

## Chapter 2

# Data and Modelling

### 2.1 Introduction

In this thesis the word model is most often used to represent a chemical transport model (CTM), which simulates chemistry and chemical transport through the atmosphere. Models of the atmosphere can be used to interpret measurements, estimate chemical concentrations at any scale, and predict atmospheric composition in the future. In remote sensing measurements, modelling is required in order to produce useful outputs. Models of ozone in the atmosphere are used broadly for international assessments of ozone precursor emissions, and estimating effects from related processes (such as radiation) (Young et al. 2017). Models provide an estimate of many trace gas concentrations, however verification is required, and generally performed using results from measurement campaigns. In situ measurements from campaigns or measurement stations can be used to examine what is happening at a particular location. These data are used to determine how accurate models or estimates are - however the utility is limited to where and when the measurements took place. In this thesis data from several campaigns are compared against model outputs and satellite datasets. Satellite datasets provide large amounts of data over most of the planet. However, they can have high uncertainty due to instrument limitations. Many datapoints can be averaged in order to reduce uncertainty. In this chapter several satellite datasets are combined to estimate biogenic HCHO amounts over Australia.

The first goal is to analyse Australia-specific HCHO concentrations measured by satellite, and determine isoprene sensitivity and any model bias. This leads into chapter 3 where biogenic HCHO columns are used to estimate isoprene emissions. The second goal is to quantify ozone transported from the stratosphere down into the troposphere (Chapter 4). The focus in this chapter is to describe and analyse model outputs and measurements along with how they are recalculated and compared. Section 2.2 details satellite and campaign datasets, and additionally describes model outputs. Measurement techniques used to retrieve the most utilised satellite dataset are outlined in Section 2.3. Section 2.4 describes the GEOS-Chem model, how it is run and what setup and outputs are used in this thesis. In Section 2.5 the process of using model outputs to recalculate satellite vertical columns is defined and analysed. In order to compare satellite data with other datasets, some work must be undertaken to avoid introducing bias (eg. Palmer et al. 2001; Eskes and Boersma 2003; Marais et al. 2012; Lamsal et al. 2014). One key step is to recalculate the satellite information using modelled data, detailed in section 2.6. The effects of these recalculations on

satellite HCHO is also examined. The creation and effects of filters used to remove non-biogenic influences are described in section 2.7.

## 2.2 Datasets

This section describes the datasets used in this thesis, along with an overview of the measurement techniques used for each. This includes modelled output, satellite measurements, and measurement campaigns. These datasets serve four purposes:

1. Model output validation in this chapter
2. Calculation of biogenic HCHO distribution over Australia in this chapter
3. Recalculated OMI formaldehyde columns are used as a basis for estimating isoprene emissions in Chapter 3
4. Extrapolation of ozone transport in chapter 4

I will also give details on filtering and interpolations which are undertaken when reading data, as each dataset has its own resolution. While I have not made any measurements myself, it is important to understand the techniques used in datasets I have utilised in order to understand possible anomalous datapoints or trends.

Uncertainty (or error) is present in each dataset and where possible the causes are explained. There are two types of error: systematic and random. Arguably the worst of these is systematic error (or bias). Bias normally indicates a problem in calculation or instrumentation. If the systematic error is known, it can be corrected for by either offsetting data in the opposite direction, or else fixing the cause. A proper fix can only be performed if the sources of error are known and there is a way of correcting or bypassing it. Random error is often reported as some function of a datasets variance, or uncertainty. It can be reduced through averaging either spatially or temporally. Temporal and/or spatial averaging decreases uncertainty by a factor of  $1/\sqrt{N}$  where N is the number of observations being averaged.

### 2.2.1 Satellite

Satellite data products are generally classed into several categories, level 0 through to level 3. Level 0 products are sensor counts and orbital swath data, level 1B data calibrates and geo-locates the level 0 data. Level 2 products additionally have temporal, spatial, solar, and viewing geometry information, as well as quality flags. To create level 2 data slant column density is determined and then translated into vertical column density. Level 3 data is a temporally aggregated subset of level 2 data, for instance monthly or yearly averages.

Satellites record near nadir (vertical) reflected spectra between around 250-700 nm split into spectral components at around 0.3 nm in order to calculate trace gases including O<sub>3</sub>, NO<sub>2</sub>, and HCHO (eg. Leue et al. 2001). Satellite measurements are generally performed using spectral fitting followed by conversion to vertical column densities. Several public data servers are available which include products from satellites, including NASAs Earthdata portal (<https://earthdata.nasa.gov/>) and the Belgian

Institute for Space Aeronomy (IASB-BIRA) Aeronomie site (<http://h2co.aeronomie.be/>).

Rayleigh and Mie scattering describe two kinds of particle effects on radiation passing through a medium. Rayleigh scattering is heavily wavelength dependent, and is the dominant form of scattering from particles up to roughly one tenth of the wavelength of the scattered light. Mie scattering more generally involves larger particles, and has less wavelength dependence. The effects of scattering are what gives us the information about substances in the atmosphere. The different particles and gases in the air have measurable properties seen by remote sensing devices such as a satellite. Although instruments will be more or less sensitive to various properties depending on altitude, radiation, and other parameters (e.g. Martin et al. 2002b).

Difficulties can arise when aerosols interfere with recorded spectra (eg. clouds, smoke, dust), however some of these can be detected and filtered out. Instruments including MODIS on board the Aqua and Terra satellites are able to determine aerosol optical depth (AOD), a measure of atmospheric scatter and absorbance. An AOD under 0.05 indicates a clear sky, while values of 1 or greater indicate increasingly hazy conditions. This is important in order to determine where measurements from other instruments may be compromised by high interference. Satellite measured AOD requires validation by more accurate ground based instruments like those of AERONET which uses more than 200 sun photometers scattered globally.

Soon much more satellite data will be available in the form of geostationary satellite measurements (Kwon et al. 2017). Geostationary satellites can provide temporally rich measurements over an area, as they are not sweeping around the earth but fixed relative to one latitude and longitude.

### 2.2.1.1 OMHCHO

OMI spectra are used in several products used in this thesis, including OMHCHO, OMNO2d, and OMAERUVd. OMHCHO data is used and modified extensively throughout this thesis, and so is discussed in more detail in Section 2.3. Calculation of column density and AMF are discussed respectively in sections 2.3.2 and 2.3.3.

### 2.2.1.2 OMNO2d

OMNO2d is a gridded daily level three product with good satellite pixels averaged into  $0.25 \times 0.25^\circ$  horizontally resolved bins. An example figure from Jan 29, 2005 is shown in figure 2.1, while an average for 2005 (global) is shown in figure 2.2. OMNO2 pixel resolution is 40 km by 130 km. NO<sub>2</sub> measured by OMI is used to check whether NO<sub>2</sub> is well represented by GEOS-Chem (see section 2.4.6 for the comparison between this product and GEOS-Chem calculations). It is also used to form the anthropogenic influence filter for OMHCHO (See section 2.7.2).

Like other satellite products, OMNO2d is influenced by a priori modelling which is required to convert slant path radiance to vertical columns. These models are generally low resolution ( $\sim 110$  km by 110 km), which leads to column smearing and difficulty detecting point sources of high NO emissions (Goldberg2018). Uncertainty in this product arises mostly from the calculation of the AMF (up to 50% of total error) (Lorente et al. 2017).

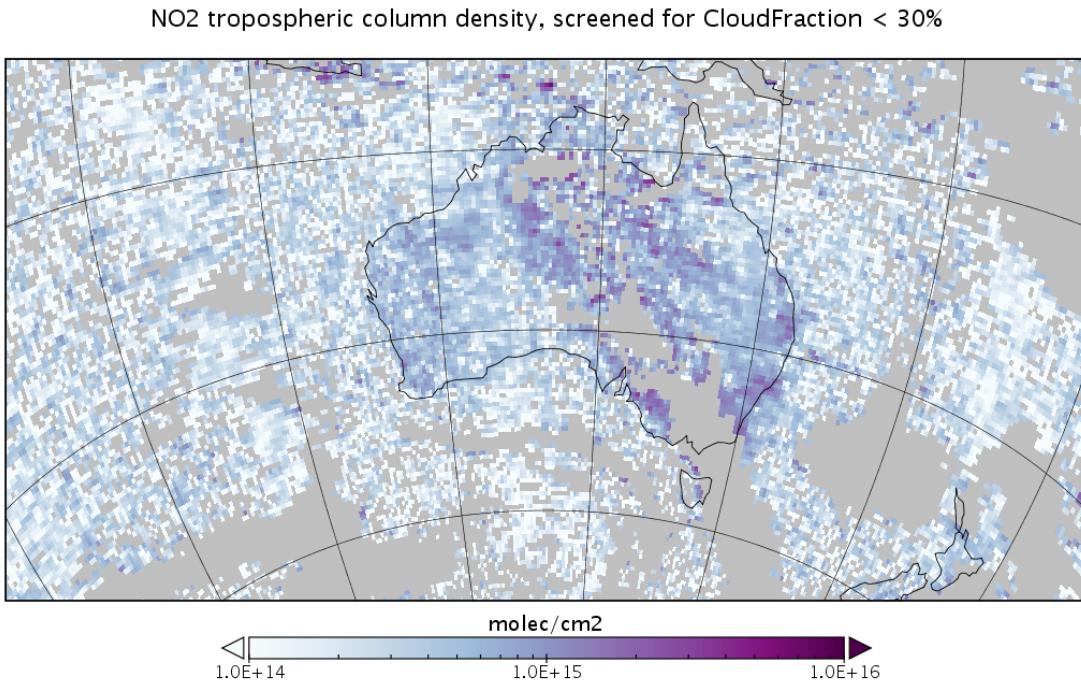


FIGURE 2.1: Example of NO<sub>2</sub> tropospheric columns taken from the OMNO2d product.

### 2.2.1.3 OMAERUVd

Aerosols in the atmosphere can be seen through their affects on light. Smoke and dust can be seen as an increase in aerosol optical depth (AOD) (see section 2.3.2). This is due these particles scattering and absorbing UV radiation (Ahn2008). A data product provided by Earthdata ([https://disc.gsfc.nasa.gov/datasets/OMAERUVd\\_V003/summary](https://disc.gsfc.nasa.gov/datasets/OMAERUVd_V003/summary)) called OMAERUVd (DOI: 10.5067/Aura/OMI/DATA3003) is used in this thesis.

OMAERUVd provides a useful dataset which allows detection of areas which may be smoke affected. The product contains AOD and aerosol absorption optical depths (AAOD) at three wavelengths (354, 388, and 500 nm), along with UV aerosol index (UVAI). The OMAERUVd product is level three gridded daily data, based on quality filtered level two swath pixels which are then gridded by averaging. The product is most sensitive to error in the form of sub-pixel scale cloud interference, so I select AAOD as the basis for my smoke filter as it is least affected by clouds (Ahn2008).

In this work AAOD is mapped from 1x1°horizontal resolution to 0.25x0.3125°using nearest value mapping. The AAOD at 500 nm wavelength is used to determine smoke influence, although any of the provided wavelengths would be affected by smoke plumes and could also be used. This daily AAOD is compared to a threshold to create a daily smoke filter, any areas with AAOD> 0.03 are considered to be potentially smoke plume affected (see section 2.7.1).

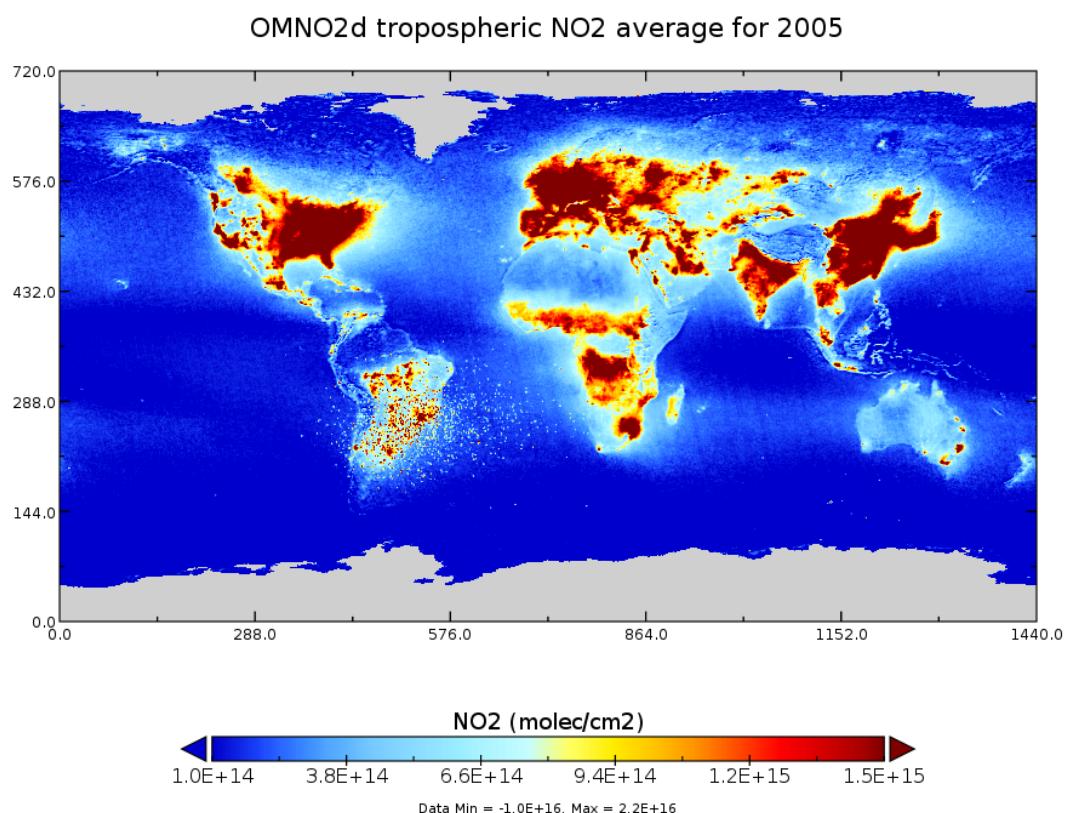


FIGURE 2.2: Average 2005 tropospheric NO<sub>2</sub> from OMNO2d with pixels screened for < 30% cloud cover.

### 2.2.1.4 MOD14A1

MOD14A1 is a gridded daily satellite based dataset of fire counts at  $1 \times 1 \text{ km}^2$  horizontal resolution. Fire observations are performed four times daily from Terra (10:30 LT, 22:30 LT) and Aqua (01:30 LT, 13:30 LT). The fire pixels are detected based on parameters including apparent pixel temperature and the nearby background temperature. The dataset is obtained from NASA Earth Observations (NEO) that is part of the EOS Project Science Office at the NASA Goddard Space Flight Center [https://neo.sci.gsfc.nasa.gov/view.php?datasetId=MOD14A1\\_M\\_FIRE](https://neo.sci.gsfc.nasa.gov/view.php?datasetId=MOD14A1_M_FIRE). This product is downloaded and binned into a lower resolution (using the sum of fire pixels) to create an active fire influence mask (see section 2.7.1).

### 2.2.1.5 AIRS CO columns

In chapter 4, potential biomass burning plumes are identified using satellite observations of CO from the AIRS (Atmospheric Infra-red Sounder) instrument aboard the Aqua satellite (Texeira 2013). CO is used as a proxy for biomass burning plumes, and used to qualitatively attribute ozone intrusion events as explained in section 4.2.5. This is a separate method of detecting fire influence near specific sites through visual analysis.

