

Chapter 3

Biogenic Isoprene emissions in Australia

3.1 Introduction

Australian forests are strong emitters of both isoprene and monoterpenes. However emissions are poorly understood due to poor measurement coverage. The lack of measurements makes it difficult to estimate important subsequent processes such as ozone and secondary organic aerosol (SOA) formation. Isoprene has a large impact on the oxidative properties of the atmosphere, as it reacts quickly with the OH radical. One frequently used model of biogenic volatile organic compound (BVOC) emissions is MEGAN (Guenther et al. 2000) which estimates $\sim 1150 \text{ Tg yr}^{-1}$ globally. The primary BVOC emission is isoprene, globally modelled at $\sim 465\text{-}500 \text{ Tg C yr}^{-1}$ (Guenther et al. 2006; Messina et al. 2016). The emission models used to derive these estimates are estimate fluxes from different plant species (phenotypes), which are seldom well understood within Australian forests. Isoprene emissions may be overestimated in Australia since they are based on measurements of taken from a few heavily emitting young eucalyptus trees which are not representative (Winters et al. 2009; Fortems-Cheiney et al. 2012).

Satellite based emissions estimates allow us to improve the models without requiring other expensive measurements over the large data sparse continent of Australia. Kefauver, Filella, and Peñuelas (2014) reviews remote sensing of BVOC emissions, examining the last 20 years of data and analysis of the satellite products. Their review reinforces the message that BVOCs affect the oxidative capacity of the atmosphere and are largely driven by and sensitive to vegetation. In the troposphere, BVOC emissions affect the hydroxyl radical (OH) cycling, ozone (O_3) and secondary organic aerosol (SOA) production, and methane longevity. The chemistry involved is complex and still suffers from relatively large uncertainties in both measurement and chemistry mechanisms.

In this chapter we use and describe a technique using satellite measurements of HCHO to estimate surface isoprene emissions. 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 González Abad 2018), and even globally (Shim et al. 2005; Fortems-Cheiney et al. 2012; Bauwens et al. 2016). HCHO is the dominant product of most BVOCs and is measured by remote sensing. HCHO products can be found in four satellite instruments: GOME on ERS-2, SCIAMACHY on ENVI-SAT, OMI on EOS AURA, and GOME2 on MetOp-A. These satellites have slightly different spectral and spatial resolutions, as well as using varied processes to estimate HCHO from detected radiances. This leads to different estimates between instruments as described in Lorente et al. (2017), and both validation and comparison become more important when using these remotely sensed data. 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

In a prior chapter (chapter 2), the OMI HCHO total columns are recalculated using an updated estimate of HCHO profiles from GEOS-Chem v10.01. These estimates are compared to available datasets of isoprene or HCHO (SPS1, SPS2, MUMBA, Wollongong FTIR). Sensitivity satellite AMF calculation is examined and quantified for some scenarios. 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. In section 3.5 we examine how these changes in emissions would affect ozone concentrations in Australia, along with some other chemical processes. Uncertainties for each step along the way are quantified in section 3.4.

Recent work suggests that modelled emissions may be overestimated in southeast Australia, while emissions of monoterpenes ($C_{10}H_{16}$, two units of isoprene) appear to be underestimated (Emmerson et al. 2016). This could lead to the unique scenario of neither emission type dominating VOC chemistry over the forests. 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 top-down estimates based on OMI HCHO measurements and modelled isoprene to HCHO yields. This a posteriori top-down estimate is used to determine if model fit against sparse available ground-based measurements can be improved. GEOS-Chem model output is examined before and after updating isoprene emissions which are used as inputs. Wellness of fit between in-situ (at Wollongong) HCHO, satellite (OMI), and modelled (GEOS-Chem) HCHO is determined with and without updated emissions estimates.

3.1.2 Top-down isoprene emissions estimates

In the remote troposphere HCHO production is dominated by methane oxidation, while in the continental boundary layer (CBL) production is largely due to non-methane VOCs (NMVOCs) (Abbot 2003; Kefauver, Filella, and Peñuelas 2014). This leads to the technique of using a linear regression between enhanced HCHO and NMVOC emissions. In the CBL, HCHO enhancement is generally driven by short lived (< 1 hr) precursors (most importantly isoprene). HCHO itself has a lifetime of a few hours

(Kefauver, Filella, and Peñuelas 2014). Isoprene is emitted and enters the atmosphere in the gas phase, where it begins a complex series of reactions. Formaldehyde is produced with high yields in many of the isoprene reactions, which are discussed in more detail in Section 1.3.3. HCHO measurements are often used as a check on how well isoprene reactions are simulated, as model output can then be compared against them (Marvin et al. 2017).

We broadly follow the method of Palmer et al. (2001) to create an emissions estimate of isoprene over Australia. In their work isoprene emissions fluxes were derived using the Global Ozone Monitoring Experiment (GOME) satellite instrument, however here we use OMI (on board the AURA satellite) as it has better temporal coverage and increased pixel counts. Palmer's method improved biogenic isoprene emissions estimates (compared with in-situ measurements) over two available inventories: the U.S. EPA Biogenic Emissions Inventory System (BEIS2) and the Global Emissions Inventory Activity (GEIA). Here we try to improve MEGAN emissions estimates over Australia and analyse some of the technique sensitivities.

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. They examine global emissions seen by three models and a top-down inversion, showing a wide range of estimated values for Australia. 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.

3.1.2.1 Bayesian

Bayesian inversion corrects biased biogenic isoprene emissions by optimising emission parameters in order to reduce the difference between observed HCHO and model output. In-depth inversions can account for effects from transport and allow source attribution (Fortems-Cheiney et al. 2012).

For example this method is used by Shim et al. (2005) who optimise GEOS-Chem isoprene emissions in areas with high HCHO concentrations to improve comparison against GOME HCHO observations. They show that the original model underestimates isoprene emissions and HCHO concentrations by 14-46%, with the corrected VOC emissions reducing model biases to 3-25%. The Bayesian inversion is also used in Curci et al. (2010), where a regional CTM (CHIMERE) simulates HCHO, which is compared against OMI observed HCHO and shown to be regionally biased. The model is run initially with emissions of BVOCs and reactive anthropogenic VOCs turned off in order to work out the background (b) values of these compounds. Curci et al. (2010) uses CHIMERE as the forward model to determine the relationship between HCHO (y), isoprene and reactive anthropogenic VOCs (x), using

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

where ϵ are the (assumed) independent errors in measurements. \mathbf{K} is the Jacobian matrix determined from CHIMERE representing the sensitivity of y to the state variable

x. This K matrix is used in conjunction with error covariance in x to determine the Maximum A Posteriori solution to calculate the optimal estimate of x.

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

3.1.2.2 Linear

This technique is the simplest, and is performed in this thesis. Using vertical columns of biogenic HCHO one can infer the local (grid space) isoprene emissions using effective molar formaldehyde yield (In other continents around 2-3, or 1 in low NO_x conditions) (Palmer 2003; Marais et al. 2012; Bauwens et al. 2016). If one assumes fast HCHO yield, so that the effect of chemical transport is minimal, and that HCHO and isoprene are at steady states, then one may calculate local yield from a CTM.

In this work yield is calculated from the modelled slope between isoprene emissions and HCHO tropospheric columns within each gridbox over Australia, as performed in Palmer (2003), using modelled values between 1300-1400 LT which is around the overpass time of the OMI. This modelled yield is then used in conjunction with the recalculated OMI measurements in order to estimate isoprene emissions. To calculate this yield we use a reduced major axis (RMA) regression between modelled average values of the total columns and isoprene emission rates, an example is shown in figure 3.1 shows 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 from four grid-squares. The top down estimation process in this thesis is further explained in section 3.2.

This technique suffers from the assumptions of fast HCHO yield and no transport, which requires filtering areas with low NO_x, high winds, or low emissions. Since we use an estimate of the yield from biogenic isoprene to HCHO, we must also filter out areas where HCHO may be coming from anthropogenic or pyrogenic sources. 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. 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.

In high NO_x environments where HCHO has a lifetime on the order of 30 minutes, it can be used to map isoprene emissions with spatial resolution from 10-100 kms. Horizontal transport smears the HCHO signal so that source location would need to be calculated using wind speeds and loss rates (Palmer et al. 2001; Palmer 2003). Smearing requires analysis and filtering due to the importance of transport and NO_x on forming robust and accurate estimates. Over Australia NO_x levels are generally not high enough to ensure quick HCHO formation and we must take care to account for resultant smearing. Details on smearing analysis and filtering are in Section 3.2.5.

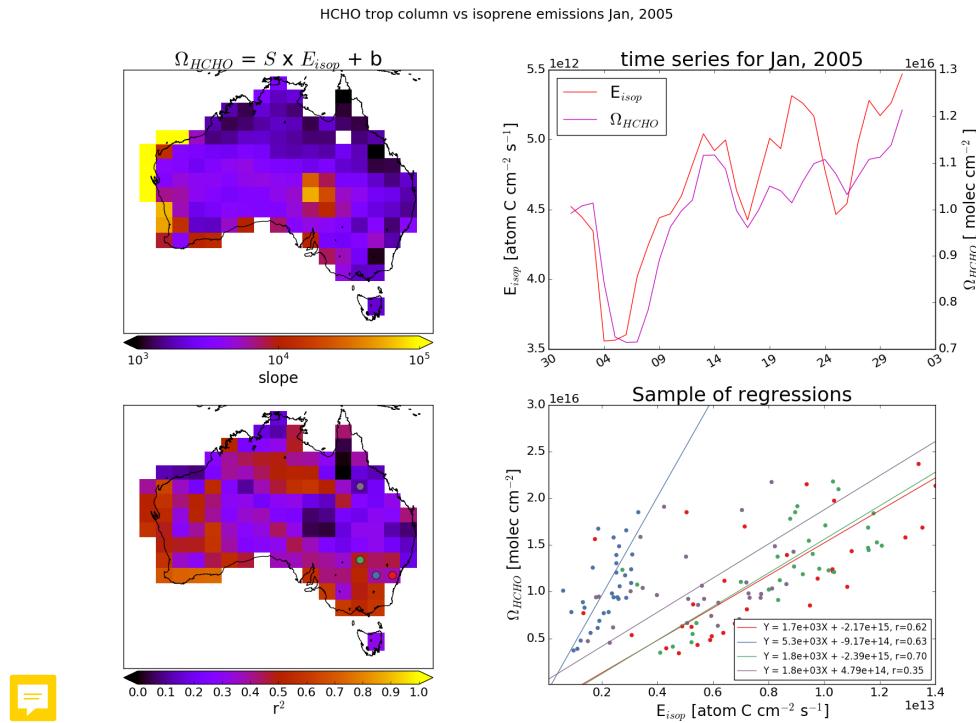


FIGURE 3.1: Top left: RMA slope between modelled tropospheric column HCHO (Ω_{GC}) and isoprene emissions (E_{GC}) using midday (13:00-14:00 LT) values over for January 2005, per gridsquare at $2 \times 2.5^\circ$ horizontal resolution. Top right: Australia-wide average of midday emissions and tropospheric 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 gridsquares.

3.1.3 MEGAN emission model



The Model of Emissions of Gases and Aerosols from Nature (MEGAN) is one of the most popular emissions inventories for biogenic isoprene. Global atmospheric studies often use MEGAN along with a chemical transport model (CTM) to examine transport, emission, deposition, and other chemical processes in the atmosphere. Emissions of Biogenic 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.

MEGAN is poorly calibrated for Australian conditions. Emissions of isoprene (C_5H_8) may be overestimated in some regions within Australia. 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). This could lead to the unique scenario of neither emission type dominating VOC chemistry over the forests.



3.1.4 satellite measurements



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 (Thomas et al. 1998), satellites have been used to provided a total column measurement of HCHO, allowing isoprene emissions estimates by proxy (Palmer et al. 2001; Bauwens et al. 2016). Using satellite information to improve biogenic emissions is pinpointed as a valuable use of satellite derived datasets (Streets2013). In this work we use NASA's OMHCCHO product, using measurements from OMI on board the AURA satellite (see section 2.3).

Satellites recording reflected solar spectra use DOAS to measure various trace gases in the atmosphere, including formaldehyde. While satellite measurements can only be used during daytime hours, HCHO lifetimes are sufficiently short that any night-time chemistry will not affect midday observations (Wolfe et al. 2016). Satellite records are often compared with modelled estimates of HCHO and also used as a proxy to estimate isoprene emissions. Total HCHO is measured by satellite over the entire world. The technique used to determine precursor emissions suffers from uncertainties, not the least of which are the uncertainties in the satellite measurements themselves. Zhu et al. (2016) use SEAC⁴RS aircraft HCHO measurements over the southeastern US as model validation. They show that a bias in the assumed OMI

shape factor leads to a bias between satellite and SEAC⁴RS measurements. Satellite based chemical concentrations often require both remote and in-situ measurements combined with modelled data for validation (Marais et al. 2014). There is less information available from satellite measurements at higher latitudes due to increased error in measurements over the more slanted column paths (De Smedt et al. 2015). Validation is important due to the various uncertainties in the satellite remote sensing process, with a priori assumptions having the greatest effect on structural uncertainty between measurements techniques Lorente et al. (2017).

