule from Equation 3.12 as follows:

$$\Omega = \frac{SC}{AMF}$$

$$\Delta\Omega = \Omega \sqrt{\left(\frac{\Delta SC}{SC}\right)^2 + (0.3)^2} \quad (3.15)$$

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

Palmer et al. (2006) calculate the error in AMF through combining estimates of error in the UV albedo database (~ 8%), model error based on in situ measurements, cloud error (20 – 30%) (Martin et al. 2003), and aerosol errors (< 20%), totalling AMF error of around ~ 30% (calculated in quadrature). TODO: calculate this uncertainty. Compare this error estimate with that of Curci et al. (2010), where the error in b) and c) 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. One method of finding overall satellite uncertainty involves using the variance over the remote Pacific (15°S to 15°N, and 180°W to 120°W) (De Smedt et al. 2012; De Smedt et al. 2015). In the remote Pacific, it can be assumed that HCHO variations are weak, with concentrations remaining steady in the short term ( $\sim$  1 month). This means the standard deviation over this region can be used as a proxy for determination of the instrument error. De Smedt et al. (2012) found satellite HCHO uncertainty to be 30 – 40% for the GOME-2 instrument by combining slant column systematic and random errors. For mid latitude winters they saw an excess of 60% uncertainty.

In order to calculate the bias or systematic error, an understanding of biases in the underlying terms is required, since there is little in the way of comparable measurements. OMHCHO has been seen to underestimate HCHO by up to 40%, although this amount is regionally dependent (Zhu et al. 2016; De Smedt et al. 2015; Barkley et al. 2013)). Satellite HCHO may also suffer from  $\sim$  13% overestimation when taking a monthly average due to only measuring on relatively cloud-free days (Surl, Palmer, and Abad 2018). Since our *a posteriori* is linearly related to the satellite HCHO, any bias is directly transferred. We note therefor that our product may be biased by up to 40% due to satellite bias. This may be complicated further if the satellite bias over Australia does not match that of measurements over the remote Pacific at corresponding latitudes. However we can not quantify bias over Australia due to insufficient measurements. GEOS-Chem biases would affect the recalculation of HCHO, but they are assumed to be minimal in this thesis. Since we lack sufficient suitable measurements to estimate satellite bias over Australia, we assume bounds on potential satellite biases based on literature of up to 40%.

#### 3.4.3.1 Sensitivity to vertical column recalculation

OMI HCHO vertical columns are recalculated using GEOS-Chem V10.01 *a priori* HCHO and air density profiles (see Chapter 2). The recalculation is assumed to be no more or less uncertain than that used to calculate the product provided AMF. The pixel SC uncertainty from the OMHCHO product is not changed by recalculating the AMF, and is used as is, as shown in Section 3.4.3. Here we examine the sensitivity of the isoprene emissions estimation technique to the AMF recalculation method. Through looking at how the emissions change based on whether we use the AMF provided ( $\text{AMF}_{\text{OMI}}$ ), the AMF with shape factor recalculated but the default scattering weights ( $\text{AMF}_{\text{GC}}$ ), or the fully recalculated AMF ( $\text{AMF}_{\text{PP}}$ ).

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

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

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

### 3.4.3.2 Sensitivity to filtering

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

## 3.5 Conclusions and implications

TODO: Brief summary of results and implications...

Our top down biogenic isoprene emission estimate shows that MEGAN overestimates emissions in summer by up to a factor of 5. The overestimation is spatially and temporally diverse and leads to model overestimation of HCHO. By applying a monthly gridded scaling factor based on the multi-year monthly difference model emissions and our top down a posteriori, HCHO bias can be improved from 50-120% down to 30-50% (TODO: update that number for summer season rather than the 2005 January it is based on).

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

What else?