### 2.2.1.6 Uncertainties

While satellite data is effective at covering huge areas (the entire earth) it only exists at a particular time of day, is subject to cloud cover, and generally does not have fine horizontal or vertical resolution. Concentrations retrieved by satellites have large uncertainties, which arise in the process of transforming spectra into total column measurements, as well as instrument degradation (satellite instruments are hard to tinker with once they are launched). Uncertainty in transforming satellite spectra comes from a range of things, including measurement difficulties introduced by clouds, and instrument sensitivity to particular aerosols (Millet et al. 2006). Many products require analysis of cloud and aerosol properties in order to estimate concentration or total column amounts (Palmer et al. 2001; Palmer 2003; Marais et al. 2012; Vasilkov et al. 2017). The main source of error in satellite retrievals of HCHO are due to instrument detection sensitivities, and calculation of the air mass factor (AMF) which converts slanted light path concentrations into a vertical profile (Millet et al. 2006). Calculations of the AMF performed by different groups tend to agree fairly well, as long as all the a priori and ancillary data is similar. Large differences can occur depending on the a priori vertical profile, trace gas concentrations, and cloud properties (Lorente et al. 2017). Choice of RTM and interpolation operations have a relatively small affect compared to the assumed state of the atmosphere, with high structural uncertainty introduced at this stage of AMF calculation - as shown in Lorente et al. (2017).

A common way of reducing satellite uncertainty is through oversampling or temporal averaging. This is done frequently for trace gases (which are often near to the detection limit over much of the globe). For example: Vigouroux et al. (2009) reduce the measurement uncertainty (in SCIAMACHY HCHO columns) by at least a factor of 4 through averaging daily over roughly 500km around Saint-Denis, and only using

days with at least 20 good measurements. Another example of this can be seen in Dufour et al. (2008), where monthly averaging is used to decrease the measurements uncertainty at the cost of temporal resolution.

In cloudy, hazy or polluted areas measurements are more difficult to analyse (e.g. Palmer 2003; Marais et al. 2014). Recent work by Vasilkov et al. (2017) showed that updating how the surface reflectivity is incorporated into satellite measurements can change the retrievals by 50 % in polluted areas.

## 2.2.2 Model datasets

### 2.2.2.1 CTM output

GEOS-Chem model output is used extensively in this thesis and is discussed in more detail in section 2.4. Section 2.4.7.1 specifically describes the model outputs used in this thesis. These are generally resolved to 47 vertical levels from the ground up to 0.01 hPa, at 2x2.5°horizontal resolution.

### 2.2.2.2 Meteorological reanalysis

Synoptic scale weather patterns are taken from the European Centre for Medium-range Weather Forecasts (ECMWF) Interim Reanalysis (ERA-I) (Dee et al. 2011). These are used in chapter 4 to determine typical weather systems for stratospheric ozone intrusions. The version used was ERA-Interim, which was the most up to date at the time (2016) but has since been superseded by ERA5.

### 2.2.2.3 CPC temperatures

The Climate Prediction Center (CPC) provides a product with maximum daily land-surface temperature at 0.5x0.5°horizontal resolution. This data is used to check the correlation between HCHO and temperature at a higher resolution than is provided by GEOS-Chem output. A full description of the data can be found at <https://www.esrl.noaa.gov/psd/data/gridded/data.cpc.globaltemp.html>. CPC Global Temperature data is provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their web site at <https://www.esrl.noaa.gov/psd/>. An example of one day of land temperature output is shown in figure 2.3.

## 2.2.3 Campaign datasets

In this thesis data from several measurement campaigns are used to examine accuracy of modelled data at specific sites. Figure 2.4 shows the locations of each of the campaigns used in this work. These took place over disparate times, and are in situ measurements which require hard to directly compare against GEOS-Chem output which is averaged over a large horizontal space.

TODO: these summaries.

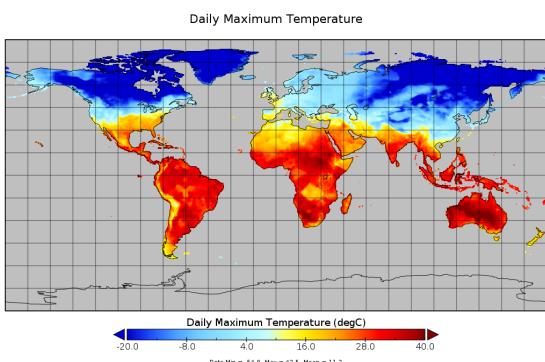


FIGURE 2.3: CPC daily maximum temperature dataset output for 1, Jan, 2005.

TABLE 2.1: Detection limits for MUMBA

Dates	HCHO (ppb)	Isoprene (ppb)	Ozone (ppb)
21/Dec/2012 - 29/Dec/2012	0.205	0.003	0.5
29/Dec/2012 - 18/Jan/2013	0.105	0.005	0.5
19/Jan/2013 - 15/Feb/2013	0.186	0.003	0.5

### 2.2.3.1 Measurements of Urban, Marine and Biogenic Air (MUMBA)

The MUMBA campaign ([PatonWalsh2017](#)) measured various compound abundances including isoprene, formaldehyde, and ozone from 21 December 2012 to 15 February 2013. These measurements took place in Wollongong, 10 m above ground level (40 m above sea level). Ozone was measured by Thermo UV absorption with 1-minute time resolution averaged into hourly outputs. Isoprene and HCHO were measured by Ionicon Proton-Transfer-Reaction Mass spectrometer (PTR-MS), with a time resolution of 3-minutes, averaged each hour. Detection limits varied due to instrument conditions, and are listed in table 2.1. The full dataset has been published on PANGAEA (DOI:[10.1594/PANGAEA.871982](#)) ([Guérette et al. 2018](#)).

In this thesis we assume uncertainty in this product is as estimated by TODO at (50%). Although the uncertainty determined through calibration measurements was only 15% ([Guérette et al. 2018](#)), this does not account for competing trace gas interference (such as furan). The readings are resampled to hourly averages. Measurements below the detection limit are set to half of the detection limit when reading this dataset.

### 2.2.3.2 Sydney Particle Studies (SPS1, SPS2)

Two trace gas measurement campaigns took place at the Westmead air quality station. Stage 1 (SPS1) from 5 February to 7 March, 2011 and stage 2 (SPS2) from 16 April to 14 May, 2012. Two instruments measured VOC concentrations: a PTR-MS, and a gas chromatograph (GC) with a flame ionisation detector (FID). The PTR-MS uses chemical ionisation mass spectrometry and can quantify VOCs at high temporal resolution (< 1 s). It was calibrated several times per day against HCHO, isoprene,  $\alpha$ -pinene, and several other VOCs, further measurement specifics can be found in Dunne et al. ([2018](#)).



FIGURE 2.4: Locations of Australian campaigns which are analysed within this thesis

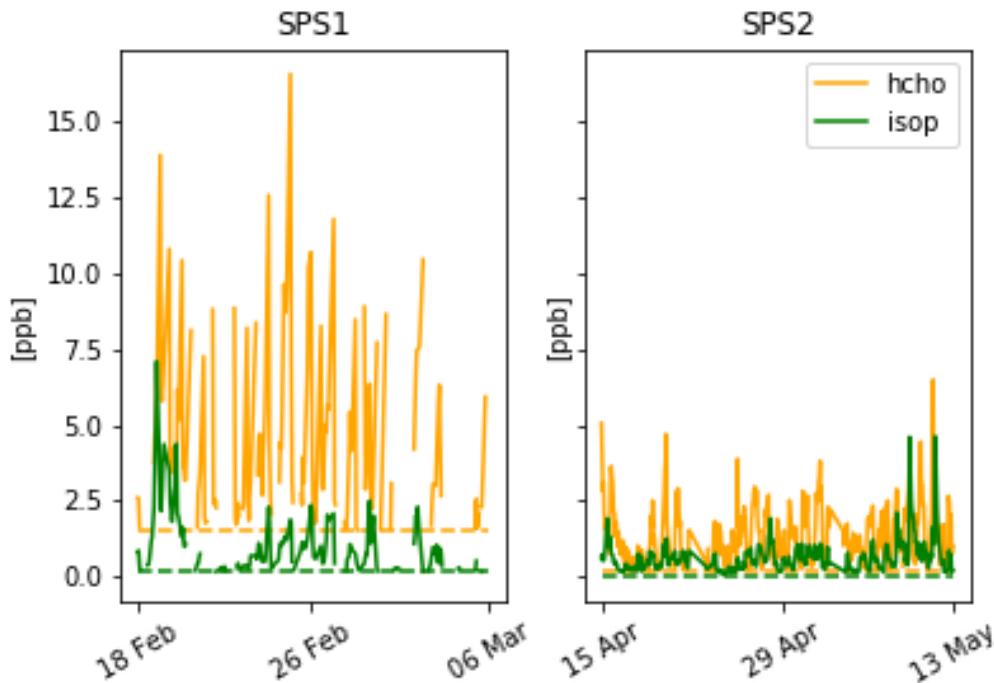


FIGURE 2.5: SPS HCHO (yellow) and isoprene (green) time series, along with detection limits (dashed). SPS 1 (left) took place in late summer 2011, while SPS 2 (right) occurred during autumn 2012.

The output lists hourly averaged ppbv concentrations of trace gases based on the mass to charge ratio ( $m/z$ ), which for isoprene is 69. It is possible that other chemicals (such as furan, with the same  $m/z$ ) interfered with this value, especially at low ambient isoprene concentrations and towards the end of autumn (SPS2) when wood fires start to become frequent (Guérette et al. 2018). The GC-FID analysed samples collected in multi-absorbent tubes, with lower temporal resolution but no interference. Further details for this method can be found in Cheng2016 GC-FID data is averaged from 0500-1000 LT, and 1100-1900 LT, while PTR-MS data is averaged hourly. This includes significant differences between measurement devices when detecting isoprene, potentially due to interfering compounds in the PTR-MS (Dunne et al. 2018).

Figure 2.5 shows isoprene and formaldehyde over the course of these two campaigns, as well as the detection limits (dashed lines), as measured by PTR-MS. In order to compare with GEOS-Chem output (see section 2.4) a daily average and an midday time (13:00-14:00 LT) average are both created from these data. In averaging, any measurements below the machine detection limit are set to half of the detection limit, as done in Lawson et al. (2015). This should minimise any introduced bias.

### 2.2.3.3 Ozonesondes

Ozonesonde data come from the World Ozone and Ultraviolet Data Centre (WOUDC). Ozonesondes are weather balloons which measure from the surface to around 35km.

Ozonesondes provide a high vertical resolution profile of ozone, temperature, pressure, and humidity. Generally the instrument will perform 150-300 measurements in the troposphere with ozone mixing ratios quantified by an electrochemical concentration cell (<http://www.ndsc.ncep.noaa.gov/organize/protocols/appendix5/>).

Ozonesondes are launched approximately weekly from Melbourne (38° S, 145° E), Macquarie Island (55° S, 159° E) and Davis (69° S, 78° E). Melbourne, a major city with more than 4 million residents (**ABS2016**), may be affected by anthropogenic pollution in the lower troposphere. Actual releases are north of the central business district in the Broadmeadows suburb. Macquarie Island is in the remote Southern Ocean and unlikely to be affected by any local pollution events. Davis (on the coast of Antarctica) is also unlikely to experience the effects of anthropogenic pollution. More information on this dataset is given in section 4.2.

#### 2.2.3.4 Uncertainties

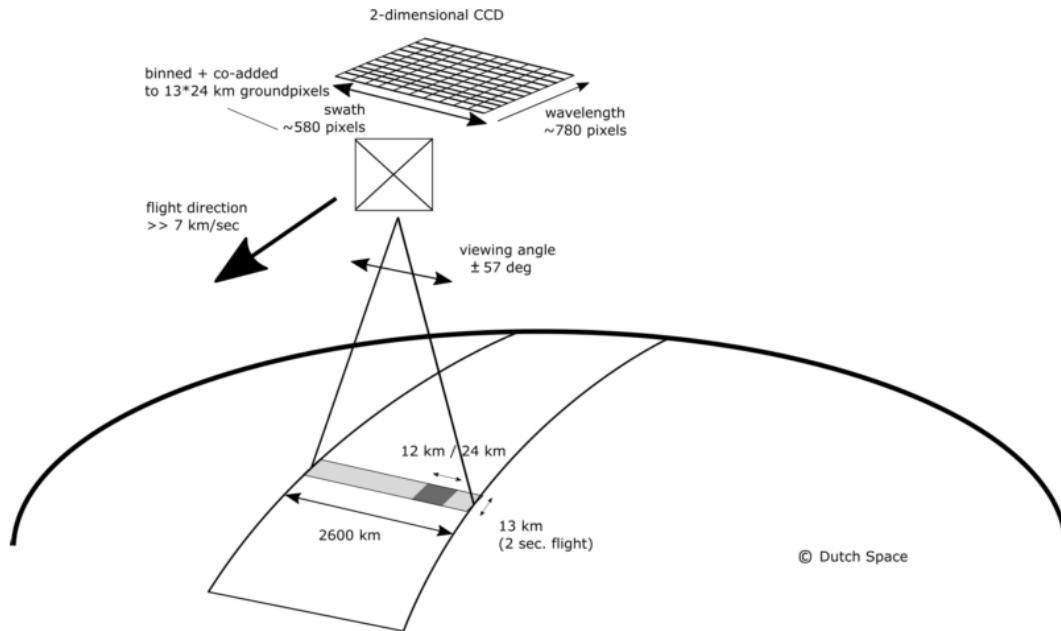
In situ measurements contain errors, and depending on the device used and chemical being measured this error can be significant. The major sources of uncertainty in measurement techniques included interference from non-target compounds and under-reporting (eg. Dunne et al. 2018; Guérette et al. 2018). Overall isoprene uncertainty in measurements analysed by Dunne et al. (2018) was a factor of 1.5 to 2. This can feed into uncertainties in modelling and satellite retrievals, as verification and correlations are affected.

### 2.3 OMHCHO

One satellite product used extensively in this thesis is named OMHCHO: from NASA’s Earth Observing System’s “Aura”, which provides several other useful datasets (products). Aura orbits the earth in a polar sun-synchronous pattern, circling the earth on a plane coincident with the sun and the poles. Aura houses the Ozone Monitoring Instrument (OMI), a near-UV/Visible Charged Coupled Device (CCD) spectrometer. The OMI instrument on board AURA has been active since July 2005, it records spectra from 264-504 nm using an array of 60 detectors with mid-resolution (0.4-0.6 nm). This band of wavelengths allows measurements of trace gases (among other quantities).

From here on the word pixel is used to describe one data point retrieved by OMI, each pixel includes a latitude and longitude within OMI’s data product. Figure 2.6 shows the details of OMI’s detector array and measurement resolutions. OMI measures atmospheric trace gases including NO<sub>2</sub>, SO<sub>2</sub>, BrO, HCHO, O<sub>3</sub>, and aerosols. OMI measurements occur from right to left on a band covering 115°, resulting in swaths of around 2600 km, with pixel sizes from 13x24 km<sup>2</sup> at nadir to 26x135 km<sup>2</sup> at the swath edges (Gonzalez Abad et al. 2015). The swaths cover Earth daily, both on the light and dark side of the planet, only daytime measurements provide useful near-UV/Visible information.