3.2 Methods

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 (Ω ; saved in OMHCHORP dataset) are calculated 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° by 0.3125° horizontal resolution (see section 2.6).
2. Level three satellite data are used to make anthropogenic, fire, and smoke influence masks (see section 2.7). These are applied to remove Ω which may be influenced by non-biogenic sources.
 - (a) The fire mask is created daily from non-zero (MODIS) fire counts over the prior 2 days which occur in local or adjacent grid squares at 0.25° latitude by 0.3125° longitude.
 - (b) Influence from transported smoke plumes is removed by filtering OMI aerosol absorption optical depth (AAOD, from OMAERUVd) greater than 0.03.
 - (c) A filter for anthropogenic influence is created daily using OMNO2d NO₂ tropospheric column amounts; masking any grid squares with greater than 2×10^{15} molec cm⁻² on any particular day, along with grid squares where the yearly average is above 1.5×10^{15} molec cm⁻².
3. GEOS-Chem modelled biogenic emissions of isoprene (E_{GC}) along with total biogenic columns of HCHO (Ω_{GC}), both averaged over $2 \times 2.5^\circ$ horizontally and 1300-1400 LT temporally, are used to calculate a linear regression slope ($\Omega_{GC} = SE_{GC} + \Omega_{GC,0}$)
 - (a) Hourly gridded model output E_{GC} atoms C cm⁻² s⁻¹ is read, and 13:00 LT daily values are extracted.
 - (b) Daily 13:00-14:00 LT Ω_{GC} molec cm⁻² output is read.
 - (c) A reduced major axis regression slope is determined between Ω_{GC} and E_{GC} using a month of modelled output (one value per day) for each gridsquare (eg. see figure 3.3)

4. Ω are used to create a top-down estimate of biogenic isoprene emissions (E_{OMI} atoms $\text{C cm}^{-2} \text{ s}^{-1}$)
 - (a) Background total columns (Ω_0) are calculated by averaging longitudinally the Ω between 180 and 120°W
 - (b) Emissions are calculated using

$$E_{OMI} = \frac{\Omega - \Omega_0}{S} \quad (3.2)$$
5. 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 $\hat{S} = \Delta\Omega_{GC}/\Delta E_{GC}$: 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 3.2.5
6. Top-down emissions E_{OMI} are compared against MEGAN estimated emissions, as well as analysed in conjunction with concentration measurement campaigns (MUMBA, SPS, and one set of airplane measurements (HIPPO)). TODO: do these things
 - (a) Top-down isoprene emissions are calculated for the time window 1300-1400 LT, matching the peak of the diurnal cycle
 - (b) A monthly scaling factor (α) for isoprene emissions is created based on the difference between peak MEGAN emissions and our top-down estimates
 - (c) GEOS-Chem is run using the a posteriori emissions, and HCHO, O_3 , isoprene, and NO_x outputs are compared to campaign measurements

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 $0.25 \times 0.3125^\circ$ horizontal bins. This forms the intermediate product OMHCHORP, which is fully described in section 2.6.1. This dataset includes gridded satellite HCHO columns (Ω), along with pixel counts (how many satellite datapoints were used for each gridbox) to allow averaging and resolution changes.

In order to determine biogenic HCHO enhancements from Ω , we require filters for non-biogenic sources. Anthropogenic, fire, and smoke influence masks are created from three satellite products: NO₂ from OMNO2d, fire counts from MOD14A1, and AAOD from OMAERUVd respectively (see section 2.7). These masks are binned using daily averaged or summed values within our $0.25 \times 0.3125^\circ$ horizontal resolution. The recalculated corrected vertical columns are saved to OMHCHORP dataset both before and after applying the filters, to allow filter analysis. While one primary source



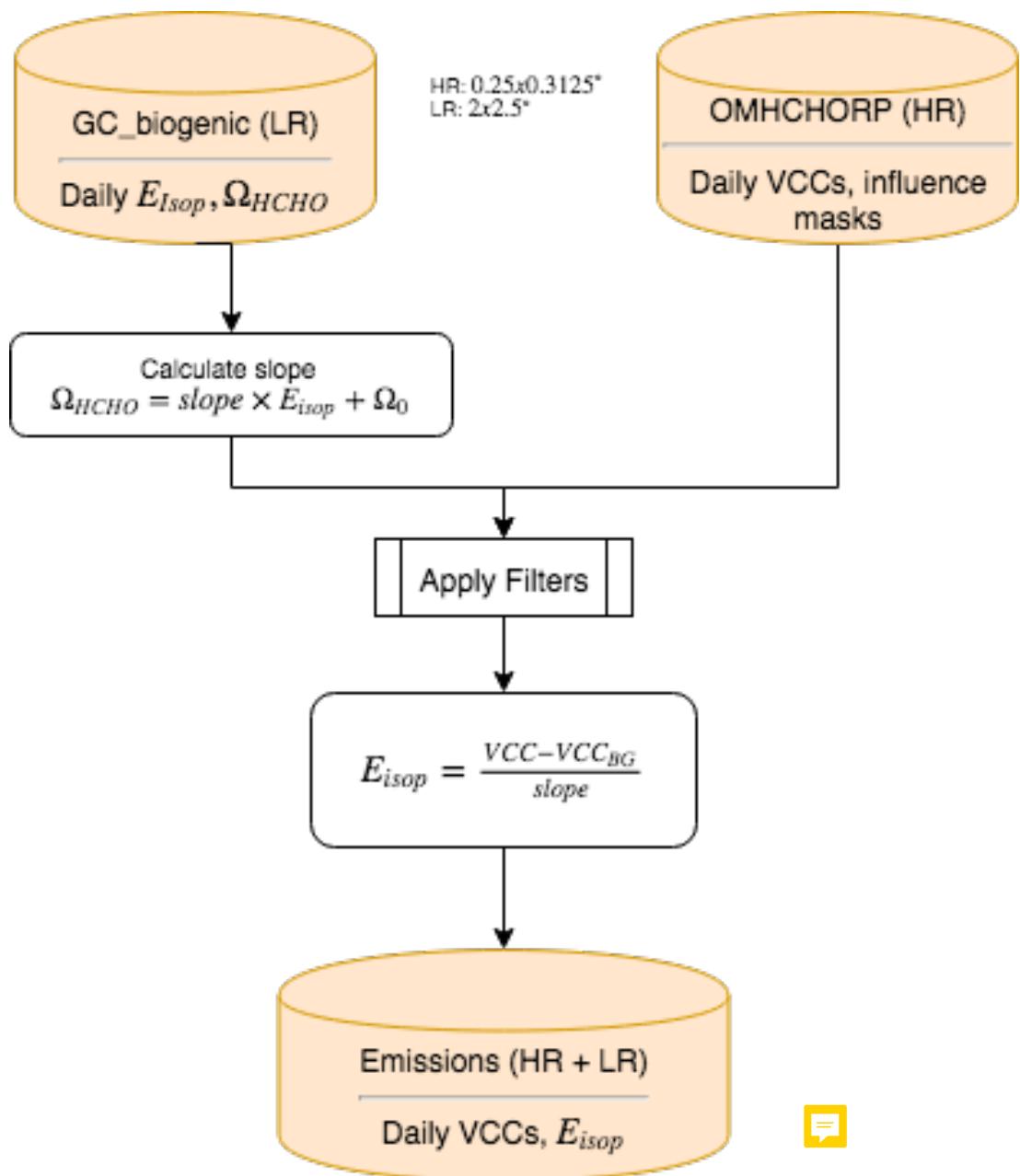
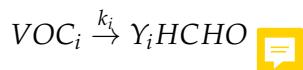


FIGURE 3.2: Creation of isoprene emissions dataset using OMH-CHORP and biogenic GEOS-Chem outputs.

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.

3.2.3 Calculating modelled slope

By assuming a simple linear steady-state relationship between HCHO and its precursors (Palmer 2003; Palmer et al. 2006; Millet et al. 2006), one may calculate isoprene emissions using measured HCHO. 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. Assuming HCHO is produced quickly from short-lived intermediates:



Where i indexes a chemical species, Y_i is HCHO yield per C atom, and k_i is the reaction rate. Then assuming a steady state of atmospheric HCHO (Ω) in molec cm^{-2} produced by oxidation of VOCs (VOC_i) and no horizontal transport:

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

Where k_{HCHO} is the HCHO loss rate due to OH and photolysis, Y_i is the molar HCHO yield from oxidation (of i) and E_i is emission fluxes (atoms $\text{cm}^{-2}\text{s}^{-1}$). We assume isoprene is the most important precursor and lump other terms into the background Ω_0 so that our equation becomes

$$\begin{aligned} \Omega_{\text{HCHO}} &= \frac{Y_{\text{isop}}}{k_{\text{HCHO}}} \times E_{\text{isop}} + \Omega_0 \\ &= S \times E_{\text{isop}} + \Omega_0 \end{aligned} \quad (3.3)$$

The equations above come from several assumptions that are important to understand. The first assumption is that HCHO is at steady state, which implies productions and losses are equivalent. Then we assume that loss is only due to photolysis and oxidation (first order, $k_{\text{HCHO}} = 1/\tau_{\text{HCHO}}$). Production of HCHO above a background is due to the sum of the precursor emissions (E_i) multiplied by their yields to HCHO (Y_i).

$$\begin{aligned} \frac{d[\text{HCHO}]}{dt} &= 0 = \text{Prod}(\text{HCHO}) - \text{Loss}(\text{HCHO}) \\ \text{Loss}(\text{HCHO}) &= k_{\text{HCHO}}[\text{HCHO}] \\ \text{Prod}(\text{HCHO}) &= \Sigma_i Y_i E_i \end{aligned}$$

Now instead of concentrations we use total column amounts for HCHO, with emissions being in units of molecules $\text{cm}^{-2} \text{ s}^{-1}$. And then lumping together non-isoprene

precursors, and assuming that enhancements above the background are only due to isoprene emissions we have:

$$\frac{k_{HCHO} \Omega_{HCHO}}{\Omega_{HCHO}} = Y_{isop} E_{OMI} + \sum_{i \neq isop} Y_i VOC_i \\ = S E_{OMI} + \Omega_0$$

Estimates of E_{OMI} , or S can be attained through modelling (e.g. Millet et al. 2006).

We look at each $2 \times 2.5^\circ$ grid box from daily GEOS-Chem (biogenic only) output of $\Omega_{HCHO} \equiv \Omega_{GC}$ and $E_{OMI} \equiv E_{GC}$ within Australia, and calculate the (RMA) correlation for each month. Modelled background concentrations can be ignored here as they do not affect slope calculation. This effectively gives us monthly gridded isoprene to HCHO slope (S) in units of molecules HCHO / atom C \times seconds (seconds). Figure 3.3 (top left) shows the RMA regression slope between modelled HCHO columns and isoprene emissions calculated within each grid square over January 2005, averaged between 1300-1400 LT each day. Also shown are the regression coefficients (bottom left), Australia-averaged E_{GC} and Ω_{GC} time series (top right), and a sample of the regressions (bottom right). Each regression is for a single grid box, over the course of one month, and is coloured by location (matching dots shown in the bottom left panel). The slopes shown in the bottom right panel can range widely due to differences in emission and yield parameters, which plays a role in the smearing filters applied in section 3.2.5. Due to the low horizontal resolution of GEOS-Chem (2.5° latitude by longitude), 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 3.3.

There are a couple of ways to determine the modelled background HCHO concentration, one of which involves running the model with isoprene emissions turned off, which allows us to see exactly how much the modelled isoprene emissions alter each vertical column of 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 involves looking at 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. 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.