The latest OMHCHO algorithm uses a shape factor determined from GEOS-Chem using 47 vertical levels at monthly temporal resolution and 2° latitude by 2.5° longitude horizontal resolution (Gonzalez Abad et al. 2015). The GEOS-Chem model has



Channel	Wavelength range	Spectral resolution	Spectral sampling	Ground pixel size
UV1	264–311 nm	0.63 nm = 1.9 px	0.33 nm px <sup>-1</sup>	13 × 48 km
UV2	307–383 nm	0.42 nm = 3.0 px	0.14 nm px <sup>-1</sup>	13 × 24 km
VIS	349–504 nm	0.63 nm = 3.0 px	0.21 nm px <sup>-1</sup>	13 × 24 km

FIGURE 2.6: Figure 1 and Table 1 from Schenkeveld et al. (2017), with the following caption “An impression of OMI flying over the Earth. The spectrum of a ground pixel is projected on the wavelength dimension of the charge-coupled device (CCD; the columns). The cross-track ground pixels are projected on the swath dimension of the CCD (the rows). The forward speed of  $7 \text{ kms}^{-1}$  and an exposure time of 2 s lead to a ground pixel size of 13 km in the flight direction. The viewing angle of  $114^\circ$  leads to a swath width on the ground of 2600 km.” The table shows the optical properties for OMIs three channels.

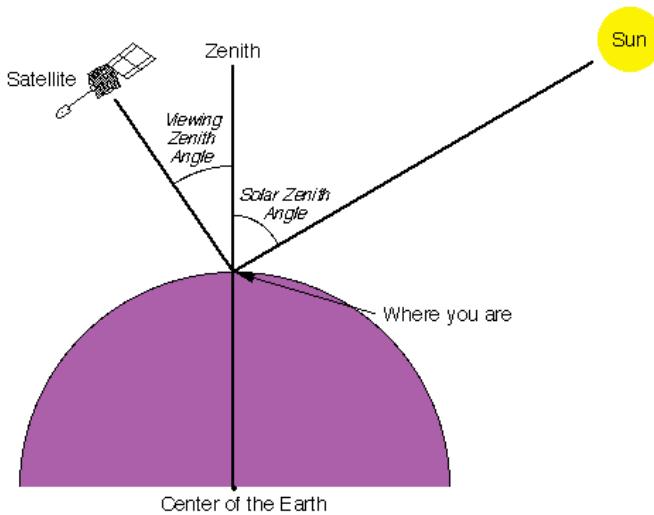


FIGURE 2.7: Solar and viewing zenith angles, image copied from Wikipedia (2016), originally from a NASA website.

been substantially updated since then. In this thesis the more recent version V10.01 is used to recalculate the vertical column HCHO (details are shown in section 2.6).

OMI uses a Differential Optical Absorption Spectroscopy (DOAS) based technique to read HCHO along the path of light which reaches the satellite instrument. The first step is to determine how much HCHO is in the path of light between the sun and detector. Measurements done using DOAS often apply a forward radiative transfer model (RTM) such as LIDORT (see section 2.3.4) in order to determine a trace gas' radiative properties at various altitudes. The forward RTM used for satellite data products also involve functions representing extinction from Mie and Rayleigh scattering, and the effect of these on spectra. These RTM are also required to account for (often estimated) atmospheric parameters such as albedo. The next step is to transform the calculated amounts along the non-vertical light path into vertical column amounts. This is done by applying an AMF. In the absence of atmospheric scattering a simple geometric AMF can be defined as a function of the solar zenith angle. The solar zenith angle ( $\theta_s$ ) and the satellite viewing angle ( $\theta_v$ ) are shown in image 2.7. However, in the UV-VIS region of the spectrum, Rayleigh and Mie scattering (see section 2.3.2) must be accounted for.

Atmospheric HCHO detected by satellite requires that other trace gases with similar features near the HCHO affected wavelengths are accounted for. A DOAS fit determines the total column amount of a trace gas along the path that the instrument views. This uses the Beer-Lambert law where radiance is reduced as light travels through a medium. I use the NASA OMHCCHOv003 data product (Gonzalez Abad et al. 2015), with HCHO determined using the spectral window 328.5 nm–356.5 nm. The algorithm used is based on direct fitting of radiances, and accounts for competing absorbers, under-sampling, and Ring effects. An OMI radiance measurement over the remote Pacific ocean is used instead of an irradiance measurement. This means that the slant columns ( $\Omega_s$ ) are formed from the spectra differential with respect to the

TABLE 2.2: OMI quality flag values table from Kurosu and Chance (2014)

Value	Classification	Rational
0	Good	Column value present and passes all quality checks; data may be used with confidence.
1	Suspect	Caution advised because one or more of the following conditions are present: <ul style="list-style-type: none"> <li>• Fit convergence flag is <math>&lt; 300</math> but <math>&gt; 0</math>: Convergence at noise level</li> <li>• Column <math>+2\sigma</math> uncertainty <math>&lt; 0 &lt;</math> Column <math>+3\sigma</math> uncertainty</li> <li>• Absolute column value <math>&gt;</math> Maximum column amount (<math>1e19</math> molec <math>\text{cm}^{-2}</math>)</li> </ul>
2	Bad	Avoid using as one of the following conditions are present: <ul style="list-style-type: none"> <li>• Fit convergence flag is <math>&lt; 0</math> : No convergence, abnormal termination</li> <li>• Column <math>+3\sigma</math> uncertainty <math>&lt; 0</math></li> </ul>
$< 0$	Missing	No column values have been computed; entries are missing

radiance reference column over the Pacific. The full method details for slant column retrieval by OMI are outlined in supplemental section A.5.1, or in the technical document (DOI: 10.5067/Aura/OMI/DATA2015). Slant columns range from  $\sim 4 \times 10^{15}$  to  $\sim 6 \times 10^{16}$  molec  $\text{cm}^{-2}$ , with uncertainties from 30% (larger columns) to over 100% (smaller columns) (Gonzalez Abad et al. 2015).

### 2.3.1 Pixel filtering

This thesis uses the level 2 product swath output from the NASA earth data web portal. OMHCHO level two data includes 14-15 daily swaths of measurements. Each swath contains roughly  $9 \times 10^4$  pixels, each of which includes latitude, longitude vertical column HCHO, along with all the ancillary data required to make the vertical column and several data quality metrics. The OMHCHO dataset has a quality flag which can be used to remove unlikely or poor satellite measurements. The states represented by this quality flag are shown in table 2.2 which is taken from Kurosu and Chance (2014). Filtering bad or missing measurement pixels is performed prior to any other filtering, this includes the datapoints affected by the row anomaly. This anomaly (`rowanomaly_url`) affects radiance data at particular viewing angles, corresponding to a row on the CCD detectors, and is dynamic over time. The slant columns affected are flagged and removed before any further processing.

Each  $\sim 90$  minutes the AURA satellite sweeps over the sunny side of the planet of which around 50 k – 80 k of the roughly 90 k pixels are classified as good. Each pixel contains several important pieces of data which are needed for recalculation of the HCHO vertical column: the total column of HCHO ( $\Omega$ ; molec cm $^{-2}$ ), cloud fraction, associated shape factor, AMF, geometric AMF, scattering weights and their vertical altitudes (hPa), viewing zenith angle, solar zenith angle, latitude, longitude, OMI sensor track, main data quality flag, cross track flag, and total column uncertainty. All of these data are needed in order to reconstruct the total vertical column using a modelled a priori shape factor rather than NASA's included a priori shape factor. Each pixel includes an estimate of the cloud fraction created using the OMI cloud product OMCLDO2. If greater than 40% of a pixel measurement is cloudy (ie. cloud fraction  $> 0.4$ ) then the pixel is removed from subsequent analysis. This removes around 30% of the pixels which remain after filtering out the bad or missing data.

One more filter is applied before any calculations, to remove unreasonable column amounts. Due to numerous highly negative vertical columns (beyond what is expected) a screen is applied to remove any pixels with vertical columns outside the range of  $-5 \times 10^{15}$  to  $1 \times 10^{17}$ . This has been performed previously in TODO

### 2.3.2 DOAS

The DOAS technique uses solar radiation absorption spectra to measure trace gases through paths of light. Beer's law states

$$T = I/I_0 = e^{-\tau} \quad (2.1)$$

with T being transmittance,  $\tau$  being optical depth, and  $I, I_0$  being radiant flux received at instrument and emitted at source respectively. The Beer-Lambert law of extinction allows spectroscopic measurement of absorbing chemical species (absorbers) in the atmosphere:

$$I_B = I_{B_0} e^{-\tau_s} \quad (2.2)$$

where  $I_B, I_{B_0}$  is backscattered intensity with and without the absorber respectively, and  $\tau_s$  is the optical thickness of the absorber along the measured path between source and instrument.

$\tau$  can be described using the scattering and absorption cross section area ( $\alpha$ , cm $^2$ ) and density ( $\eta$ , molec cm $^{-3}$ ) of an absorber as follows:

$$\tau = \int \alpha(s) \eta(s) ds \quad (2.3)$$

$\tau$  through a medium is the sum of optical thicknesses of each absorber within the measured path (s) substituting equation 2.3 into equation 2.2 leads to

$$I = I_0 \exp \left( \sum_i \int \eta_i \alpha_i ds \right)$$

Where  $i$  represents a chemical species index, and the integral over ds represents integration over the path from light source to instrument.

Another way of describing optical depth (also called optical thickness) is the natural logarithm of the ratio of incident radiant power to transmitted radiant power through a material (from equation 2.2). In the atmosphere we are interested in the optical depth of various chemical species, and we use incoming solar radiation to determine this. The difference between solar radiation at the top of the atmosphere and the earth's surface defines the atmospheric optical depth along the path of observation.

$$\tau = \ln \frac{\phi_e^i}{\phi_e^t}$$

where  $\phi_e^i$  is radiant flux seen at the earth surface,  $\phi_e^t$  is the solar radiant flux which arrives at the top of the atmosphere. In the atmosphere, optical depth can be due to several factors including scattering, chemical absorbance, and aerosols.

### 2.3.3 Air mass factor (AMF)

To convert the trace gas profile from a reflected solar radiance column (slanted along the light path) into a purely vertical column requires calculations of an air mass factor (AMF). In satellite data, the AMF is typically a scalar value for each horizontal grid point which will equal the ratio of the total vertical column density to the total slant column density. This value requires calculations to account for instrument sensitivities to various wavelengths over resolved altitudes, and is unique for each trace gas under consideration. An AMF characterises measurement sensitivity to a trace gas at various altitudes Palmer et al. 2001, e.g. Lorente et al. (2017) show that AMF calculations can be the largest source of uncertainty in satellite measurements. Another way of describing AMFs are as measures of how radiance at the top of the atmosphere (TOA) changes with trace gas optical depths at specific altitudes (Lorente et al. 2017). Calculation of the AMF is important as it is multiplied against the estimated slant columns in order to give vertical column amounts.

DOAS column retrievals are an integration of a trace gas over the instruments viewing path, in order to convert this total to a vertically distributed column a few assumptions and estimates are required. The vertical profile of a trace gas is assumed or estimated via a CTM, while its scattering and radiative properties are calculated at prescribed altitudes using an RTM. These properties are combined into a single array called the AMF. Two examples of this are GOME-2 (on the MetOp-A satellite) products ([http://atmos.caf.dlr.de/gome/product\\_hcho.html](http://atmos.caf.dlr.de/gome/product_hcho.html)), and OMI products which respectively use LIDORT combined with IMAGESv2 and GEOS-Chem for processing (Instrument 2002; Gonzalez Abad et al. 2015). AMFs are unique to each trace gas and due to their complexity and the influence of cloud cover they remain one of the largest error sources in remote sensing of BVOCs (Palmer et al. 2001; Millet et al. 2006)). Lam-sal et al. (2014) recommends that when comparing satellite data to models, the AMF should first be recalculated using the model as an a priori. This is in order to remove any a priori bias between model and satellite columns.

Related to the AMF is the averaging kernal (AK), which is used to handle instrument measurements which are sensitive to gas concentrations at different altitudes through the atmosphere. DOAS methods can be heavily influenced by the initial estimates of a trace gas profile (the a priori) which is often produced by modelling,

so when comparing models of these trace gases to satellite measurements extra care needs to be taken to avoid introducing bias from differing a priori assumptions. One way to remove these a priori influences is through the satellite AK (or AMF), which takes into account the vertical profile of the modelled trace gas and instrument sensitivity to the trace gas (Eskes and Boersma 2003; Palmer et al. 2001). This process is called deconvolution ( $\Omega = AK \times VC_{satellite} + \times(I - AK)VC_{apriori}$ ) of the AK of the satellite instrument. The AK represents sensitivities to each species at multiple altitudes through the atmosphere and in the case of OMI, can be approximated from the scattering weights ( $\omega(z)$ ) function as follows:

$$AK(z) = \frac{\omega(z)}{AMF} \quad (2.4)$$

This is an approximation for the OMI product, which does not include the AK but does include the  $\omega$  and AMF, as explained in Gonzalez Abad et al. (2015).

#### 2.3.4 LIDORT

LIDORT is a model of LInearized Discrete Ordinate Radiative Transfer, used to determine backscatter intensities and weighting functions at arbitrary elevation angles (Spurr, Kurosu, and Chance 2001). The model solves radiative transfer equations and can be used to determine various atmospheric column measurement attributes such as optical depth, ring effects, and scattering. These radiative properties (or at least estimates thereof) are required when measuring trace gases in the atmosphere through a long path such as seen by satellites (eg. Palmer et al. 2001; Martin et al. 2002a; De Smedt et al. 2015; Gonzalez Abad et al. 2015).

#### 2.3.5 Uncertainty

Uncertainty in a single pixel for OMHCHO is roughly the same magnitude as HCHO background levels. Each pixel has  $\sim 2 \times 10^{14}$  molec  $\text{cm}^{-2}$  uncertainty, which is  $5\times$  higher than GOME. However, there are  $\sim 100 - 200\times$  as many measurements allowing a greater reduction of uncertainty with averaging. This is due to the smaller footprint and better temporal resolution of OMI (Instrument 2002; Millet et al. 2008). The finer nadir resolution of OMI (13 by 24  $\text{km}^2$ ) compared to other satellites also reduces cloud influence (Millet et al. 2006; Millet et al. 2008). The top row in figure 2.8 shows OMI HCHO columns binned to at  $0.25^\circ$ longitude by  $0.3125^\circ$ latitude averaged over one day and one month (with and without filtering). Row two shows uncertainty of the satellite data after averaging. It is clear that one day of satellite data is too uncertain when binned at  $0.25 \times 0.3125^\circ$ horizontal resolution, however after a month (with or without filtering) the uncertainties become manageable. If we assume the uncertainty is random error, and not bias introduced through calculation techniques, then we are able to reduce the uncertainty through averaging. High resolution low detection limit estimates can be built up using “oversampling”, which averages satellite measurements over time (eg. Zhu et al. 2014). Uncertainty in satellite recalculations, along with other factors is analysed in section 3.5.

### OMI HCHO and uncertainty for 200501

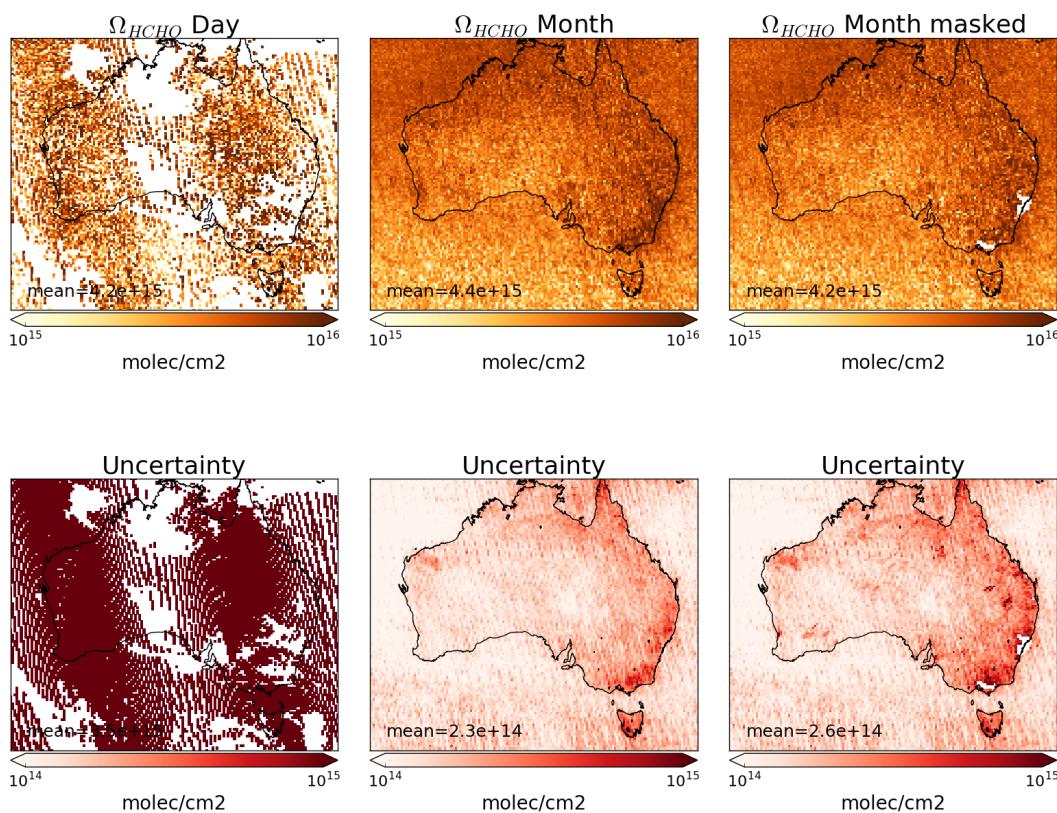


FIGURE 2.8: Top row shows  $0.25^{\circ}$  by  $0.3125^{\circ}$  binned OMHCHO columns with one day, one month, and one month with non-biogenic masking applied from left to right respectively. Bottom row shows the uncertainty for each gridsquare after averaging.

In this thesis, HCHO columns (pixels) with cloud fractions over 40% are filtered as done in Palmer et al. (2001), which introduces a clear-sky bias to any monthly averages. This is due to HCHO being lower on unrecorded cloudy days. This bias has been measured as a 13% positive monthly mean bias (Palmer et al. 2001; Surl, Palmer, and González Abad 2018).

Recently Schenkeveld et al. (2017) analysed the performance over time of the instrument and found irradiance degradation of 3-8%, changed radiances of 1-2%, and a stable wavelength calibration within 0.005-0.020 nm. These changes are measured excluding the row anomaly (RA) effect, which is relatively stable since 2011, although it is still growing and remains the most serious concern. Their analysis of OMI concludes that data is still of high quality and will deliver useful information for 5-10 more years, with radiance only changing by 1 – 2% outside of RA impacted areas. An analysis of the row anomaly by Huang et al. (2017) state that measurements remain suitable for scientific use, with recommendation for further evaluation. The RA began in June 2007, with some cross-track rows seemingly blocked. The most likely cause is some instrument insulation partially obscuring the radiance port (Schenkeveld et al. 2017).

In satellite HCHO products, concentrations over the remote pacific ocean are sometimes used to analyse faulty instrument readings. This is due to the expected invariance of HCHO over this region. For instance GOME (an instrument which measures trace gases on board the ERS-2) corrects for an instrument artefact using modelled HCHO over the remote pacific (Shim et al. 2005). OMI HCHO products use a similar technique to account for sensor plate drift and changing bromine sensitivity (Gonzalez Abad et al. 2015). Uncertainty in the OMI satellite instrument is calculated by the Smithsonian Astrophysical Observatory (SAO) group using the uncertainty in backscattered radiation retrievals (Gonzalez Abad et al. 2015; Abad et al. 2016). Another method of calculating the uncertainty is used by the Belgian Institute for Space Aeronomy (BIRA) group, who determine uncertainty from the standard deviation of HCHO over the remote pacific ocean (De Smedt et al. 2012; De Smedt et al. 2015).

For many places the tropospheric column HCHO measured by satellite is biased low, Zhu et al. (2016) examine six available datasets and show a bias of 20 - 51% over south east USA when compared against a campaign of aircraft observations (SEAC<sup>4</sup>RS). De Smedt et al. (2015) also found OMI and GOME2 observations were 20 - 40% lower than ground based vertical profiles, and Barkley et al. (2013) determine OMI to be 37% low compared with aircraft measurements over Guyana. These bias can be corrected by improving the assumed a priori HCHO profiles which are used to calculate the AMFs of the satellite columns. Millet et al. (2006) examine OMI HCHO columns over North America and determine overall uncertainty to be 40%, with most of this coming from cloud interference. Millet et al. (2008) shows that there also exists some latitude based bias, as well as a systematic offset between the OMI and GOME instruments. This does not appear to be due to the different overpass times of the two instruments.

AMF calculation often dominates the total uncertainty in satellite retrievals, especially in polluted regions (Lorente et al. 2017). In scenarios where the gas is enhanced in the lower troposphere, AMF calculation is the largest uncertainty in satellite measurements. In polluted environments the structural uncertainty is estimated at 42 %,

or 31 % over unpolluted environments (Lorente et al. 2017). Another impact often not included in uncertainty calculations is the structural uncertainty of retrieval methods. The structural uncertainty of AMF calculation approaches used by different retrieval groups comes from how the AMF is calculated, rather than uncertainty in the calculation components. The importance of a priori and ancillary data (such as surface albedo and cloud top height) sharply affects the structural uncertainty (Lorente et al. 2017).

## 2.4 GEOS-Chem

### 2.4.1 Overview



GEOS-Chem is a well supported global, Eulerian CTM with a state of the science chemical mechanism, with transport driven by meteorological input from the Goddard Earth Observing System (GEOS) of the NASA Global Modeling and Assimilation Office (GMAO). Chemistry, transport, and meteorology are simulated at 15 minute time steps within a global set of 3-D boxes. Emissions are either prescribed by inventories or modelled (eg. biogenic emissions are created using the Model of Emissions of Gases and Aerosols from Nature (MEGAN)).

GEOS-Chem simulates more than 100 chemical species from the earth's surface up to the edge of space (0.01 hPa) and can be used in combination with remote and in situ sensing data to give a verifiable estimate of atmospheric gases and aerosols. It was developed, and is maintained, by Harvard University staff as well as users and researchers worldwide. In this thesis I use version 10.01 of GEOS-Chem, which outputs up to 66 chemical species (tracers) in the standard run, at 2 by 2.5° horizontal resolution, with 47 levels up to the top of the atmosphere (TOA at 0.01 hPa).

Global CTMs are often run using one or several emission models (or the output from them) to determine boundary conditions. Some of the inventories used by GEOS-Chem are described here. Meteorological fields are taken from NASA's GEOS-5 dataset ( $0.5^\circ \times 0.666^\circ$ ) (Chen et al. 2009), which exists up to April 2013. GEOS-5 meteorological fields are used as the boundary conditions driving transport. Fire emissions come from the GFED4 product (Miglio, Randerson, and Van Der Werf 2013). Anthropogenic VOC emissions come from the EDGAR inventory, while biogenic VOC emissions are simulated using the MEGAN model (see section 2.4.5). MEGAN is used to determine biogenic emissions for our default GEOS-Chem simulation. The estimated biogenic VOC emissions are important for accurately simulating chemistry within models, as discussed in Section 1.1.2.

### 2.4.2 Running GEOS-Chem

#### 2.4.2.1 Installation and requirements



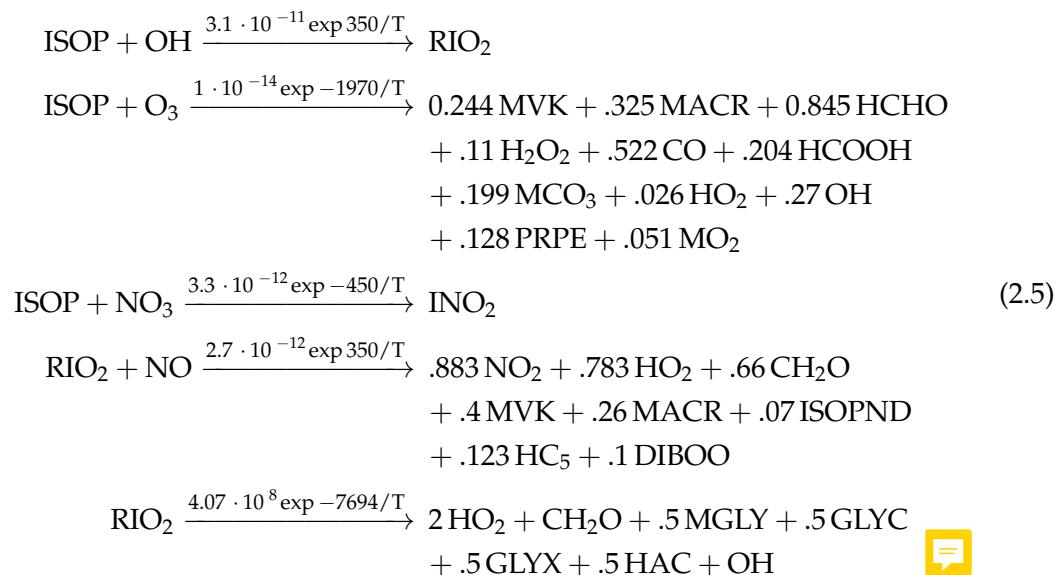
GEOS-Chem instructions for download, compilation, and running can be found in the user guide provided by Harvard: <http://acmg.seas.harvard.edu/geos/doc/man/>. In order to build and run GEOS-Chem a high-speed computing system is optimal, as globally gridded chemical calculations can take a long time to perform. I installed GEOS-Chem onto a suitably configured workspace on the National Computational

Infrastructure (NCI, <http://nci.org.au/>). This workspace included access to compilers and libraries which are needed to build the Fortran based GEOS-Chem source code, and IDL, Python, and various editors and scripting languages to read, run, edit, and analyse both GEOS-Chem and its output. After downloading GEOS-Chem, the code can be compiled with different options for resolution and chemical mechanisms.

### 2.4.3 Chemical Mechanism

Chemical reactions are turned into systems of differential equations (DEs) to be solved by the CPU for each gridbox in GEOS-Chem. A chemical mechanism is the name for a closed system of chemical reactions and the rates of each reaction. Simplifications are required due to the massive amount of reactions which occur in the atmosphere, and the coupled and stiff nature of these reactions which serve to slow down computation of the solutions (Brasseur and Jacob 2017). Stiffness in chemical systems of differential equations is due to the massively differing reaction rate time scales - for instance hydroxy radicals react within seconds while methane has an atmospheric lifetime of 8-10 years (Wuebbles2002).

Some of the important reactions involving isoprene are copied here, including reaction rates in the form  $k = A \exp(-ER/T)$ . The tracers have multiple names in order to allow the model to keep track of extra information such as the tracers source. For instance  $\text{INO}_2$  is  $\text{RO}_2$  specifically created by the  $\text{ISOP} + \text{NO}_3$  reactions. The full current mechanism is described online at [http://wiki.seas.harvard.edu/geos-chem/index.php/New\\_isoprene\\_scheme](http://wiki.seas.harvard.edu/geos-chem/index.php/New_isoprene_scheme).



In these reactions T is temperature,  $\text{INO}_2$  is  $\text{RO}_2$  formed from isoprene and  $\text{NO}_3$ , ISOPNB and ISOPND are  $\beta$  and  $\delta$  isoprene nitrates respectively.

#### 2.4.4 GEOS-Chem isoprene modelling

The isoprene reactions simulated by GEOS-Chem were originally based on Horowitz et al. (1998). This involved simulating NO<sub>x</sub>, O<sub>3</sub>, and NMHC chemistry in the troposphere at continental scale in three dimensions, with detailed NMHC chemistry with isoprene reactions and products. The mechanism was subsequently updated by Mao et al. (2013), who change the isoprene nitrates yields and add products based on current understanding as laid out in Paulot et al. (2009a) and Paulot et al. (2009b). Further mechanistic properties, like isomerisation rates, are based on results from four publications: (Peeters, Nguyen, and Vereecken 2009; Peeters and Muller 2010; Crounse et al. 2011; Crounse et al. 2012).

Crounse et al. (2011) examines the isomerisations associated with the oxidation of isoprene to six different isomers of ISOPOO formed in the presence of oxygen through  $ISOP + OH \rightarrow ISOPOO$ . They determine rates and uncertainties involved in these reactions, and study the rate of formation of C<sub>5</sub>-hydroperoxyenals (C5-HPALD) by isomerisation. Prior to 2012 oxidation chamber studies were performed in high NO or HO<sub>2</sub> concentrations, giving peroxy lifetimes of less than 0.1 s (Crounse et al. 2012; Wolfe et al. 2012). In most environments NO and HO<sub>2</sub> concentrations are not so high, GEOS-Chem uses production rates for different NO concentrations and peroxy radical lifetimes determined by Crounse et al. (2012). OH regeneration through photolysis of HPALDs in areas with high isoprene emissions are included from Peeters and Muller (2010). Photolysis of photolabile peroxy-acid-aldehydes creates OH and improved model agreement with continental observations. OH and HPALD interactions are central to maintaining the OH levels in pristine and moderately polluted environments, which makes isoprene both a source and a sink of OH (Peeters and Muller 2010; Taraborrelli et al. 2012).

Formation of isoprene nitrates (ISOPN) have an effect on ozone levels through NO<sub>x</sub> sequestration, and the yields and destinies of these nitrates is analysed in Paulot et al. (2009a). In a chamber with clean air and high NO concentrations, isoprene photooxidation is initially driven by OH addition, followed by NO<sub>x</sub> chemistry (150 min - 600 min), and finally HO<sub>x</sub> dominated chemistry. GEOS-Chem uses these the yields of various positional isomers of isoprene nitrates, and pathways of their oxidation products, and reactions within its suite of chemical mechanisms determined by Paulot et al. (2009a) and Mao et al. (2013).

GEOS-Chem models both high and low NO<sub>x</sub> scenarios using NO<sub>x</sub> dependent reactions as derived by Paulot et al. (2009a) and Mao et al. (2013). In low NO<sub>x</sub> ISOPOO reacts with HO<sub>2</sub> (70% yield of hydroxy hydroperoxides, ISOPOOH), RO<sub>2</sub> (producing mainly MACR, MVK, and HCHO), or isomerises (1,5-H shift producing MACR, MVK, HCHO, or 1,6-H shift producing HPALDs). ISOPOOH can be oxidised (by OH) to produce epoxydiols (IEPOX), recycling OH (Paulot et al. 2009b). HPALDs can photolyse to regenerate OH and small VOCs (Crounse et al. 2011; Wolfe et al. 2012; Jozef et al. 2014). Under low NO<sub>x</sub> conditions production of HCHO, MVK, and MACR is 4.7%, 7.3%, and 12% respectively. Refer to section 1.3.3 for more information. Under high NO<sub>x</sub> conditions, isoprene undergoes OH addition at the 1 and 4 positions, becoming  $\beta$  (71%) or  $\delta$  (29%) hydroxyl peroxy radicals (ISOPO<sub>2</sub>). The  $\beta$ -hydroxyl reacts with NO<sub>x</sub> and produces HCHO (66%), methylvinylketone (40%) (MVK), methacrolein (26%), and  $\beta$ -hydroxyl nitrates (6.7%) (ISOPNB). The  $\delta$ -hydroxyl reacts with NO to

form  $\delta$ -hydroxyl nitrates (24%) (ISOPND), and ISOPNB (6.7%). ISOPNB and ISOPND yield first generation isoprene at 4.7% and 7% respectively.