3.2.4 Calculation of Emissions

As is done in Palmer (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. Emissions of precursors are easy to calculate using the slope S calculated in the prior section from modelled HCHO and isoprene columns:

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

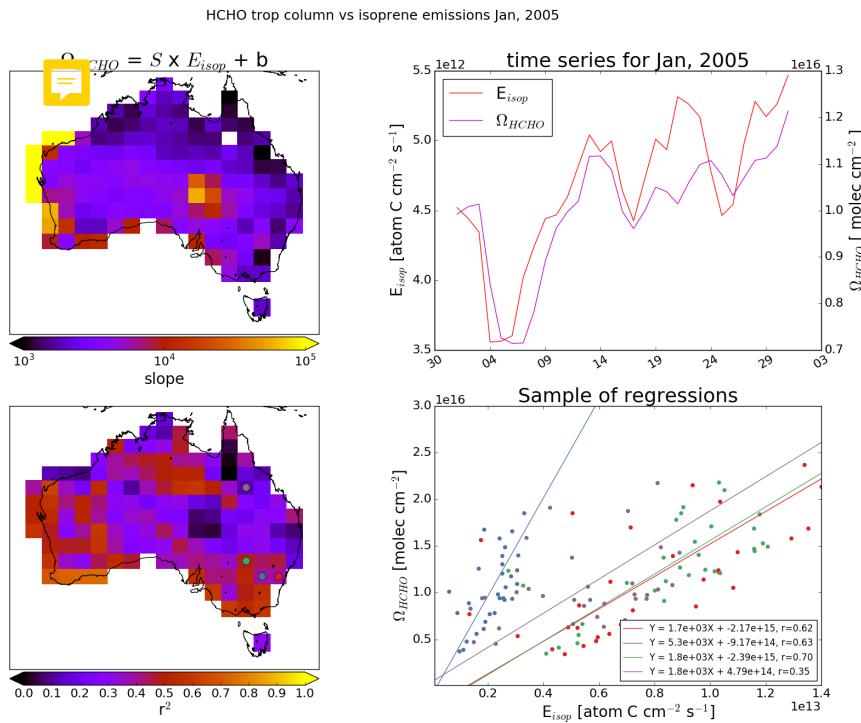


FIGURE 3.3: Top left: the RMA regression slope between modelled HCHO columns and isoprene emissions calculated within each grid box over January 2005. Bottom left: the square of the regression coefficients. Top right: Australia averaged time series. Bottom right: a sample of the regressions coloured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

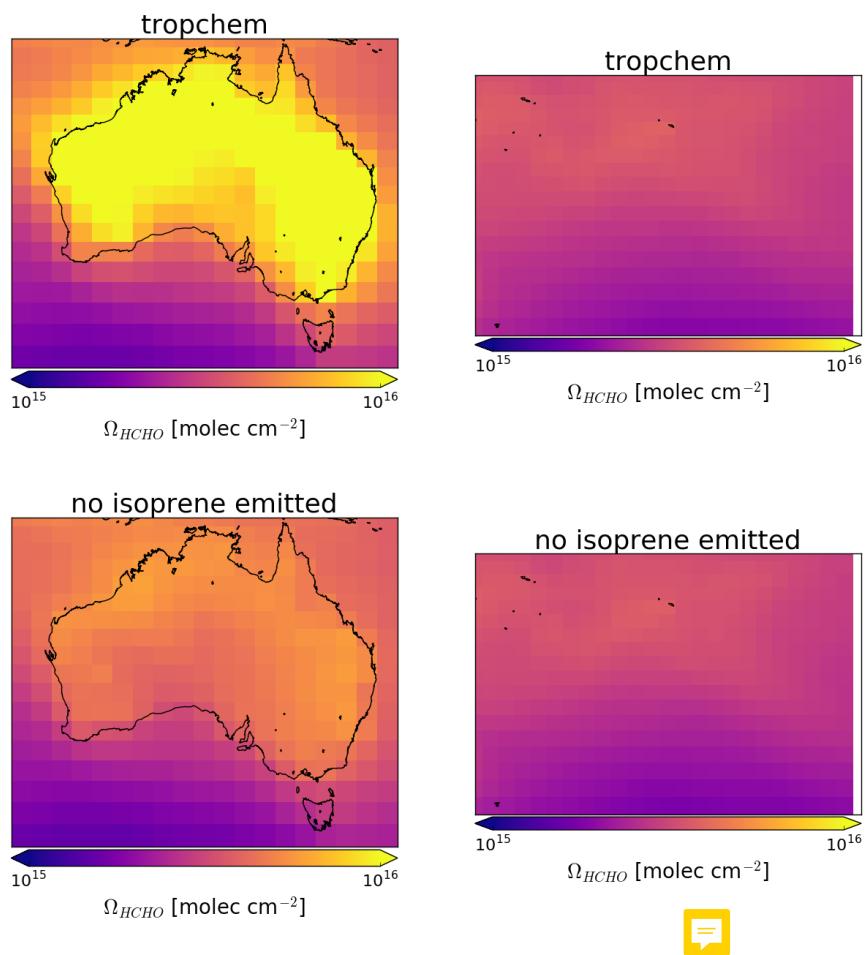


FIGURE 3.4: Total column HCHO over Australia (left) and the remote pacific region (right) using GEOS-Chem with (top) and without (bottom) isoprene emissions.

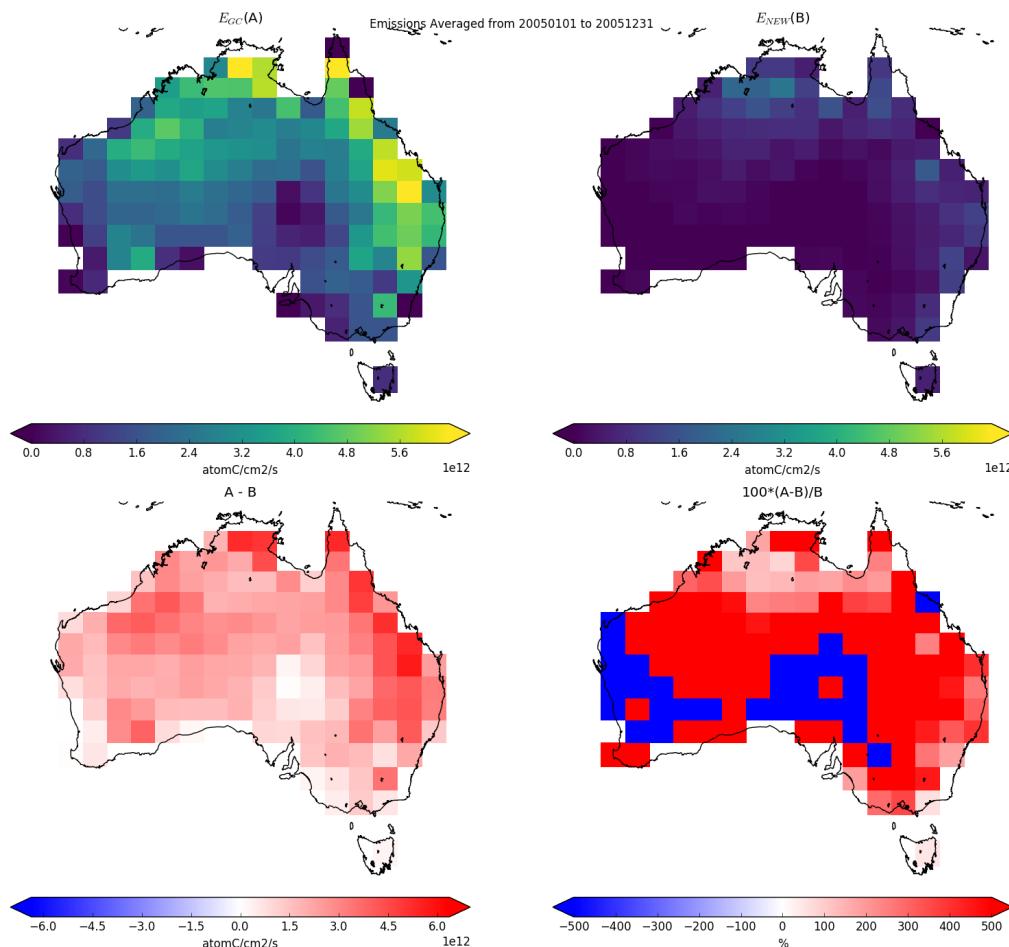


FIGURE 3.5: Top row is isoprene emissions for the month of January, in 2005, from GEOS-Chem and estimated from OMI respectively. Bottom row shows the absolute and relative differences between the two.

This is the same equation 3.3, except now we use the modelled slope along with satellite HCHO (Ω , and Ω_0). The background HCHO is calculated using measurements in the remote **pacific** at the same time and latitude as Ω . 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 the emissions calculated this way along with the modelled emissions output by GEOS-Chem E_{GC} averaged over January, 2005.

The **Background** (Ω_0) from OMI is determined using the mean column HCHO measured over the remote pacific ocean. For this term we use remote ocean measurements averaged monthly and longitudinally. This gives us a background which is appropriate for any latitude. Figure TODO: figure with background region highlighted and a time series of background values. One potential issue is the low number of valid measurements at **high latitudes** (especially during winter). When calculating the E_{OMI} from our modelled slope, negative emissions result when using OMI measured columns are lower than the background amounts (as $E_{OMI} = \frac{\Omega - \Omega_0}{S}$). These are

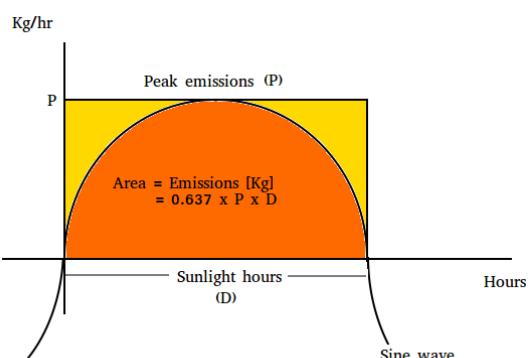


FIGURE 3.6: How daily isoprene emissions (in kg) are calculated.

set to zero, which increases the average TODO: ~X %.

Top-down emission rates calculated in this work are in units of molecules $\text{cm}^{-2} \text{s}^{-1}$. In order to calculate the emissions in kg, each gridsquare is multiplied by the area, and then daily emissions are assumed to follow a sine wave peaking at the estimated rate Figure 3.6 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 hrs (Jul).

3.2.5 Accounting for smearing

Accounting for transport of the precursors is important, especially in low NO_x conditions in which isoprene has a longer lifetime (hours-days). ~~When estimating emissions of isoprene using one of its products, it is often assumed that isoprene has a short lifetime.~~ This is a problem in Australia where low NO_x environments abound, as detected formaldehyde may not be directly above its emission source (as assumed in this top-down estimation technique). An analysis of spatial smearing (smearing from here forwards) can be used to mitigate potential emissions estimation biases by filtering out affected gridsquares. Smearing in this case is a measure of how much formaldehyde is created from isoprene emissions in non-local grid boxes Smearing has been estimated in various works (eg. 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 González Abad 2018), often implementing the method designed in Palmer (2003). This method involves calculating smearing 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 (eg. Stavrakou et al. 2009) involves the analysis of an adjoint CTM, however this is computationally expensive

In this work a run of GEOS-Chem using globally halved isoprene emissions (with no other changes) is performed to create a smearing filter. Consider halving the isoprene emitted globally and rerunning the model, one would expect HCHO enhancement (above background levels) to be halved in isoprene-emitting gridboxes if no transport has occurred. This idea is behind the method of testing the correlation between HCHO enhancement and isoprene emissions. By assuming no transport and

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) ^a	1300	1800	Smearing limits for South America
Surl, Palmer, and González Abad (2018)	2200	4900	Slope ranges seen in India

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

negligible yield and lifetime changes between model runs, an equation can be derived and tested, finding where the assumptions lead to unlikely yields.

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))^T; one can set a simple bound on \hat{S} of $[0.2 \times 1.5, 0.4 \times 4]$ hrs or 1080 to 5760 seconds. As NO_x levels across Australia are relatively low, 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 5200 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 NO_x concentrations. TODO: Figure todo shows \hat{S} over Australia for one year, along with where and when the filter has most impact. Table 3.1 shows the smearing filters or typical slopes seen in other works.

3.2.5.1 Calculation of smearing

In order to understand the smearing calculation the underlying equations and assumptions must first be understood. From section 3.2.3, equation 3.3 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:

$$\begin{aligned} Run_1 : \Omega_{HCHO} &= SE_{isop} + \Omega_0 \\ Run_2 : \Omega'_{HCHO} &= S'E'_{isop} + \Omega'_0 \end{aligned} \tag{3.4}$$

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^A; this is the basis for equations 3.4. It is assumed that background values (Ω_0) are from oxidation of methane and other long lived VOCs, so that $\Omega_0 = \Omega'_0$. Between these two runs we are only changing the E term, we do not change any chemistry and so we can expect that the

yield and loss rate is not changing between the two runs $S = \frac{Y_{isop}}{k_{HCHO}}$ which leads to us being able to combine the runs in equation 3.4 as follows:

$$Run_1 - Run_2 : \Omega_{HCHO} - \Omega'_{HCHO} = SE_{isop} - S'E'_{isop} + \Omega_0 - \Omega'_0 \quad (3.5)$$

$$\Delta\Omega_{HCHO} = S\Delta E_{isop} \quad (3.6)$$

$$\hat{S} \equiv \frac{\Delta\Omega_{HCHO}}{\Delta E_{isop}} \approx \frac{Y_{isop}}{k_{HCHO}} \quad (3.7)$$

And using the output from our two runs we can see if the calculated \hat{S} is wildly different from expected values for S .

Similarly to smearing sensitivity calculations in Marais et al. (2012), we run GEOS-Chem with isoprene emissions halved, then calculate $\hat{S} = \frac{\Delta\Omega_{HCHO}}{\Delta E_{isop}}$. Here Δ represents the departure (daily over 1300-1400 LT) from default run values. If \hat{S} is large, then you can infer sensitivity to non-local isoprene emissions. A relatively large change in Ω_{HCHO} compared to local emissions suggests that HCHO is being formed from non-local isoprene emissions. Alternatively, a relatively low value of \hat{S} infers that emissions from a particular grid square are being exported before they form HCHO, which also informs us that local HCHO levels are not due to local emissions.