The isoprene mechanism in GEOS-Chem includes OH regeneration from oxidation of epoxydiols and slow isomerisation of ISOPO<sub>2</sub> (Mao et al. 2013). Older models isoprene produced ISOPOOH which then titrated OH, however, the loss of OH had not been seen in measurements (Paulot et al. 2009b; Mao et al. 2013). Mao et al. (2013) show that a lower (factor of 50) rate constant for ISOPO<sub>2</sub> isomerisation leads to better organic nitrate agreements with ICARTT. The chemistry updates have led to more accurate modelling of OH concentrations, especially in low NO<sub>x</sub> conditions common in remote forests. Prior to Mao et al. (2012), measurements of OH in high VOC regions may have been up to double the real atmospheric OH levels, due to formation of OH inside the instrument. The updates to isoprene chemistry by Mao et al. (2013), and those shown in Crounse et al. (2011) and Crounse et al. (2012) are the last before version 11.

#### 2.4.5 Emissions from MEGAN

MEGAN is a global model with resolution of around 1 km, and is used to generate the BVOC emissions used in various global chemistry models such as GEOS-Chem. MEGAN uses leaf area index, global meteorological data, and plant functional types (PFTs) to simulate terrestrial isoprene emissions. The model includes global measurements of leaf area index, plant functional type, and photosynthetic photon flux density, from remote sensing databases (Kefauver, Filella, and Peñuelas 2014). The various PFTs are used to generate emission factors which represent quantities of a compound released to the atmosphere through an associated activity. For example, the emission factor for isoprene is tied to available sunshine, suitable temperature, etc. The schematic for MEGAN, taken from Guenther (2016), is shown in figure 2.9.

GEOS-Chem V10.01 uses MEGAN V2.1 with biogenic emissions from Guenther et al. (2012). It computes some emissions using predefined EFs from MEGAN source code, and others using PFT maps and associated EFs. MEGAN “is a modelling framework for estimating fluxes of biogenic compounds between terrestrial ecosystems and the atmosphere to account for the major known processes controlling biogenic emissions” (Guenther et al. 2012). It allows parameterisation of various BVOC emissions, with descriptions given in Guenther et al. (2012).

MEGAN was developed as a replacement for two earlier canopy-environment emission models (BIES and GEIA), and initially included a simple canopy radiative transfer model, which parameterised sun-lit and shaded conditions through a canopy. Early models did not account for abiotic stresses, such as drought, prior rainfall and development processes. These stresses influenced species-specific emissions by more than an order of magnitude (Niinemets et al. 1999). Isoprene emissions were based on temperature, leaf area, and light, but have since been updated to include leaf age activity (Guenther et al. 2000), and a leaf energy balance model (Guenther et al. 2006) in MEGANv2.0. This update included a parameter for soil moisture, to account for drought conditions, however this parameter is currently (as of version 2.1) not applied to isoprene (Sindelarova et al. 2014). Soil moisture effects on isoprene emission are important, and can affect estimates.

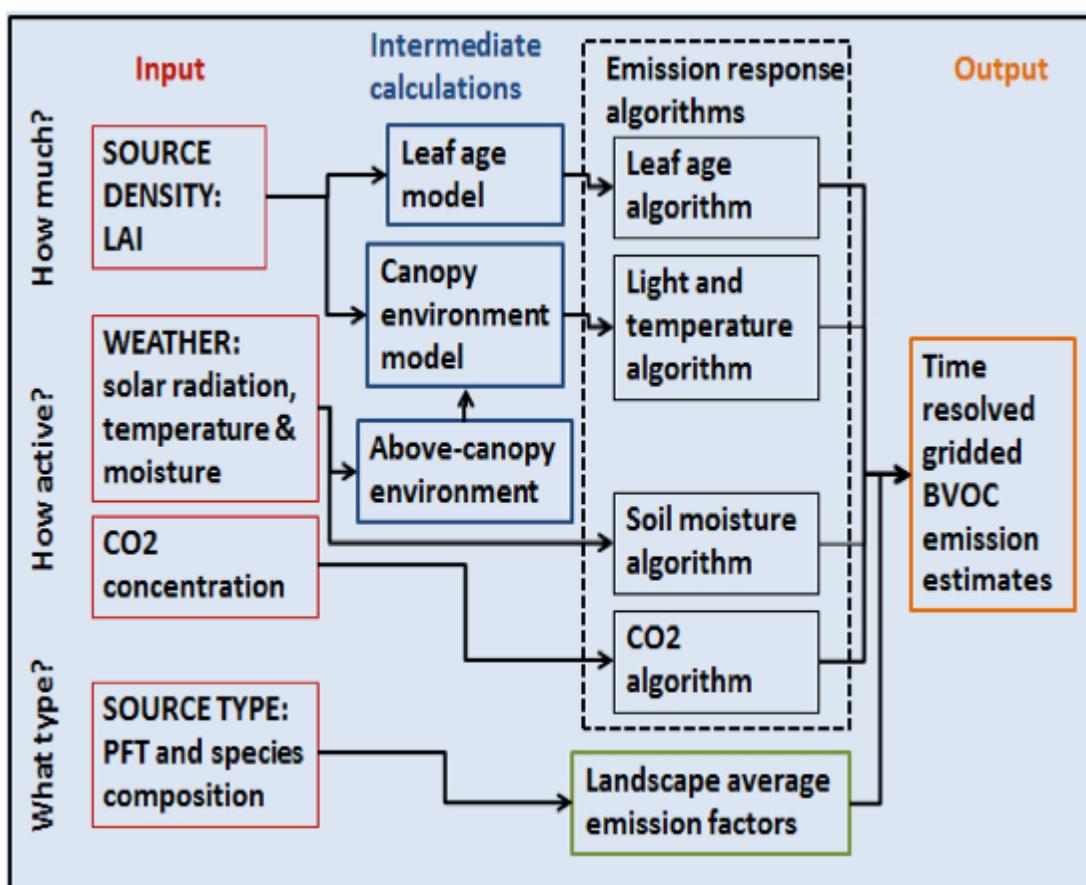


FIGURE 2.9: MEGAN schematic, copied from Guenther (2016)

Instructions to run version 2.1 are available at [http://lar.wsu.edu/megan/docs/MEGAN2.1\\_User\\_GuideWSU.pdf](http://lar.wsu.edu/megan/docs/MEGAN2.1_User_GuideWSU.pdf), and a version using the Community Land Model (CLM) is available at <http://www.cesm.ucar.edu>. It uses meteorological fields from the Weather Research and Forecasting (WRF) modelling system. Version 2.1 (updated from 2.0 (Guenther et al. 2006)) includes 147 species, in 19 BVOC classes, which can be lumped together to provide appropriate output for mechanisms in various chemical models.

#### 2.4.6 Rescaling NO<sub>x</sub>

NO<sub>x</sub> concentrations affect atmospheric oxidative capacity, which changes many factors important in estimating isoprene emissions including isoprene to HCHO yield, isoprene lifetime, and isoprene oxidation pathways. This means that if the model is poorly simulating NO<sub>x</sub>, yields and the effects of transport may be poorly estimated. In order to determine if re-scaling the NO emissions over Australia is necessary in GEOS-Chem, modelled NO<sub>2</sub> amounts are compared to satellite data for most of

Simulated GEOS-Chem tropospheric NO<sub>2</sub> columns averaged from 1300-1400 LT are compared against OMNO2d data. Figure 2.10 shows the direct comparison between these datasets averaged over the months of January and February, 2005. The top row shows (from left to right) GEOS-Chem NO<sub>2</sub>, OMI NO<sub>2</sub> at 0.25x0.3125°, and OMI NO<sub>2</sub> at 2x2.5°. The bottom row shows the difference (absolute, and relative) between GEOS-Chem and OMI, as well as the RMA and correlation. The OMNO2d product shows Sydney and Melbourne as NO<sub>2</sub> hotspots, which are underestimated by GEOS-Chem due to averaging over the 2x2.5°horizontal resolution. Over much of the country GEOS-Chem overestimates NO<sub>2</sub> by 10-60%, except in NA and northern Queensland where up to 50% underestimation occurs. The comparison is repeated for winter (JJA) of 2005 in Figure 2.11.

This comparison is expanded, including against modelled emissions, and repeated for autumn (MAM), winter (JJA), and spring (SON) in figures 2.12 to 2.19. These show an analysis of GEOS-Chem NO emissions and their correlations with the bias between GEOS-Chem NO<sub>2</sub> mid-day columns and the OMNO2d product, averaged over each season in 2005. The scatter plots have one datapoint for each land square over Australia. The correlation between the bias (GEOS-Chem - OMNO2d) with anthropogenic or soil emissions ( $E_{NO}$ ) is shown in the bottom right of these figures. The correlation between model and satellite NO<sub>2</sub> columns is OK throughout the year over Australia, with some overestimation in the north during non-summer months. There is also slight underestimation over Sydney and Melbourne throughout the year. These figures show the visible biases are not driven by modelled emissions of NO. While the correlation between column NO<sub>2</sub> and emitted NO is clear, emissions do not appear to bias the model in either direction away from the satellite data. The poor correlations for anthropogenic NO suggest that blanket alterations over Australia would not lead to improved NO<sub>2</sub> fit.

The conclusion drawn is that modelled anthropogenic and soil NO emissions do not show sufficient evidence of biasing GEOS-Chem NO<sub>2</sub> columns away from satellite measurements over Australia. For this reason modelled NO emissions are not scaled in model runs in this thesis.

## GC NO vs OMNO2d 20050101-20050228

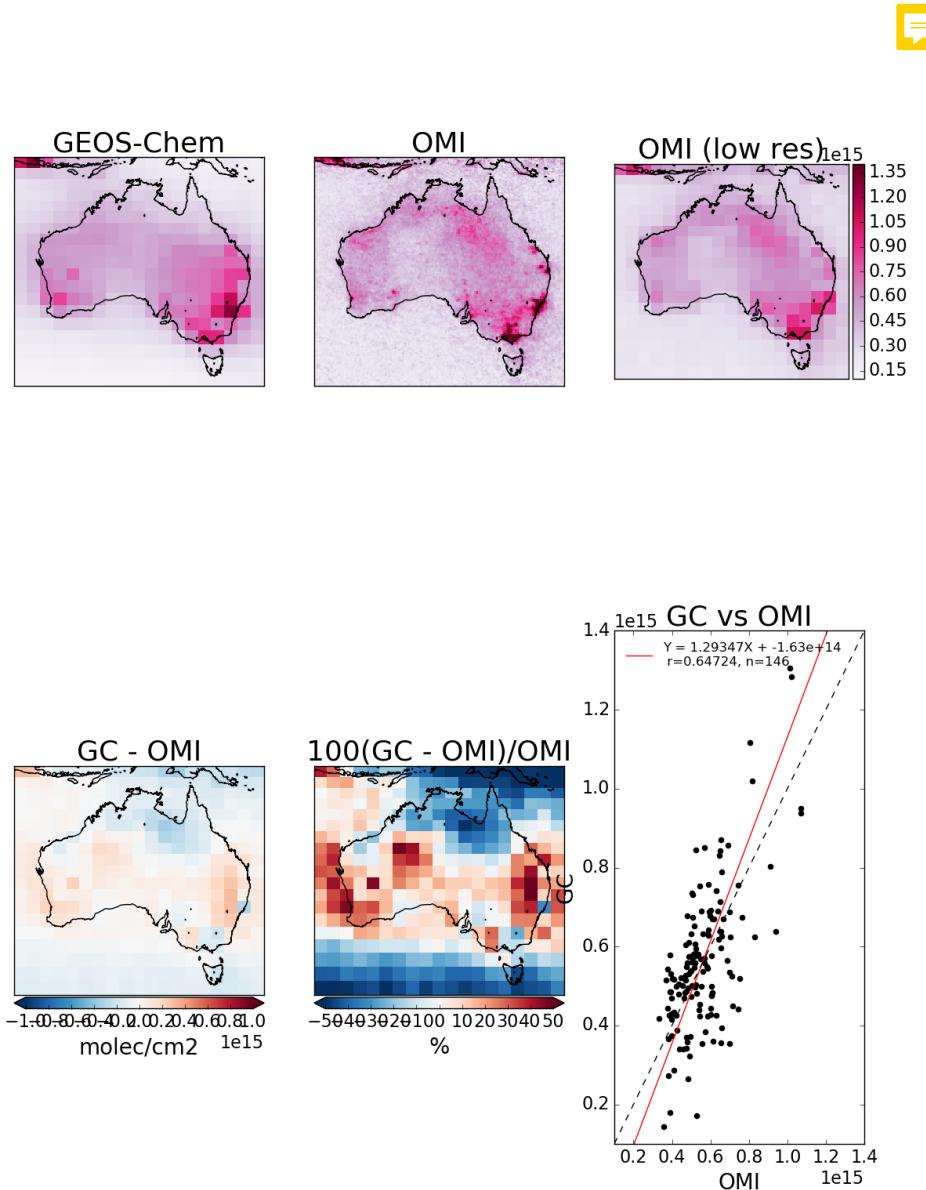


 FIGURE 2.10: Row 1 shows the tropospheric columns in molec  $\text{cm}^{-2}$ , GEOS-Chem, OMNO2d, and OMNO2d averaged onto the lower resolution of GEOS-Chem from left to right. Row 2 shows the correlations of GEOS-Chem (X axes) between daily anthropogenic emissions, and mid-day OMNO2d columns. Row 3 shows the differences with OMNO2d columns averaged into the lower resolution of GEOS-Chem.

## GC NO vs OMINO2d 20050601-20050831

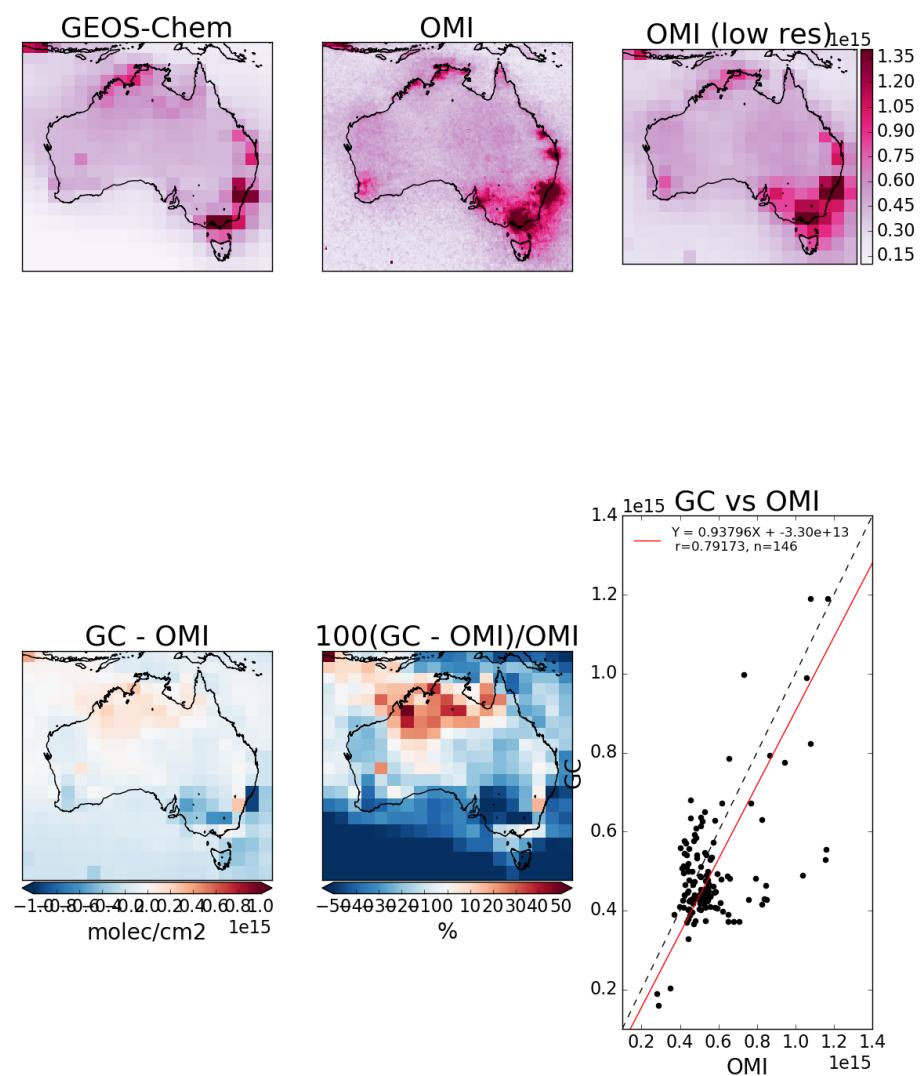


FIGURE 2.11: As figure 2.10, for winter 2005.