Smearing is sensitive to how E_{isop} is determined, figure 3.7 shows smearing over two seasons defining E_{isop} as the daily averaged (left column) and midday (1300-1400 LT) isoprene emissions (right column). Essentially the midday isoprene emissions are at the peak of their daily cycle (shown later in figure 3.10) which means the effect of smearing is relatively smaller during these hours. Figure TODO: shows averaged isoprene emissions with added markers showing where the threshold of 800-5200 affects at least one day within the season (cyan or pink diamonds) and where it removes all data for that gridbox (blue or red x).

3.2.5.2 Sensitivity to smearing

Smearing can be dependent on local or regional weather patterns, as greater wind speeds will reduce the time any emitted compound stays within the local grid box. As such smearing sensitivity is both spatially and temporally diverse. Figure TODO: is a picture of the smearing sensitivity over Australia in Summer and Winter, along with where the smearing threshold is exceeded (for one or all days). Large smearing values can be seen near many coastlines as only a fraction of the grid squares actually emit isoprene, which makes transported isoprene relatively more important in these gridboxes. Once the smearing-sensitive grid squares are filtered out, application of equation 3.3 is used to estimate isoprene emissions across the nation.

When limiting smearing (\hat{S}) to within 800-4600 s, GEOS-Chem correlation between isoprene emissions and HCHO columns should improve (TODO analysis). A problem arises due to the loss of datapoints used to create monthly gridsquare sessions, and a secondary filter is applied where the confidence interval of the slope exceeds 200%. Figure TODO shows GEOS-Chem midday HCHO columns compared against GEOS-Chem emissions of isoprene, over land squares with (red) and without (grey) filtering for smearing. This smearing range captures isoprene to HCHO yields of around 0.16 to 0.32 C per C if HCHO lifetime is assumed to lie within 1.5 to 4 hours.

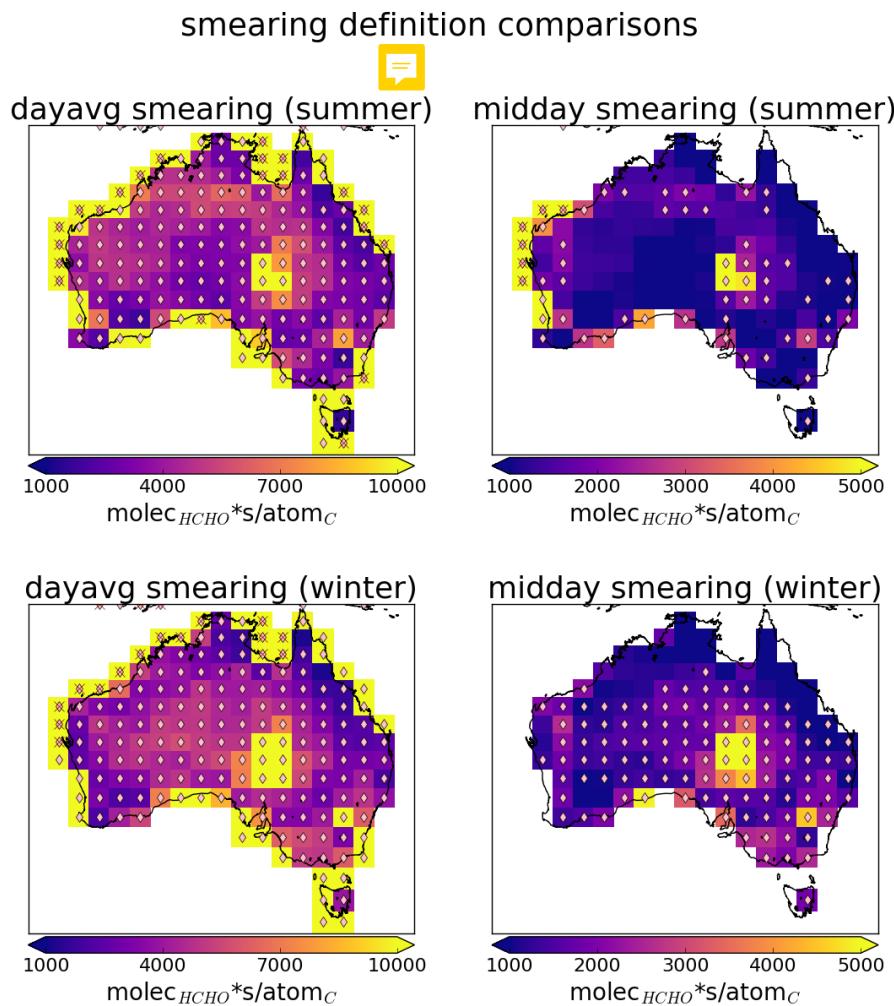


FIGURE 3.7: Smearing (\hat{S} , see text) in summer (DJF, top row) and winter (JJA, bottom row) using averaged isoprene emissions daily (left column) and 1300-1400 LT (right column). The scale changes between left and right columns.

3.2.5.3 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 ($L_{d,i}$) downwind (d) by a precursor (**i**) before becoming 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 a fraction ($1 - 1/e$) of the precursor is completely transformed into HCHO. This equation uses the initial VOC column concentration ($[VOC]_0$) at the point of emission and mass balance equations as follows:

$$\frac{1}{k_{HCHO} - k_i} \left(k_{HCHO} \exp \left[\frac{-k_i L_{s,i}}{U} \right] - \kappa_i \exp \left[\frac{-k_{HCHO} L_{s,i}}{U} \right] \right) = \frac{1}{e} \quad (3.8)$$

with limiting values $L_{s,i} \rightarrow U/k_i$ for $k_i \ll k_{HCHO}$, and $L_{s,i} \rightarrow U/k_{HCHO}$ for $k_{HCHO} \ll k_i$. Figure TODO: shows a rough estimate of isoprene smearing length (L_{isop}) in Australia using wind speeds from TODO:, and reaction rates k_{isop} , k_{HCHO} from GEOS-Chem.

GEOS-Chem daily averaged HCHO lifetime (τ) is shown for 2005 in figure 3.8. This lifetime is calculated using daily averaged surface loss rates and concentrations of HCHO:

$$\tau = \frac{[HCHO]}{\text{Loss}}$$

The expected lifetime of HCHO is determined by assuming loss is linear (first order) and dividing grid box daily averaged concentrations of GEOS-Chem HCHO ($[HCHO]$ in molecules cm^{-3}) by their modelled losses (Loss 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 (JJA) is shown in figure 3.8. Additionally lifetimes coloured by land grid squares (dots in top right panel) are shown over time in the bottom panel. The problem with this approximation is that we are not interested in daily averaged lifetime, but the midday (13:00-14:00 LT) lifetime. This figure highlights the seasonal nature of HCHO loss rates, although midday numbers are expected to have less seasonality. 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.

3.2.5.4 NO_x dependence

Isoprene production of HCHO depends on several factors, importantly NO_x levels directly affect the fate of VOCs in the atmosphere. Isoprene first reacts with the hydroxy radical, producing an organic peroxy radical (RO₂). At higher NO mixing ratios (at least a few hundred pptv), RO₂ react mostly with NO. At low NO (less than 50 pptv), RO₂ is more likely to either isomerise, or react with HO₂, or another RO₂. In low NO_x environments, reported HCHO yields from isoprene are around 0.2 - 0.3 C per C, while in high NO_x environments this value becomes two to three times higher (Palmer 2003; Wolfe et al. 2016).

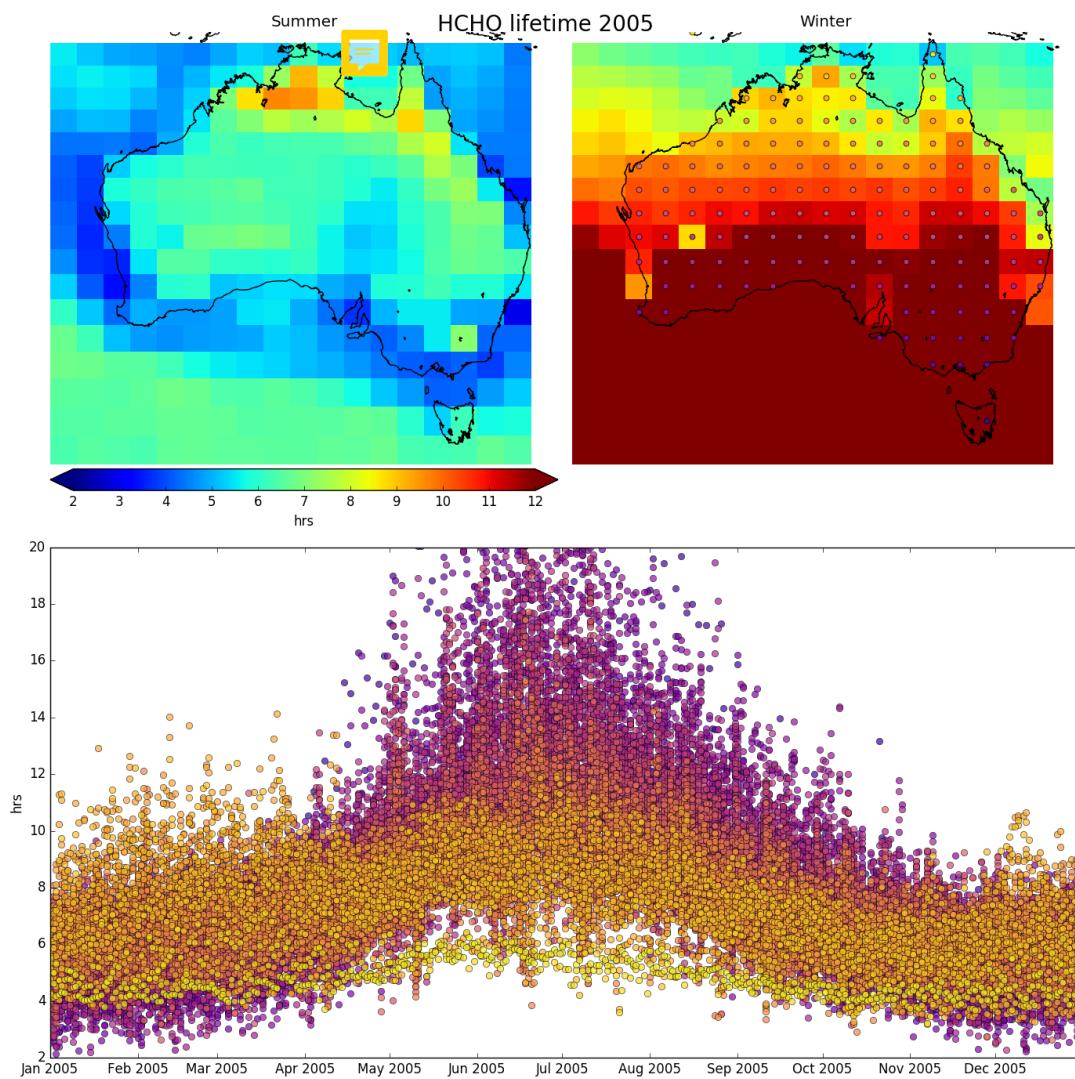


FIGURE 3.8: Top left, right: Summer (Jan., Feb.) and winter (JJA) averaged daily surface HCHO lifetime (τ). Bottom panel: τ over the year, coloured by grid square (see dots in top right panel).

NO_2 measured by OMNO2d gives us a daily mid-day measurement which we can compare to output from GEOS-Chem to determine how well the model does at simulating NO_2 (see section 2.4.6). The effect of NO_2 on smearing can be seen in figure 3.9. This plot shows how smearing over Australia compares to satellite NO_2 levels, with smearing distributions binned by NO_2 both with and without applying a filter for smearing. One feature of the figure is that at lower NO_2 levels the smearing is often 2-4 orders of magnitude above the upper threshold. This abruptly decreases at around 5×10^{14} molec cm^{-2} NO_2 . There is also a higher number of data points below the lower threshold before that same NO_2 level, suggesting that transport is a bigger issue at $\text{NO}_2 < 5 \times 10^{14}$ molec cm^{-2} .

The half-life of HCHO to photo oxidation with hydroxyl radicals is around 1 hr depending on environmental conditions (WHO_hcho_guidelines_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 NO_x levels) 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.

Conversions between HCHO per unit C yield and molar % yield from species X are given by the equation $Y_{\text{molar}\%} = 100 \times C_X \times Y_{\text{HCHO C}^{-1}}$, where C_X is how many carbon are within species X (5 for isoprene). For instance a 200% molar yield of HCHO from isoprene implies 1 mole of C_5H_8 becomes 2 mole HCHO which is a 0.4 HCHO per unit C yield.