### GEOS-Chem vs OMNO2d Jan-Feb 2005

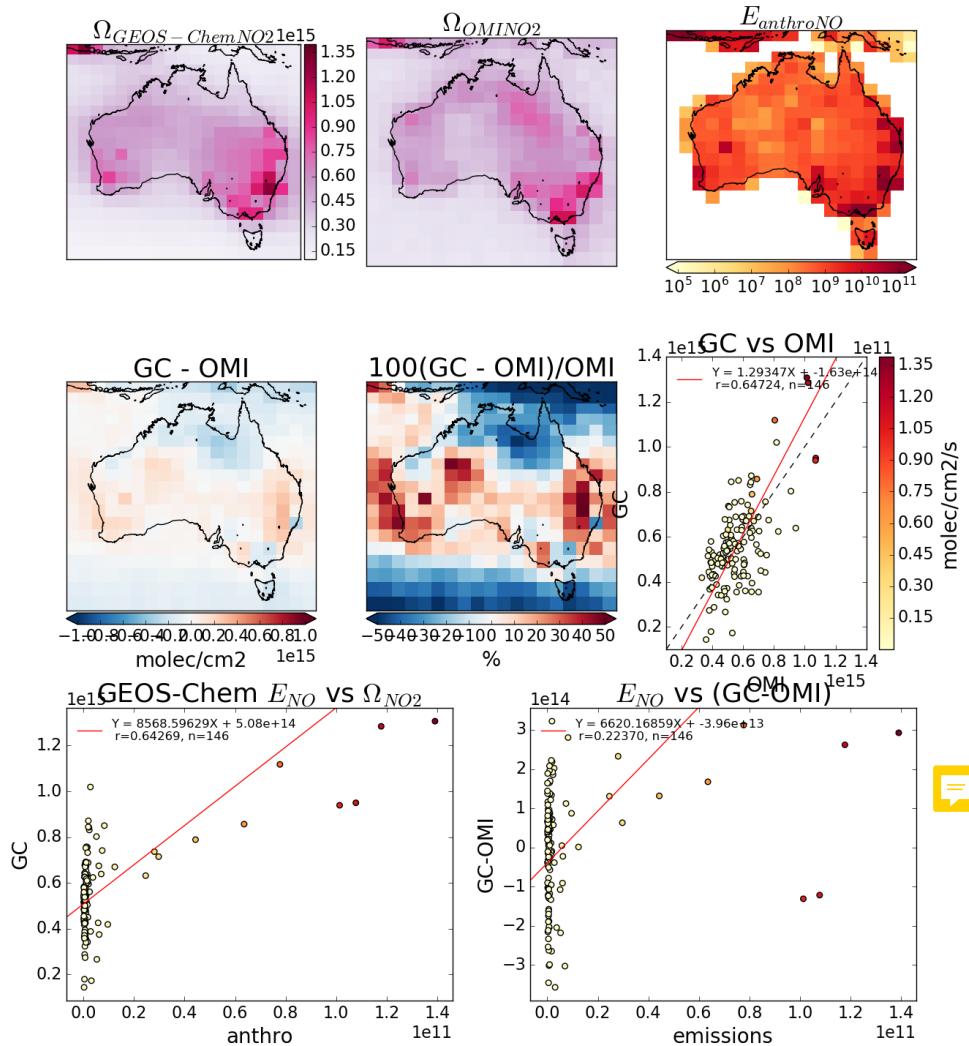


FIGURE 2.12: Top row (left to right): GEOS-Chem NO<sub>2</sub> mid-day tropospheric columns, OMNO2d NO<sub>2</sub> columns, modelled anthropogenic NO emissions. Second row: absolute and relative difference between GEOS-Chem and OMI NO<sub>2</sub> data, and the correlation. Third row: correlation between GEOS-Chem tropospheric column NO<sub>2</sub> and emitted NO, then between the model-satellite bias and the emissions. All correlation plots are coloured by NO emission rates.

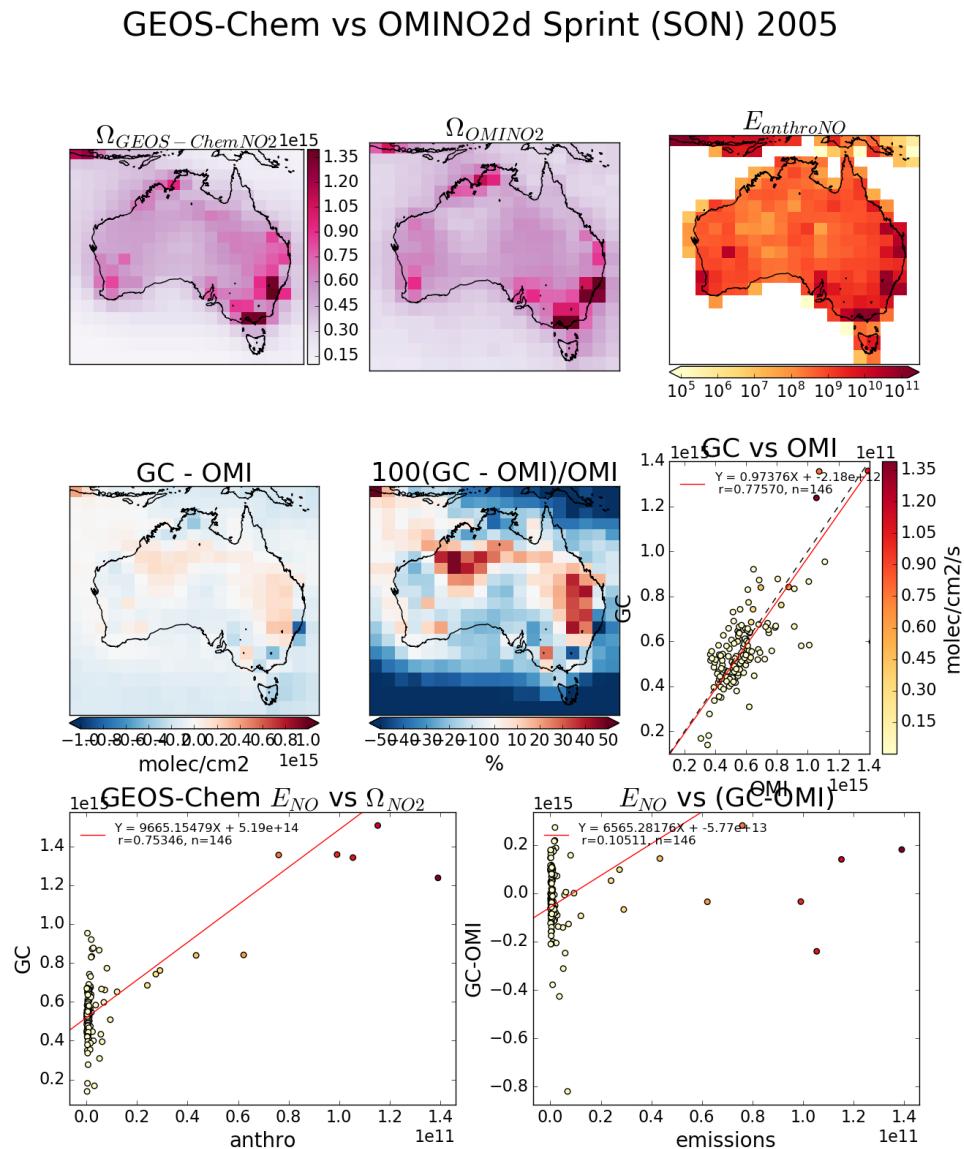


FIGURE 2.13: As figure 2.12, for Autumn 2005.

### GEOS-Chem vs OMINO2d Winter (JJA) 2005

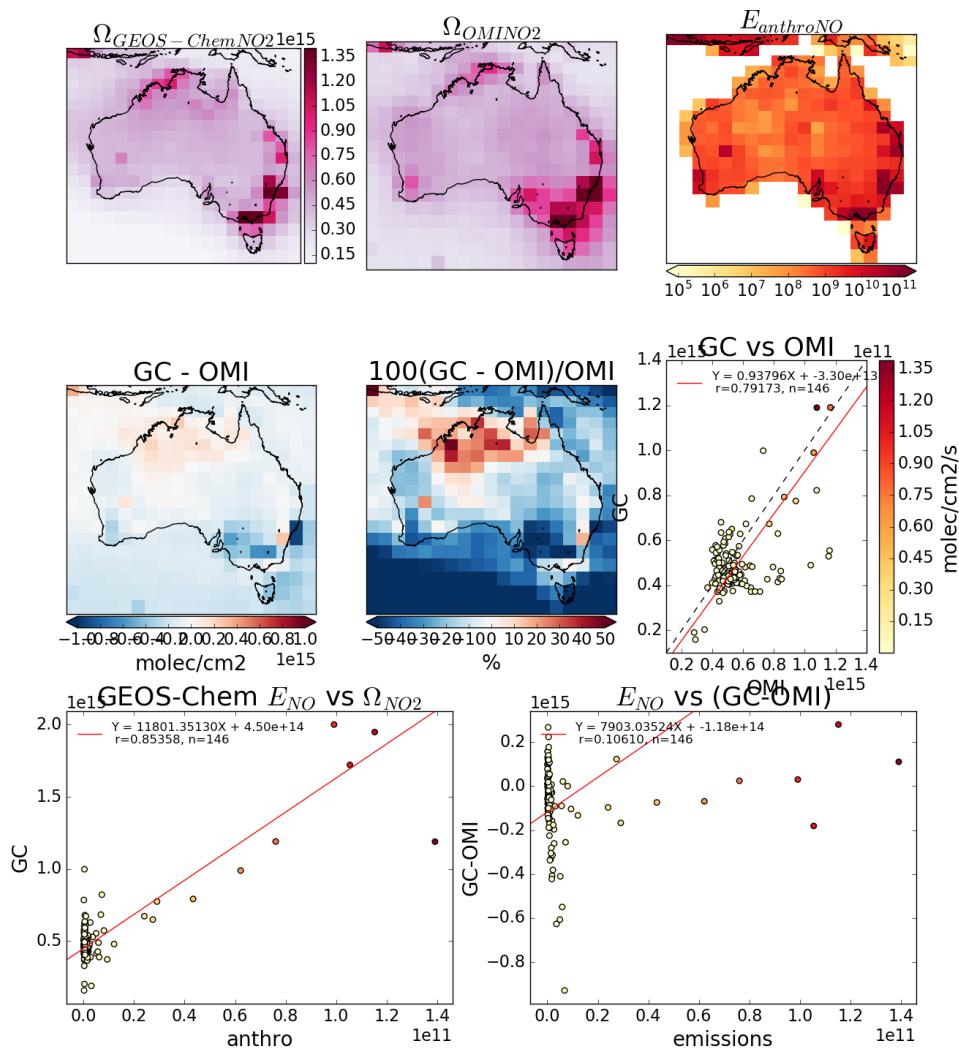


FIGURE 2.14: As figure 2.12, for Winter 2005.

GEOS-Chem vs OMINO2d Autumn (MAM) 2005

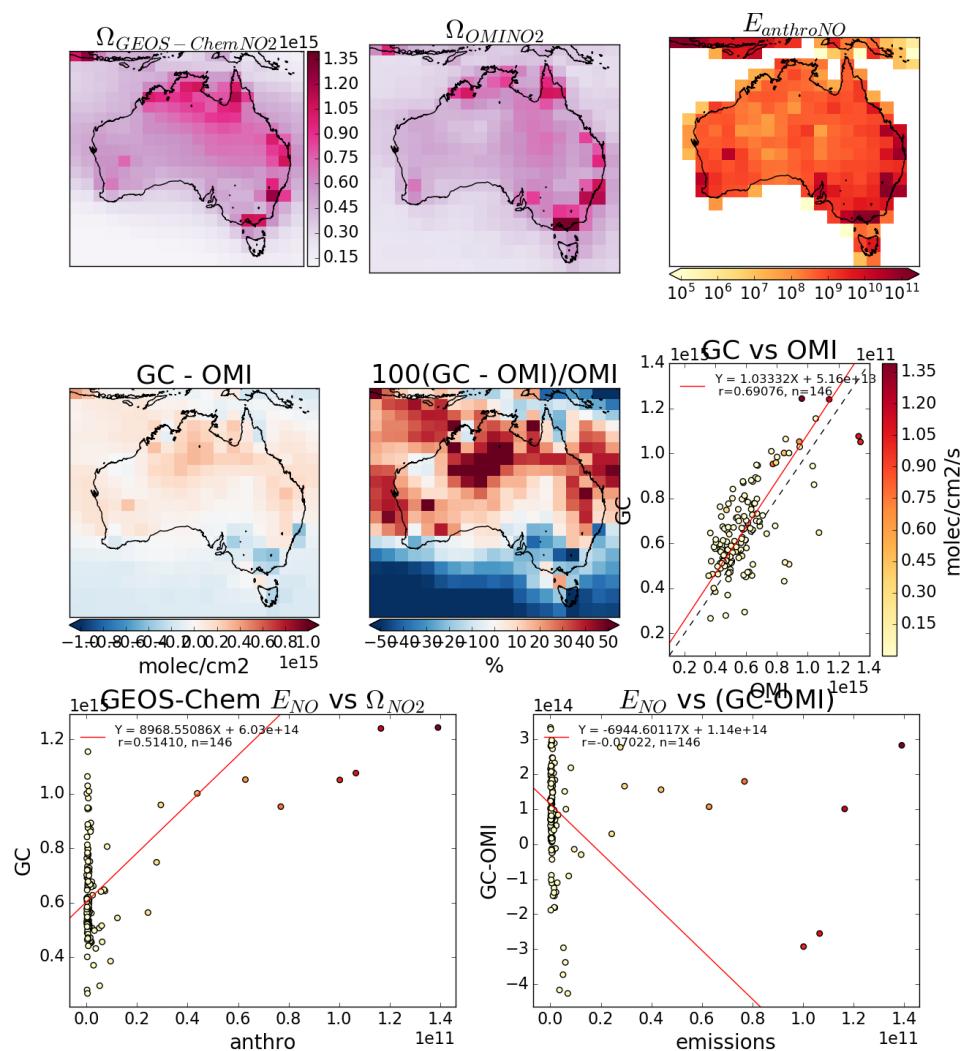


FIGURE 2.15: As figure 2.12, for Spring 2005.