3.2.6 Running GEOS-Chem using a posteriori emissions

After creating the top-down estimation of isoprene emissions, we run GEOS-Chem again with emissions scaled to match the new estimate. This is done through taking all the new midday (13:00-14:00 LT) emissions (per grid box) and forming a multi year monthly mean, which can be compared to the MEGAN equivalent. A monthly factor (α at $2 \times 2.5^\circ$) that scales MEGAN to match the top-down emissions is then applied within HEMCO (the emissions module in GEOS-Chem). Fig 3.10 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. First hourly biogenic isoprene emissions are retrieved from GEOS-Chem (estimated using the MEGAN model). Then the midday emissions for each month per gridbox are averaged (see figure TODO) and the multi-year average of these is compared against the top-down estimate.

A scaling factor α is derived which when multiplied with E_{GC} produces the top down emissions estimate $E_{\text{OMI}} : \alpha = \frac{E_{\text{GC}}}{E_{\text{OMI}}}$. This is performed through a small modification of GEOS-chem source code which applies α after calculating emissions for each grid-square based on the default emission factors and meteorology. This scale factor is set to one wherever top-down emissions are not calculated.



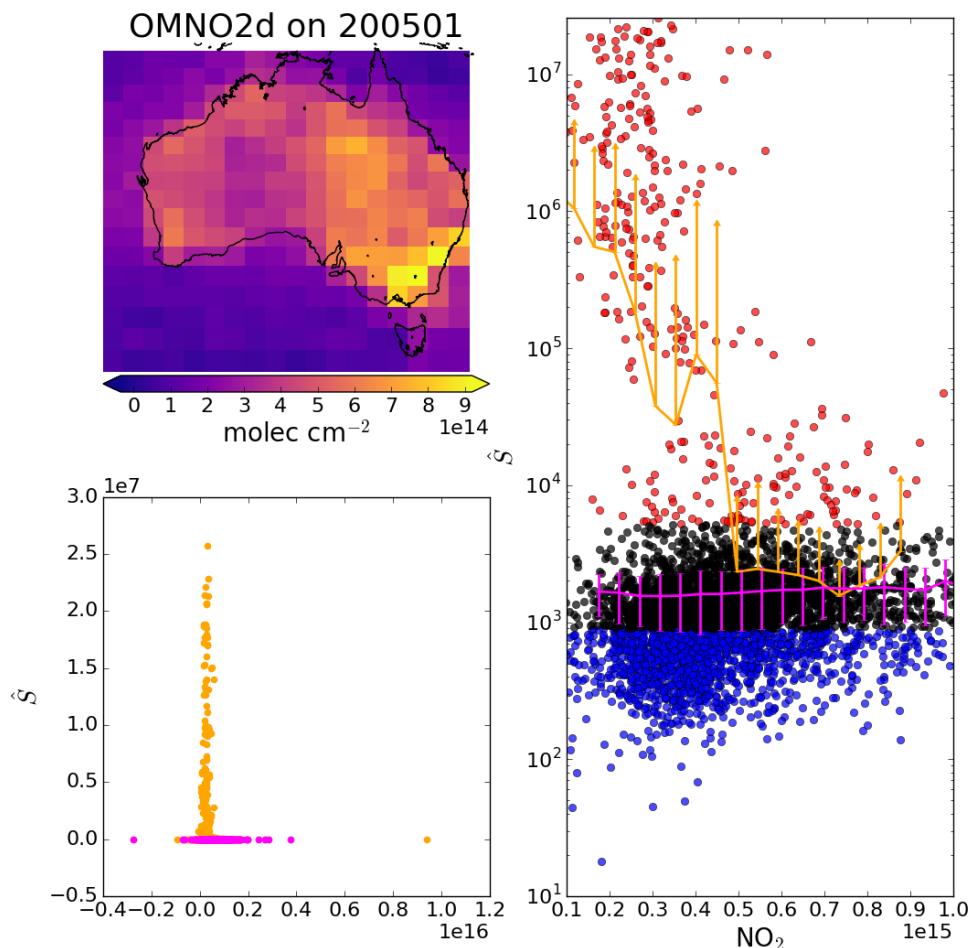


FIGURE 3.9: Top left: OMNO2d tropospheric NO₂ columns (NO₂: molec cm⁻²) averaged into 2²⁵ horizontal bins for over Jan, 2005. Right: Scatter plot of NO₂ against smearing calculations from GEOS-Chem (\hat{S}), with points above and below the smearing threshold range of 900-5200 s coloured red and blue respectively. Points are binned by NO₂ with and without having the smearing filter applied (orange and magenta respectively). Overplotted is the mean and standard deviation (error bars) within each bin. Due to the logarithmic Y scale we only show the positive direction of standard deviations for unfiltered data. Bottom left: Daily NO₂ scattered against smearing with (magenta) and without (orange) applying the smearing filter. This plot is a zoomed out version of the right panel.



TABLE 3.2: HCHO yields from isoprene, and lifetime against oxidation by OH.

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

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

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

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

d Wolfe et al. (2016): “prompt yield”: change in HCHO per change in ISOP₀. $[ISOP]_0 = [ISOP] \exp(k_1[OH]t)$; where k_1 is first order loss rate. Effectively relates HCHO abundance with isoprene emission strength.

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

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

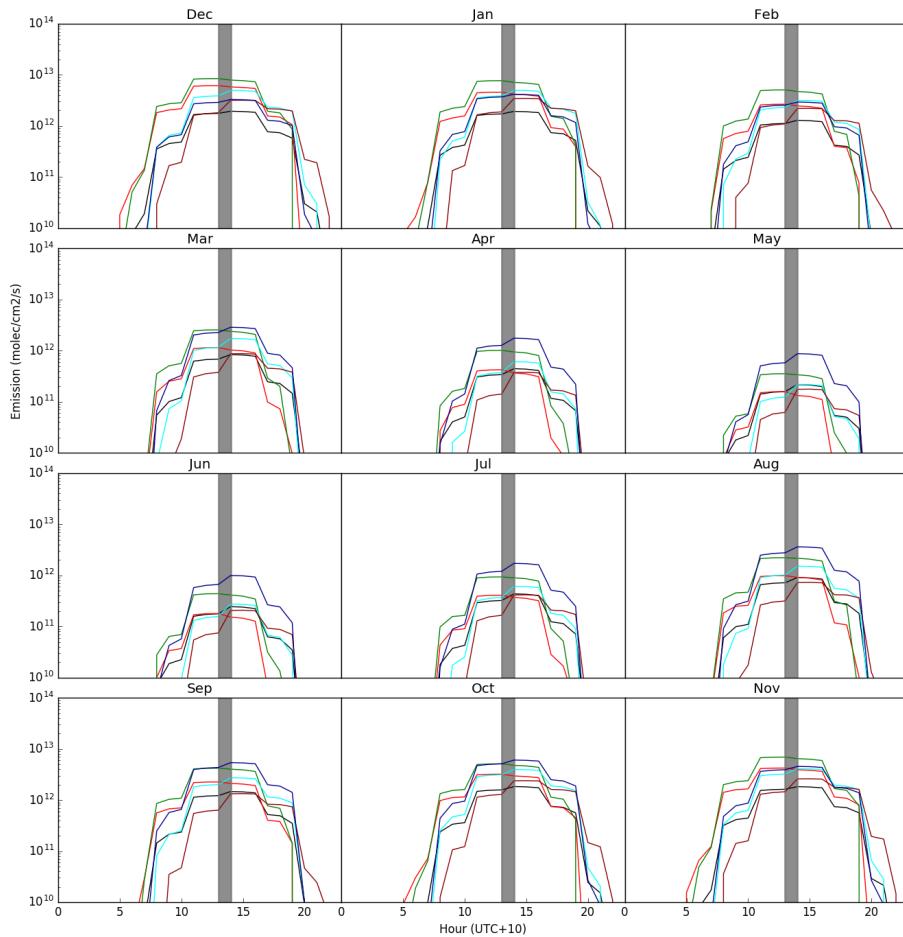


FIGURE 3.10: The diurnal cycle of MEGAN emissions averaged by month over 1, Jan, 2005 to TODO, 2013 are shown with lines, while top-down emissions estimates are shown with plus symbols. MEGAN emissions are estimated hourly per $2 \times 2.5^\circ$ horizontal grid box. Shown here are the averages within several areas (denoted by colour, see figure todo). Top-down estimates are similarly grouped by colour, and shown at the 13:00 LT mark for each month. Rows 1-4 match seasons from austral summer (DFT) through to spring (SFT).

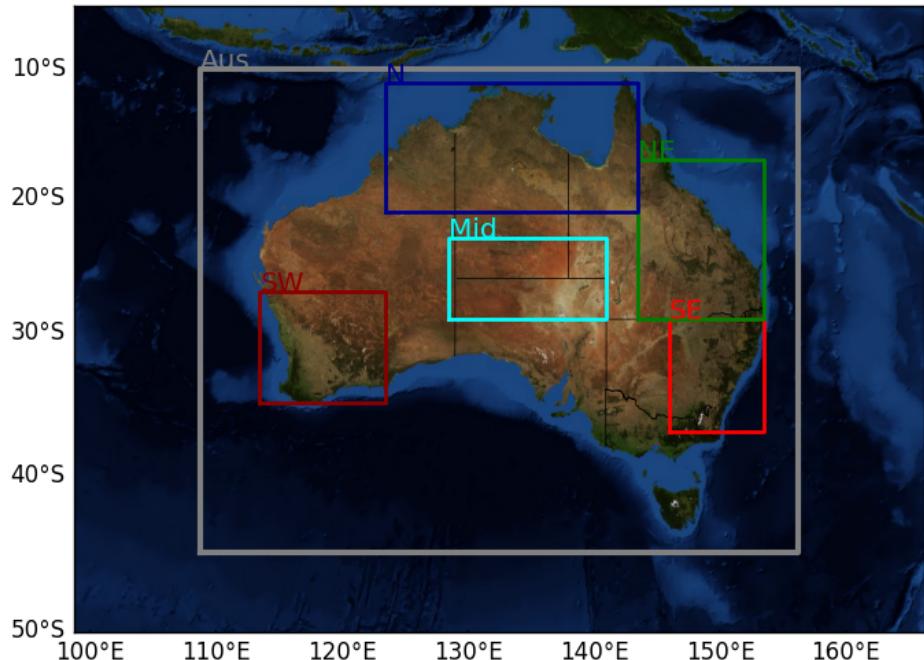


FIGURE 3.11: Sub-regions used in subsequent figures. Averages taken ~~within~~ Australia will be black or grey, while averages from within the coloured rectangles will match the colour shown here.



3.3 Results

Australia is a large country, roughly 7.7 million km², 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 3.11. Top-down emissions estimates (E_{OMI}) shown in this section 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).



3.3.1 HCHO Products and yield

Isoprene reaction chains are diverse, with many branches forming HCHO. HCHO production yields are often classed into ~~at least~~ two categories: ~~first~~ generation HCHO yield refers to the amount of HCHO produced per unit isoprene consumed by initial oxidation, while total (or molar) yield refers to time-dependent yield of HCHO over multiple oxidation stages (Wolfe et al. 2016). Wolfe et al. (2016) define prompt yield as the change in formaldehyde per unit change in initial isoprene emissions. In this work yield (Y_{isop}) is approximately the total yield within 4 hours.



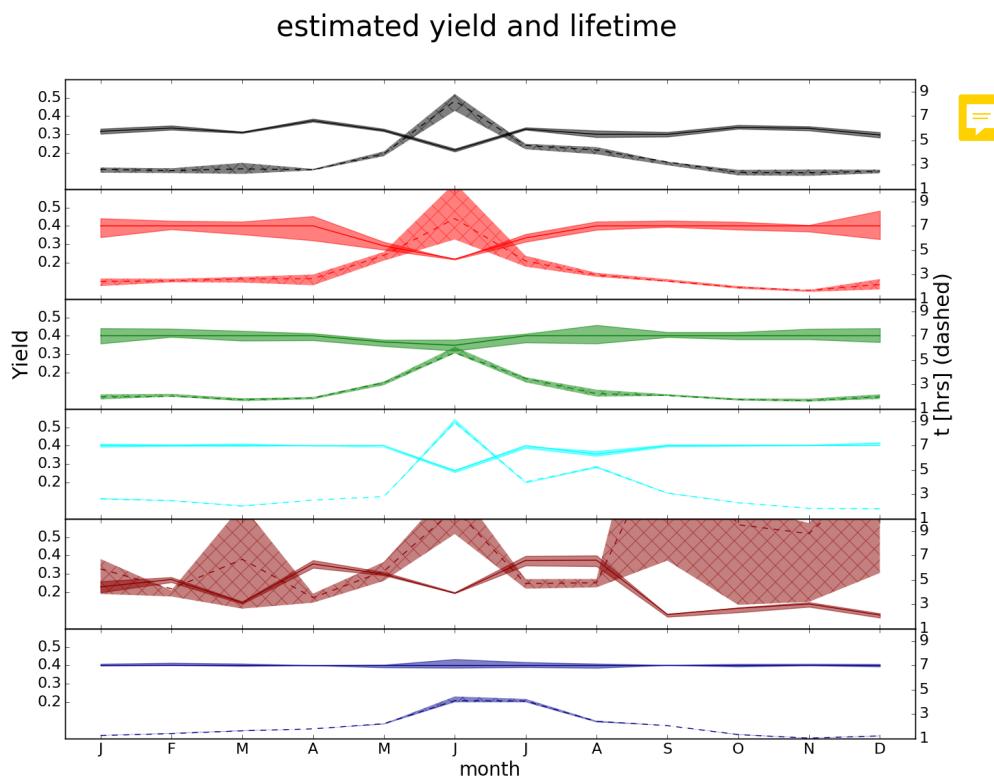


FIGURE 3.12: Monthly area averaged isoprene to HCHO midday yield and HCHO lifetime. Coloured by regions shown in figure 3.11. Shaded areas show the yield (plain) and lifetime (hatched) IQR.