### GEOS-Chem vs OMINO2d 20050101-20050228

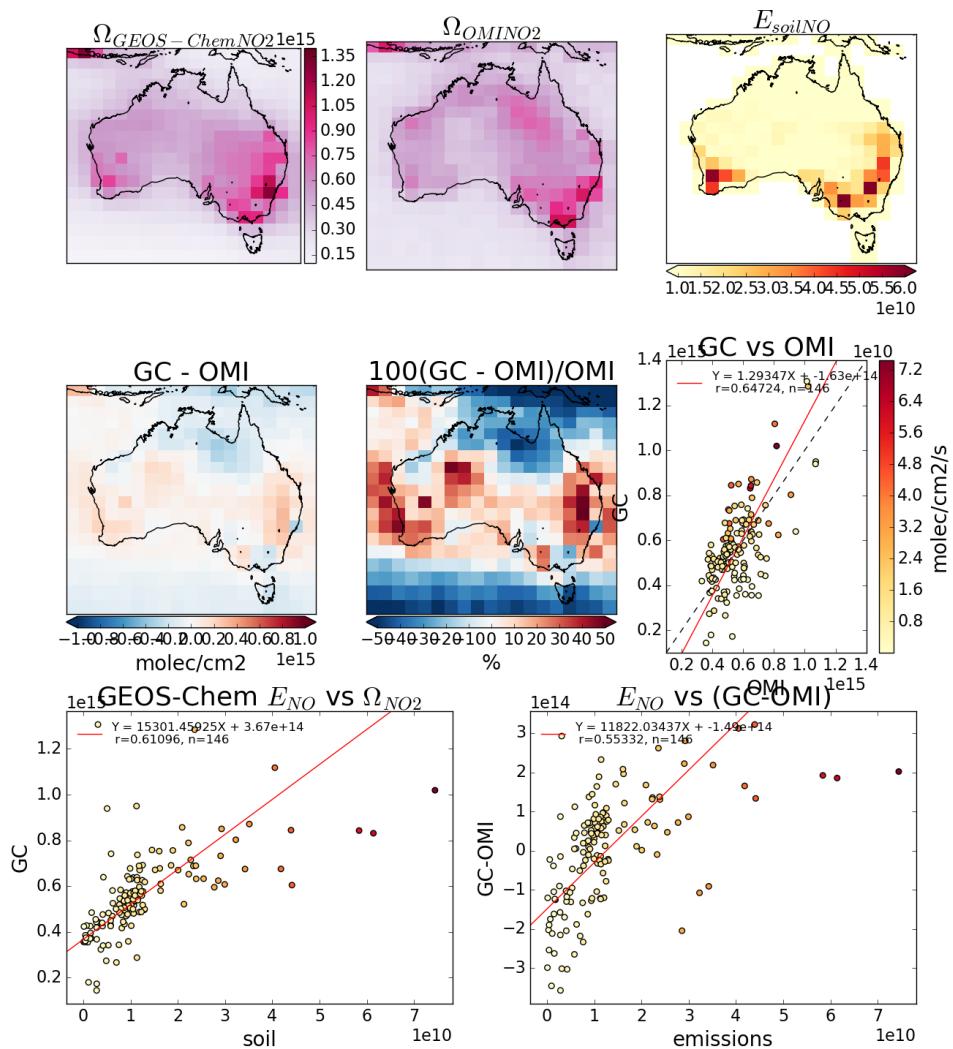


FIGURE 2.16: As figure 2.12, except anthropogenic NO emissions are replaced by soil NO emissions.

## GEOS-Chem vs OMINO2d 20050301-20050531

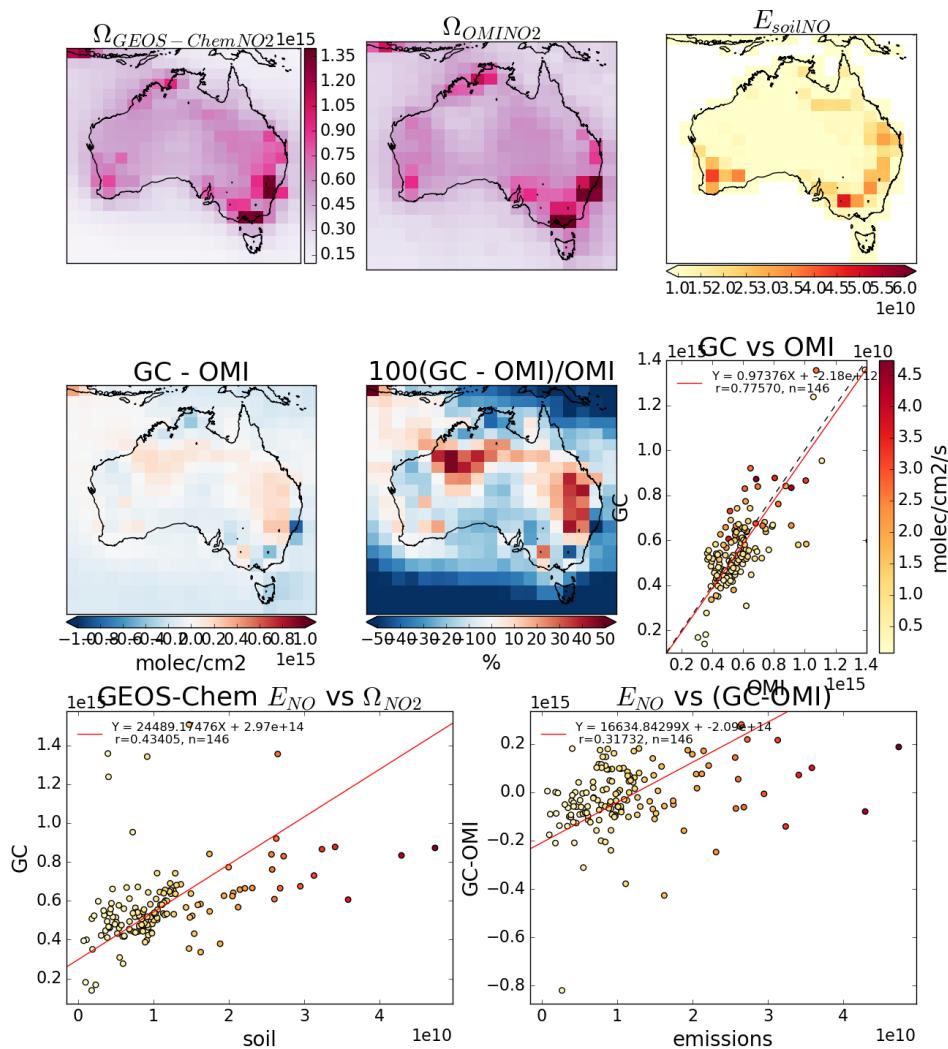


FIGURE 2.17: As figure 2.12, for Autumn 2005, with soil NO emissions replacing anthropogenic NO emissions.

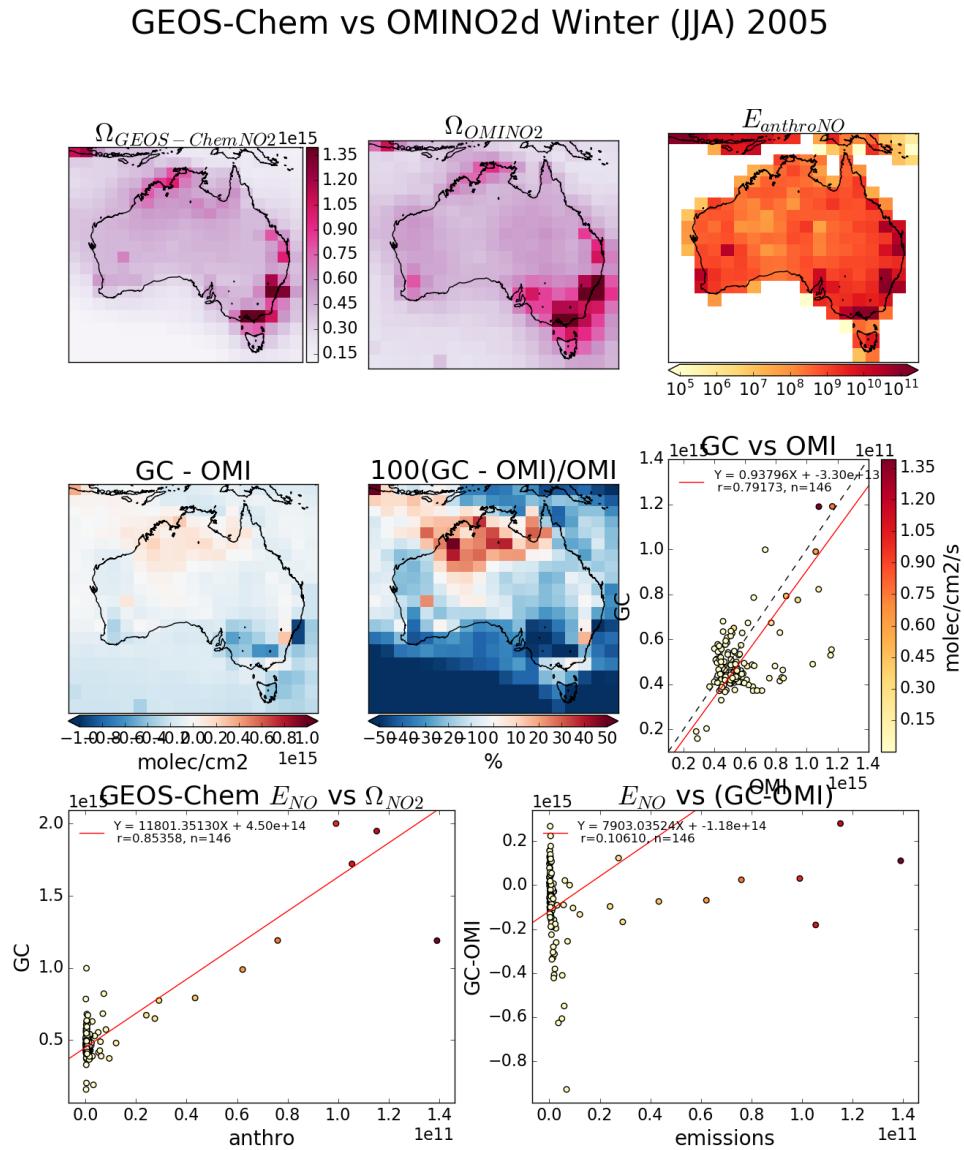


FIGURE 2.18: As figure 2.12, for Winter 2005, with soil NO emissions replacing anthropogenic NO emissions.

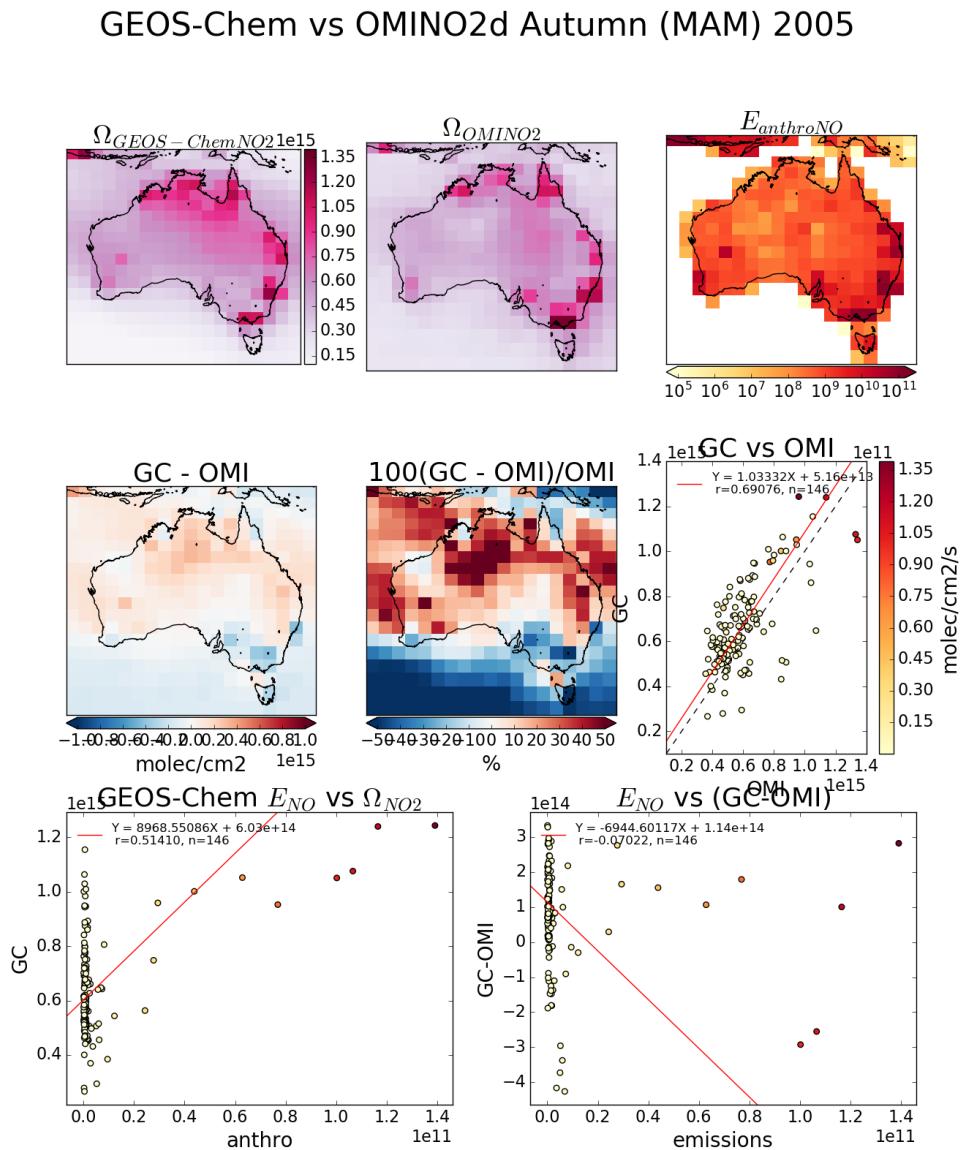


FIGURE 2.19: As figure 2.12, for Spring 2005, with soil NO emissions replacing anthropogenic NO emissions.

### 2.4.7 GEOS-Chem simulations

GEOS-Chem is run four independent times in this thesis, with different outputs from each simulation used to determine specific information. In this section the different output types are first described (section 2.4.7.1). Following this is the list of model runs, including a summary of the run, which outputs are created, and a summary of how they are used (section 2.4.7.2). Finally a brief comparison between a subset of the runs is performed (section 2.4.7.3).

#### 2.4.7.1 GEOS-Chem outputs

There are various outputs available when running GEOS-Chem, some of which require manipulation in order to compare against observations. GEOS-Chem in this thesis is run with a 15 minute time step for both chemistry and transport, at 2x2.5°horizontal resolution over 47 vertical levels. Output is the average of these time steps either over an entire month, or else per day, or else for the daily local time window of 1300–1400. For example: in this thesis, estimation of model isoprene to HCHO yields uses daily averaged HCHO columns compared against colocated isoprene emissions from MECAN. Optionally one can save high temporally resolved data for a single (or list of) column(s). This has been used here to compare modelled ozone with ozonesonde profiles at three sonde release sites discussed in Chapter 4.