Figure 3.12: shows the GEOS-Chem Y_{isop} and HCHO lifetime (τ) estimated throughout the year. By using an assumed constant Y_{isop} of 0.4, we estimate the midday lifetimes of HCHO using $S = \frac{Y_{isop}}{k_{HCHO}}$ from equation 3.3. Then dividing the slope by this monthly mean lifetime returns an estimate of the Yield. Both of these terms are heavily influenced by the assumed yield, and should not be taken as results, however this technique shows the seasonal cycle and spread of the HCHO yield and lifetime. A clear June (and sometimes March, July and August) increase in HCHO lifetimes is shown, with a matching drop in yield. These are the winter months, when midday temperature and insolation is reduced. Noise in the southwest region may be indicative of heavy filtering, potentially driven by westerly winds which can lead to both smearing and transported pollution.

3.3.2 Emissions comparisons

Guenther et al. (2012) Estimate global biogenic isoprene emissions at roughly 535 Tg yr^{-1} , using MEGAN. Sindelarova et al. (2014) Estimate around 594 Tg yr^{-1} using MEGAN with MACC, showing isoprene as 69.2% of the total BVOC emissions, with monoterpenes at 10.9 Tg yr^{-1} (10.9%). They show 41 Tg yr^{-1} decrease in Australia when introducing soil moisture parameterisation. When comparing the GEOS-Chem (which

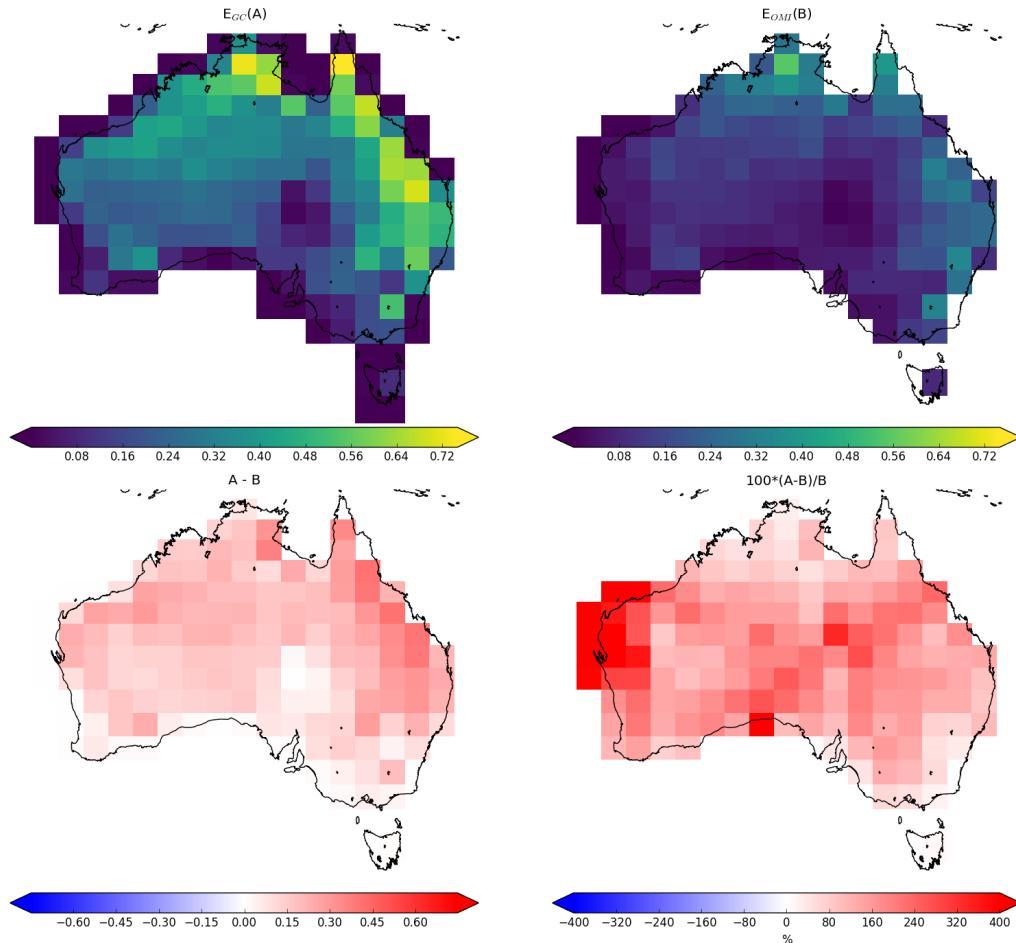


FIGURE 3.13: Top row: multiyear in emissions in $Tg\ yr^{-1}$ from E_{GC} (GEOS-Chem; running MEGAN) and E_{OMI} (top-down emissions) respectively. E_{OMI} uses an assumed sinusoidal daily cycle, with daylight hours prescribed for each month: see section 3.2.4 . Bottom left and right shows the absolute and relative differences respectively.

runs MEGAN) emissions to those calculated using our top-down inversion, we see a decrease of around 29 $Tg\ yr^{-1}$ (66%). Table 3.3 shows yearly isoprene emissions from this work and some other works for Australia and globally. Figure 3.13 shows how this decrease is distributed spatially, with E_{GC} and E_{OMI} in $Tg\ yr^{-1}$ calculated as a multi-year mean. Across all of Australia we see large reductions of total emissions using the new top-down estimate.

Figure 6 shows emissions over Australia calculated using the OMI top down estimate (column 1: E_{OMI}) and GEOS-Chem simulated emissions (column 2: E_{GC}). The first row shows the time series (daily midday averages), and the final row shows the absolute differences ($E_{GC} - E_{OMI}$). This figure is repeated using monthly means (of the daily midday estimates) in figure 3.17.

Figure 3.18 shows the multi-year monthly mean and IQR of daily midday isoprene emissions estimates, averaged over several regions (see figure 3.11). Generally months

TABLE 3.3: Isoprene emissions (Tg/yr)

Australia	Global	notes
43(2)	445(18)	a) GEOS-Chem: 2005-2010
19(2)		b) Top-down: 2005-2010
	535	Guenther et al. (2012):
	594	Sindelarova et al. (2014):
26-94	272-570	c) 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: Range shown here based on 3 different models and one top-down inversion

outside of May to August show the a posteriori lower than the a priori, except in the south eastern portion of Australia.

Figures 3.14 and 3.15 show how the distributions of top-down emissions compare to those of MEGAN in each region during summer months (DJF) with zeros and negatives removed from both distributions. Figure 3.14 shows the daily midday distributions, binned hexagonally to show data-point frequency. There is only weak correlation apparent between daily top-down and MEGAN estimations, however daily values suffer from large uncertainty. Figure 3.15 displays the regressions between monthly averages of the same data. In the monthly averages more 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.

3.3.3 Comparison with measurements

TODO: Analyse comparison of gridbox with campaigns of measurements

Comparison between ground-based measurements and large ($2 \times 2.5^\circ$) averaged grid squares suffers from heavy representational errors. Figure 3.19 shows the SPS and MUMBA measurement sites, along with the outline of the $2 \times 2.5^\circ$ model gridbox. The urban footprint of Sydney and Wollongong are clearly shown, along with some ocean, forest, and rural regions, which are all averaged within calculations made here. Due to high uncertainty in components of the top-down emissions estimate, temporal resolution is also limited. MUMBA, SPS1 and SPS2 each provide only a couple of comparable data points, and these campaigns measured isoprene concentrations while our estimate is of emissions.

Figure TODO shows how isoprene compares against measurements from the MUMBA and SPS campaigns (section 2.2.3.1, 2.2.3.2) both before and after scaling isoprene emissions to match the top-down estimation. TODO: discuss results and differences, does isoprene improve?

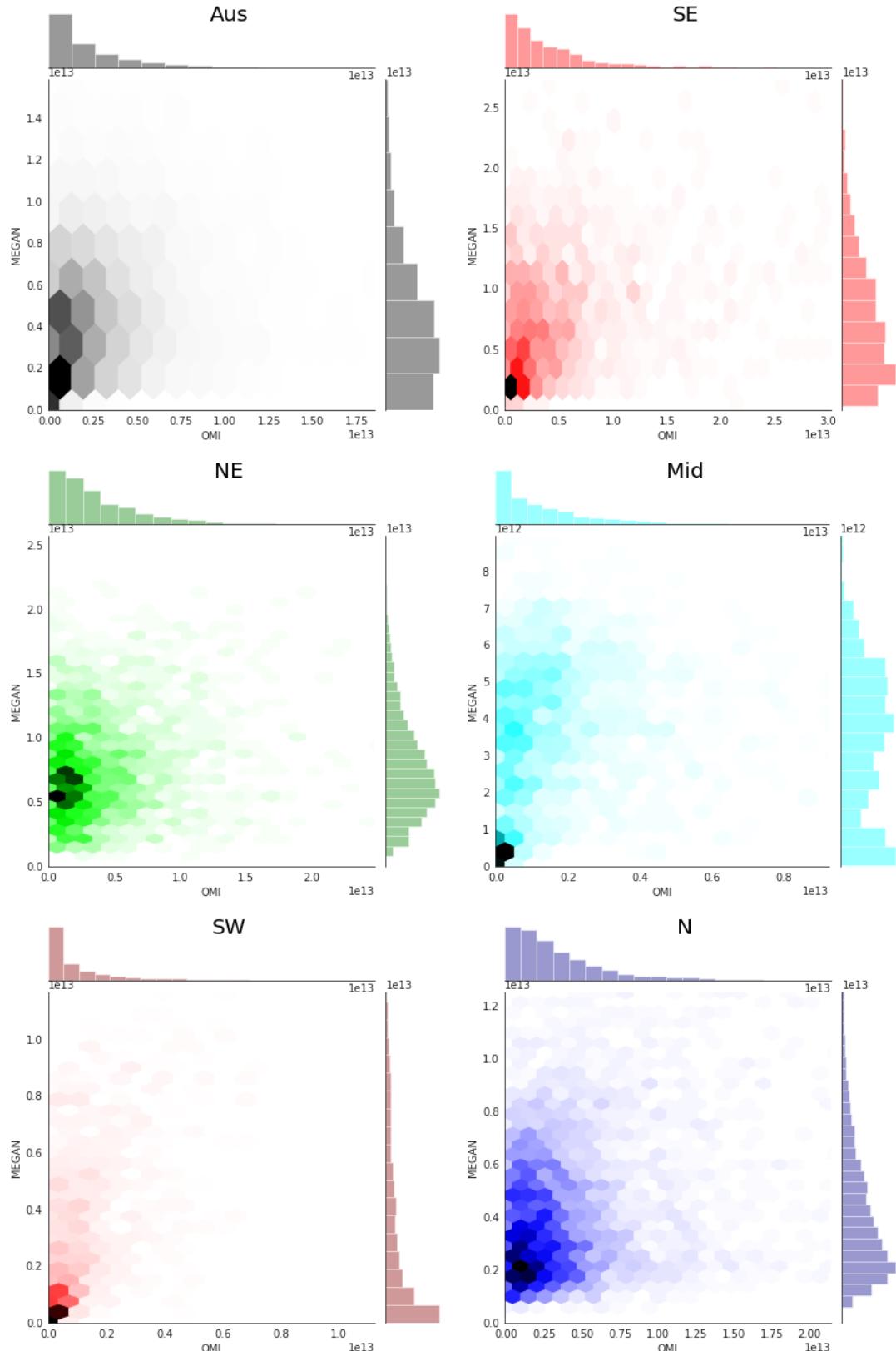


FIGURE 3.14: Scatter plot (binned hexagonally to show data-point frequency) along with the distributions of MEGAN (y axis) and the top down estimate (x axis). This figure is based on summer (DJF) midday values over multiple years. Coloured by regions shown in figure 3.11.

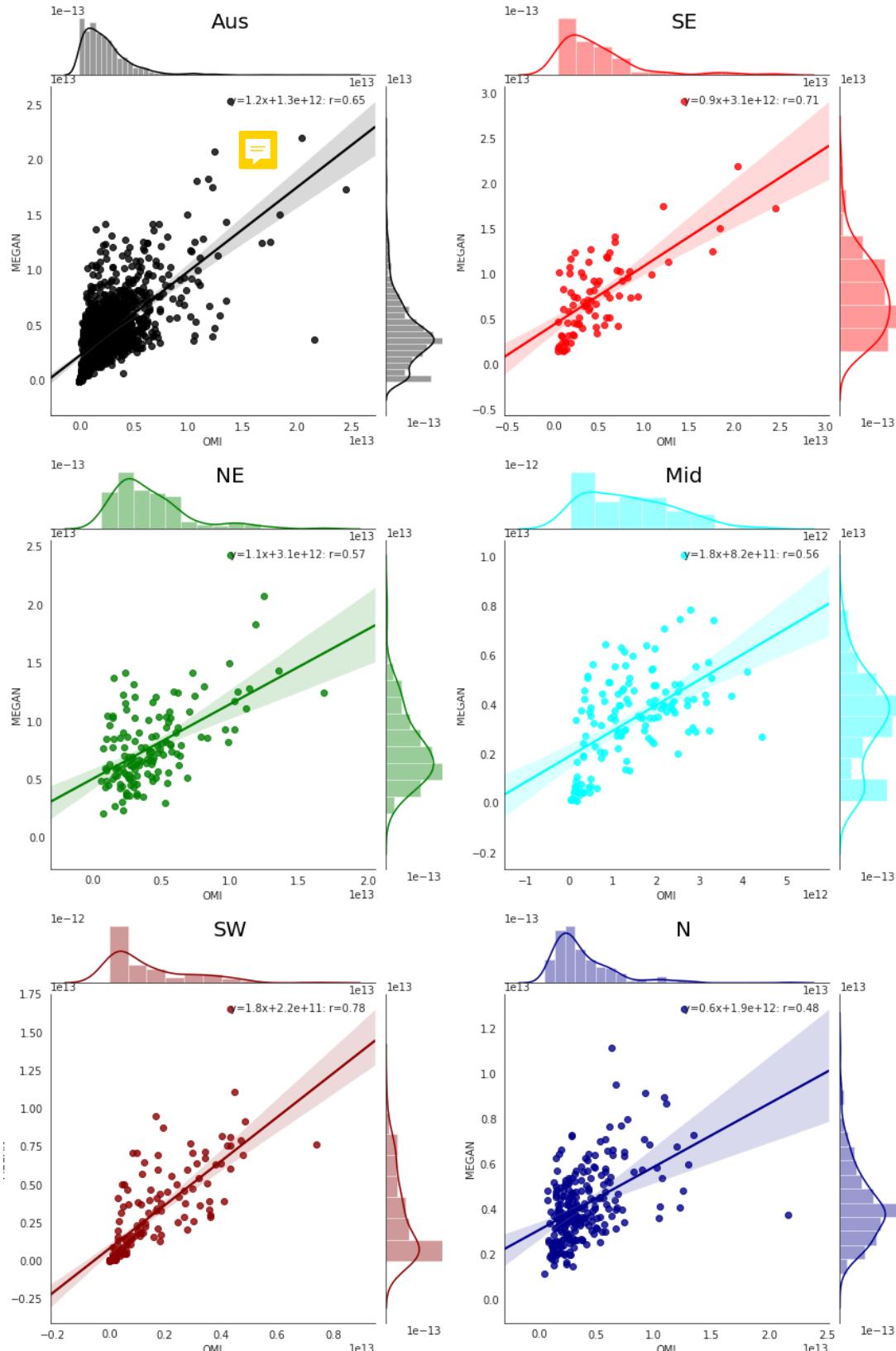


FIGURE 3.15: Scatter plot of MEGAN emissions against top down emissions using monthly averaged gridsquares as regression datapoints. Figures use multiple years of summer (DJF) midday values averaged monthly within each region shown. Coloured by regions shown in figure 3.11.

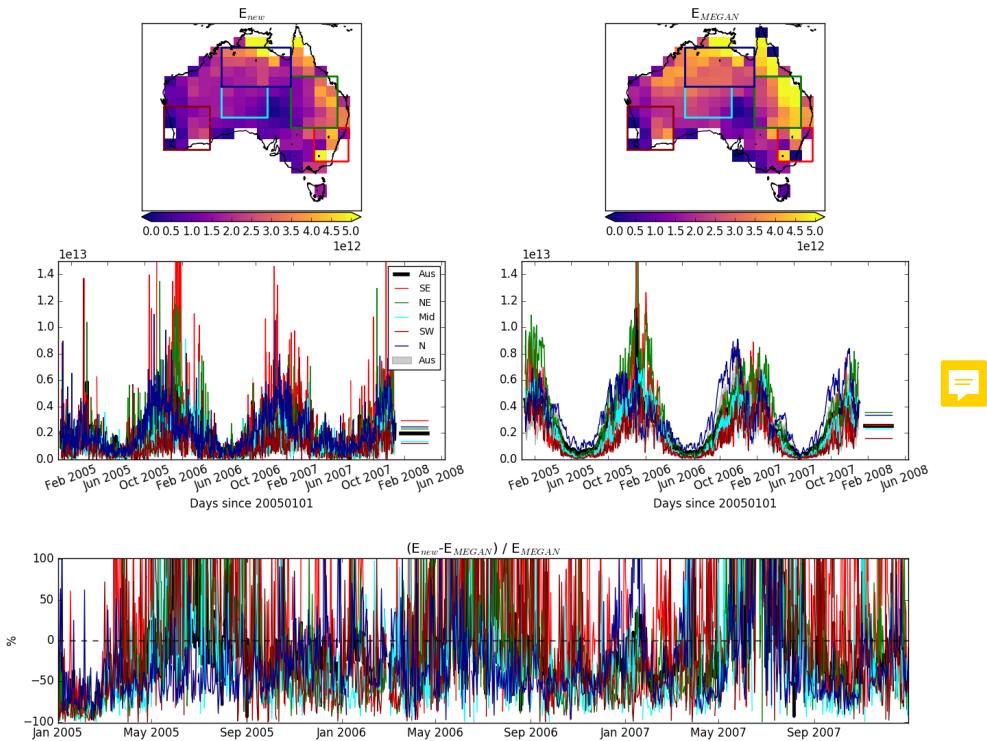


FIGURE 3.16: Emissions of isoprene estimates using OMI top down inversion (column 1: E_{OMI}) and MEGAN (column 2: E_{GC}). Row 1: overall averaged daily midday emission rates, from 1, Jan, 2005 to 1, May, 2013. Row 2: time series of daily midday averages for all of Australia along with several subregions (shown in row 1). Row 3: relative differences: $\frac{E_{MEGAN} - E_{OMI}}{E_{MEGAN}}$. The black lines and grey areas show the Australian mean and inter-quartile range respectively, while the coloured lines show the mean within the rectangles (of matching colours) shown in the first row.

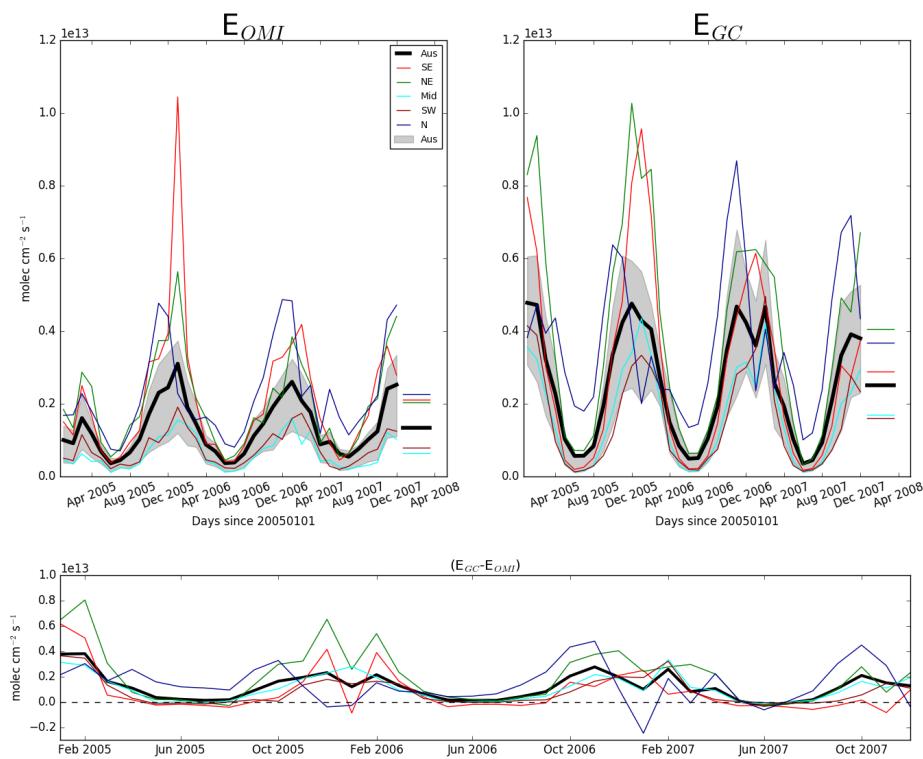


FIGURE 3.17: As figure 3.16 using monthly medians of the daily mid-day emissions estimates.

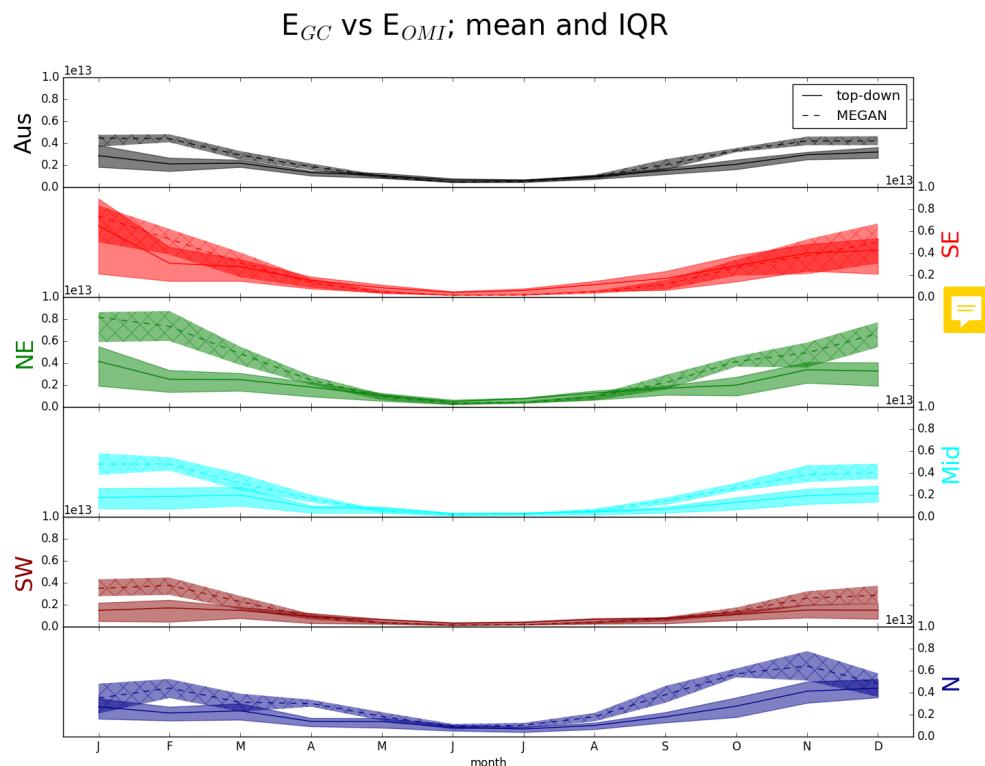


FIGURE 3.18: The multi-year monthly mean (lines) and IQR (shaded) of midday (13:00-14:00 LT) isoprene emissions estimates. Estimates come from MEGAN run by GEOS-Chem (E_{GC}), and the OMI top-down technique (E_{OMI}). The mean E_{GC} is shown by the dashed lines and hatched shaded areas show the IQR. E_{OMI} means 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 shown in Figure 3.11.

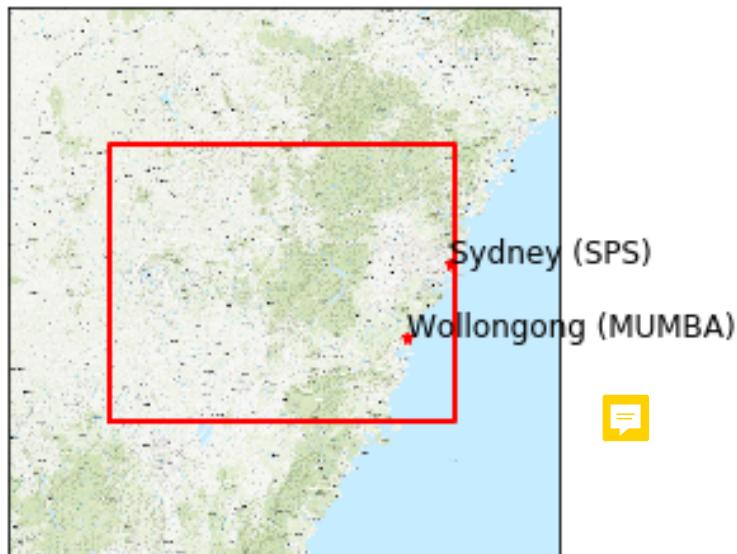


FIGURE 3.19

3.4 Uncertainty

3.4.1 Summary

Uncertainties introduced through the inversion process are hard 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 the overall uncertainties of calculating isoprene emissions using OMHCHO and GEOS-Chem in the top-down method. The major source of uncertainty lies in TODO, and calculations are more or less uncertain in the Winter. This limits temporal resolution of isoprene emissions estimates. Table TODO shows each term calculated in this work and the corresponding uncertainty estimate in summer and winter.

3.4.2 Top-down emissions

There are several factors which need to be considered when looking at the uncertainty in our emissions estimate. Things with their own inherent uncertainty include the modelled a-priori, modelled relationship between HCHO and isoprene, and satellite measurements. Important factors which need to be analysed for confidence in results

TABLE 3.4: Uncertainties in literature and here.

product	uncertainty	location	notes
satellite HCHO top-down E_{OMI}	40%	North America	(a) mostly due to cloud interference
	X%	where	(b)
	Y%	where	(c)
	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:

include the steady state assumptions, filtering techniques for fire and human influences, and the regression model for determining the isoprene to HCHO yield.

Uncertainty in satellite HCHO, along with top down emissions estimates E_{OMI} from literature is listed in table 3.4. The final determination of top-down emissions comes from equation 3.2: $E_{OMI} = \frac{\Omega - \Omega_0}{S}$. Assuming each term is independent, we use the following equations to estimate random error in E_{OMI} :

$$\begin{aligned} z = x + y : \Delta z &= \sqrt{(\Delta x)^2 + (\Delta y)^2} \\ z = x/y : \Delta z &= z \sqrt{\left(\frac{\Delta x}{x}\right)^2 + \left(\frac{\Delta y}{y}\right)^2} \\ \Omega - \Omega_0 = \Phi & \\ \Delta \Phi &= \sqrt{(\Delta \Omega)^2 + (\Delta \Omega_0)^2} \\ \Delta E_{OMI} &= E_{OMI} \times \sqrt{\left(\frac{\Delta \Phi}{\Phi}\right)^2 + \left(\frac{\Delta S}{S}\right)^2} \end{aligned}$$

In order to quantify ΔE_{OMI} we need to find the uncertainty in the underlying terms: ΔS , $\Delta \Omega$, and $\Delta \Omega_0$. For ΔS ($\Omega_{GC} = S \times E_{isop} + \Omega_0$ from equation 3.3) we examine related GEOS-Chem output in section 3.4.3, since S comes directly from the monthly linear regression of modelled isoprene emissions and column HCHO. Uncertainty in terms Ω , and Ω_0 come from both the OMHCHO slant columns (SC) and the AMFs calculated to transform them into vertical columns ($VC = SC / AMF$ from equation 2.6). Section ?? describes these calculations.

The summation of these uncertainties through standard quadrature rules provides an estimate of random error in the calculation of E_{OMI} . 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. Known biases:

OMHCHO up to 40% underestimated HCHO in the OMI satellite product (pixel bias, (Zhu et al. 2016; De Smedt et al. 2015; Barkley et al. 2013))

OMHCHO around 13% overestimation of monthly averaged HCHO (cloud-free bias, (Surl, Palmer, and González Abad 2018))

GEOS-Chem HCHO TODO under/over estimation of modelled HCHO due to coarse resolution (over prediction of low-NO_x oxidation pathway TODO:citation)

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.

3.4.3 Model Uncertainty

E_{OMI} estimation depends partly on the product it is trying to improve, using modelled yields based on MEGAN. Model uncertainty is difficult to accurately ascertain, generally an analysis of the model compared to in-situ measurements is performed, however there are few of these measurements over Australia. Here GEOS-Chem output is compared against the campaign datasets with the caveat that in-situ and point measurements are quite different to modelled (large) area averages.

Uncertainty in modelled yield is estimated through uncertainty in the regression slope (TODO:). Prior works use flight campaigns and in-situ data to verify HCHO columns in various locations (TODO: redo cite list from lit review). Figures X to D TODO: make figures comparing campaigns to model HCHO Yield calculations are performed at low resolution ($2^\circ \times 2.5^\circ$) which may lead to overestimation (Yu et al. 2016). TODO: how do I check this? Isoprene to HCHO yield between low and high NO_x (0.1 to 1 ppbv respectively) conditions has been estimated through box modelling to be 1.9 to 2.4 mol mol⁻¹ in Bauwens et al. (2016).

3.4.4 Satellite Uncertainty

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

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

a) is available in the OMI product and reduced through spatial and temporal averaging. Taking the eight day gridded average with horizontal resolution of 0.25 by 0.3125 degrees (latitude by longitude) typically reduces uncertainty by a factor of 1.5 to 4. b) could be determined through comparison of GEOS-Chem output to measured HCHO columns, if they existed. 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). Here we use 30% as a rough estimate of error in this term. TODO: Paul palmer calculation and combination for overall Satellite VC uncertainty per pixel and gridded. c) is also determined through a study of GEOS-Chem output, in relation to in-situ measurements. Since we expect oceanic background HCHO to be invariant, then variance in remote ocean HCHO can be used as a rough estimate of background uncertainty. 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 (1σ) of 25 – 27% in HCHO vertical columns with calculated AMFs where cloud fraction < 0.2 .

Two simple methods of looking at overall uncertainty from satellite measurements are performed here.

1. using the variance over the remote pacific ocean to provide relative uncertainty globally (e.g. De Smedt et al. 2012)
2. scaling up HCHO columns by 40% as an upper bound on satellite uncertainty

Analysing variance over the remote  pacific gives a quick measure of uncertainty if we assume HCHO levels within the region are stable. This region should be relatively invariant throughout any month so each month the standard deviation of the midday total column HCHO amounts from 15°S to 15°N , and 180°W to 120°W . Other literature has found satellite HCHO columns to be up to 40% too low, so scaling them up 40% is another way of quickly analysing how sensitive our calculations are to the satellite HCHO columns.

3.4.4.1 OMI Retrieval

TODO: Calculate remote pacific variance (also bin by latitudes) and plot over time.

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. BIRA use another method, and calculate the standard deviation of HCHO over the remote  ocean as the uncertainty (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 (~ 1 month). This means the standard deviation over this region can be used as a proxy for determination of the instrument error.

 TODO: uncertainty calculation on remote pacific OMI.

Satellite measured HCHO has been found to be biased low in several studies (eg. Zhu et al. 2016; De Smedt et al. 2015; Barkley et al. 2013). These papers use in-situ data to scale up the satellite HCHO columns for  their areas of interest, however Australia lacks sufficient HCHO measurements to do this. Satellite bias is seen to be as high as 40%, which we use as a simplistic method of quantifying potential satellite uncertainty. If satellite HCHO was scaled up by 40% our isoprene emissions estimates would increase by the same, as our estimate is proportional to satellite HCHO.

OMI is scaled up by up to 40% in several papers (cite) we consider HCHO scaled by 1 and 1.4 to be boundaries for modelled yield. If we infer from this that there is 40% bias and random uncertainty remains unchanged, then in the S term, we find TODO increase or decrease in ΔE_{OMI} of some amount through changing $\Delta S =$.

3.4.4.2 Satellite vertical column recalculations

OMI HCHO vertical columns are recalculated using GEOS-Chem V10.01 a priori HCHO and air density profiles (see chapter 2). The recalculations are considered assumed to be

no more or less uncertain than that used to calculate the default AMF, and uncertainty in the AMF recalculation is not used to alter the pixel uncertainty from the OMH-CHO product. 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 (AMF_{OMI}), the AMF with shape factor recalculated but the default scattering weights (AMF_{GC}), or the fully recalculated AMF (AMF_{PP}) 

Figure todo shows the emissions over Australia averaged within January 2005. This figure shows estimates from MEGAN (top), and top down estimates using OMH-CHO Ω s (row 1: OMI), Ω recalculated using GEOS-Chem shape factors (row 2: GC), and using the code from Paul Palmer's group (row 3: PP). Column 2 is emissions without applying anthropogenic or pyrogenic filters, Column 3 is calculated at the lower resolution of $2 \times 2.5^\circ$.

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

3.4.5 Fire Filtering

Figure 3.20 shows emissions estimates for January 2005, using three different HCHO columns as the basis: the original OMI satellite HCHO columns (Ω), those with AMF recalculated using a new a priori (Ω_{GC}), and those with AMFs recalculated using PP code (Ω_{PP}). The first row shows top-down emissions estimates, while the second row runs the same calculations without applying any fire or smoke filter. The Third row is the absolute difference between them: fire filtered minus standard emissions. 

3.5 Conclusions and implications

3.5.1 Effects from scaling emissions

Using our top-down emissions estimate to recalculate MEGAN isoprene emissions for Australia is detailed in section 3.2 How these changes affect HCHO and ozone model outputs are discussed in this section. 

3.5.1.1 HCHO levels

We examine the affect 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. We interpolated or something (TODO) the emissions over Australia into the inventories used by GEOS-Chem which reduced the emissions by X% per year (over Australia). The resulting simulation output shows that HCHO was reduced by X%, although if we boost monoterpenes by X% where the isoprene emissions were lowered then 

Wollongong IR measurements (see section ??) 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

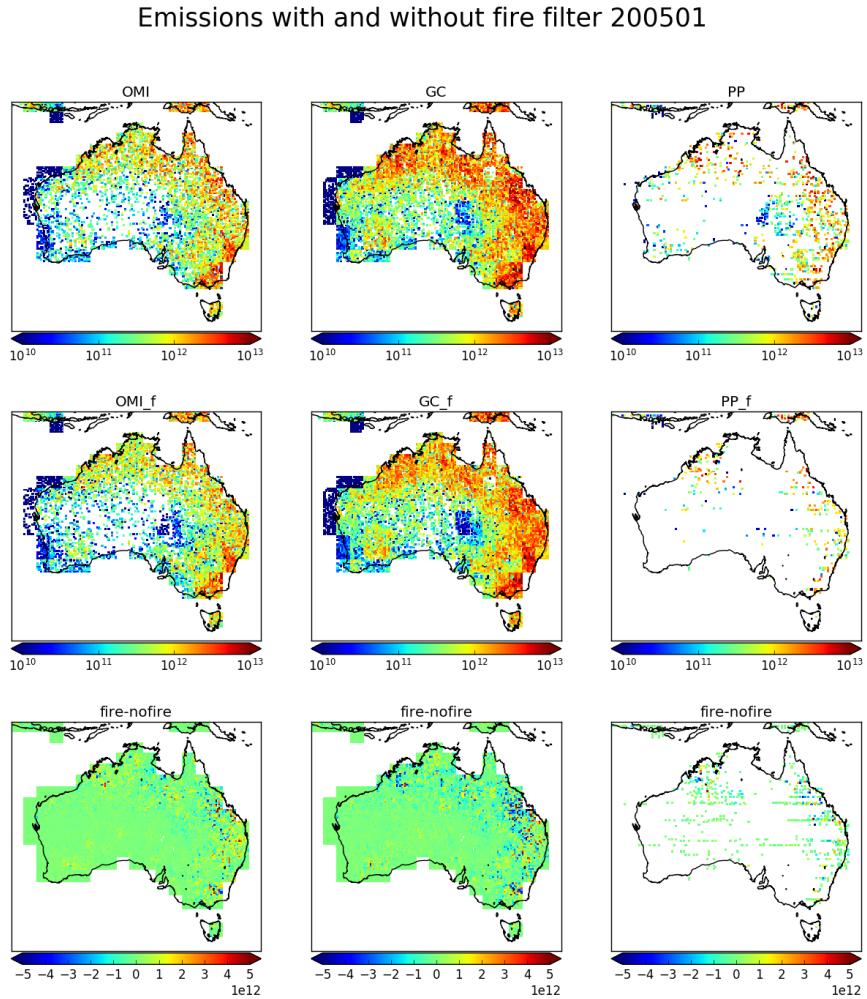


FIGURE 3.20: Emissions estimates using OMI satellite columns (column 1) recalculated with updated shape factor (column 2) and scattering weights (column 3). Turning off the fire and smoke filters gives emissions in row 2, while the difference between row 1 and row 2 is shown in row 3.

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.

A long term regression of anomalies from the multi-year monthly average shows that TODO. Figure TODO shows the trend for each month along with co-located HCHO from GEOS-Chem outputs. Figure TODO shows how campaign data compares against GEOS-Chem modelled HCHO before and after isoprene scaling is performed. The regressions show improvement/no improvement TODO however there are relatively few data points for comparison as the campaigns only lasted a few months.

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

3.5.1.2 Ozone levels

TODO: compare ozone after changing isoprene emissions in GEOS-Chem Changing isoprene emissions over Australia impacts the cities/coasts TODO determine this. This is likely due to the relatively clean atmosphere over the majority of Australia, which is likely NO_x limited outside of major population centres.

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).