**Satellite output** is output from averaging over a window of local time for each grid-box. Output averaged between 1300–1400 LT is saved to match with Aura satellite measurements, as Aura overpasses at ~1330 LT each day. Satellite output is saved both for comparison with, and recalculation of, satellite measurements.

**HEMCO diagnostics** are output by the HEMCO module, which deals with emissions inventories used in GEOS-Chem. When working with globally gridded data, handling local time offsets becomes more important. The hourly averaged output emissions of isoprene is saved using GMT, which needs to be offset based on longitude in order to retrieve local time. This is done by setting up a latitude by longitude array which matches the horizontal resolution of the data, filling each gridbox with its local time offset. This offset is determined as one hour per 15°(as 360°is 24 hours), and then used to retrieve global data at any specific local time. The retrieval of a daily local time global array is done by index matching the GMT+LT (modulo 24) with the desired hour on this grid over the 24 GMT hours.

**Tracer averages** are daily or monthly averaged gridbox concentrations.

**Time series** list of vertical columns with diagnostics saved at a temporal resolution of up to 15 minutes.

#### 2.4.7.2 GEOS-Chem runs

The following list summarises each model run as well as enumerating the outputs (described above), and how the run is used in the thesis.

**UCX** Universal tropospheric-stratospheric Chemistry eXtension (UCX) mechanism is run with 72 vertical levels from the surface to the top of the atmosphere (TOA $\sim$ 0.1 hPa). UCX runs a chemistry mechanism with combined tropospheric and stratospheric reactions, with an increased number of stratospheric calculations performed online (Eastham, Weisenstein, and Barrett 2014).

1. Satellite output, daily tracer averages, time series over three stations
2. Satellite output used to check how shape factors for AMF recalculation are affected by vertical resolution and stratospheric chemistry TODO: Make a plot of AMF with and without 72 vertical levels
3. Check if the stratospheric chemistry has much influence over tropospheric isoprene or HCHO concentrations.
4. daily tracer averages are used for ozone intrusion quantification (section 4.5.1), and ozone seasonality (section 4.4).
5. Time series are compared against ozonesonde releases (section 4.4) both over time and vertically.

**Tropchem (standard)** default settings for GEOS-Chem 10.01, using 47 vertical levels and  $2.5^{\circ}$  horizontal resolution. Additional satellite output is created to allow pp code to run on OMI satellite measurements (section 2.6.5).

1. Satellite output, daily tracer averages, HEMCO diagnostics
2. Used in recalculation of the satellite AMF (section 2.5), and the modelled background HCHO over the remote pacific which is used in the reference sector correction for OMI column retrievals (section 2.6.7)
3. Also used to determine isoprene to HCHO yield, after removing days with high biomass burning emissions (section 3.2.3)
4. Additional satellite outputs used to determine a separate AMF.
5. Satellite output is combined from two different runs in order to determine smearing (section 3.2.5) 
6. HEMCO diagnostics output used to determine isoprene to HCHO yields (section ??), and to compare against top-down estimations of isoprene emissions (section 3.3.2)
7. TODO: Compare total yearly ozone concentrations before and after implementing top-down estimate.

**Tropchem (isoprene emissions halved)** identical to standard tropchem except isoprene emissions are halved.

1. Satellite output, monthly tracer averages
2. Check modelled ozone sensitivity to isoprene emissions TODO: plot ozone between the two runs and reference here
3. Combined with standard run to determine model sensitivity transport (section 3.2.5)

**Tropchem (biogenic emissions only)** identical to standard tropchem except all non-biogenic emissions inventories are disabled.

1. Satellite output, hourly biogenic emissions from MEGAN
2. Used to determine the biogenic yield from isoprene to HCHO over Australia, described in section ??
3. Compared to run with updated emissions (3.3)

**Tropchem (altered MEGAN scaling factor)** . Satellite output, time series, daily averaged tracers

2. Compare to campaign datasets after altering isoprene emissions (see chapter 3)
3. Compare against satellite column HCHO as a sanity check on improving isoprene emissions

#### 2.4.7.3 UCX vs tropchem

Here we compare the model output with and without enabling the Universal tropospheric-stratospheric Chemistry eXtension (UCX). Both runs use  $2^{\circ}$  latitude by  $2.5^{\circ}$  longitude, however the UCX mechanism is run with 72 vertical levels from the surface to the top of the atmosphere (TOA~0.1 hPa), while the standard (tropchem) run uses 47 levels. The extra vertical levels are added in the stratosphere, providing finer vertical resolution from around 70 hPa to the top of the atmosphere. For both runs the input parameters such as MEGAN emissions and GEOS-5 meteorological fields are identical.

Figure 2.20 shows an example of surface HCHO amounts, averaged over Jan and Feb, 2007, with (A) and without (B) UCX turned on. Surface HCHO (first model level; up to  $\sim 100$  m) does not differ much between runs. differences do not exceed 3% over Australia, and absolute differences are exceedingly minor (note the scale in A-B). The major notable difference occurs over northern Africa, where HCHO is around 20% lower in the UCX run. Additionally a slight ( $<-8\%$ ) decrease in  $\Omega_{HCHO}$  over the oceans can be seen. The comparison is repeated using total columns (instead of surface values) in figure 2.21, showing that differences affecting HCHO between the model run are spread over the entire vertical column.

Figure 2.23 shows the differences in surface isoprene concentrations over Australia, averaged for 1, Jan to 28, Feb, 2007. Here we start to see a higher relative difference in concentrations, although this is generally over the areas with less absolute concentrations. Very little isoprene is seen away from the continents (4-5 orders of magnitude less), due to its short lifetime and lack of oceanic sources. Generally isoprene is 0-30% higher over mid to western Australia when the UCX mechanism is turned on, however this increase is lower in the regions with high isoprene emissions (north-east to south-east coastline). This enhancement can be seen throughout the entire tropospheric column as shown by Figure 2.23. There is a greater effect in Africa and South America in the tropics, with high relative differences in many regions with low absolute amounts.

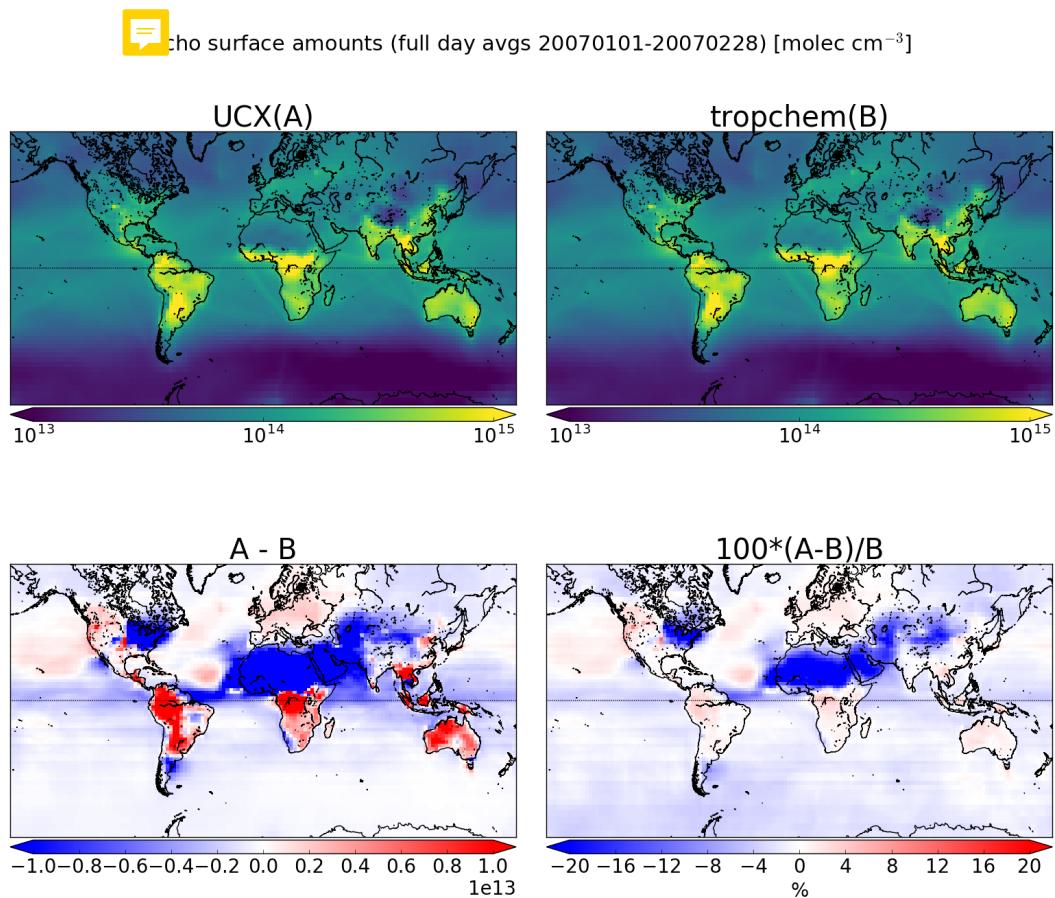


FIGURE 2.20: Surface HCHO simulated by GEOS-Chem with UCX (top left), and without UCX (top right), along with their absolute and relative differences (bottom left, right respectively). Amounts are the average of all times between 1, Jan and 28, Feb, 2007.

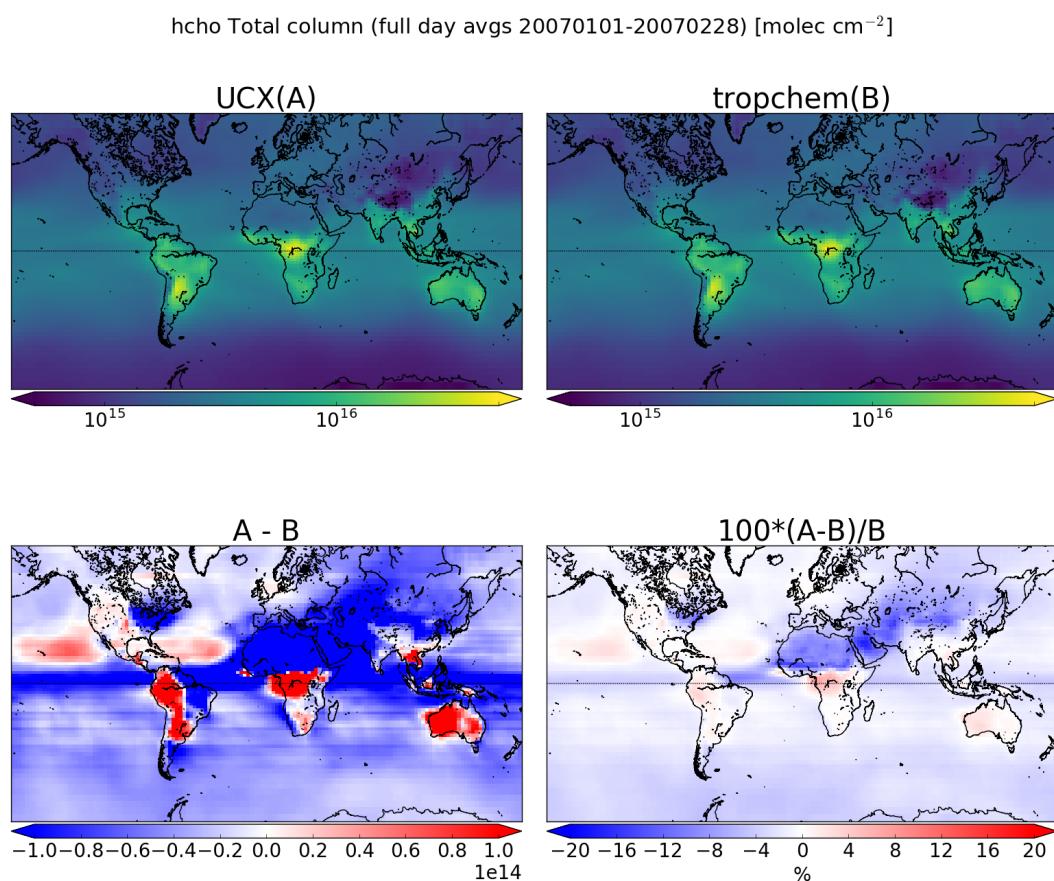


FIGURE 2.21: As figure 2.20 except using total column amounts instead of surface concentrations.

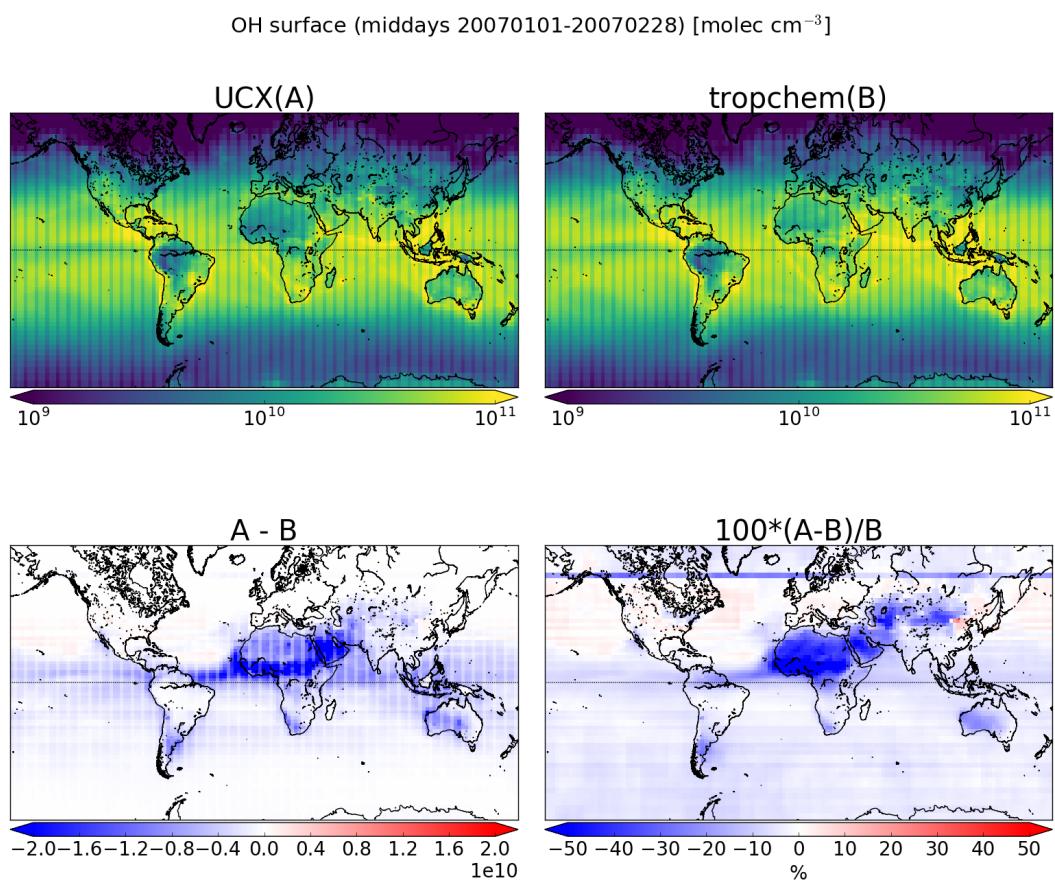


FIGURE 2.22: Midday (1300-1400 LT) surface OH concentrations averaged over Jan-Feb, 2007. Absolute and relative differences are shown on the bottom row.

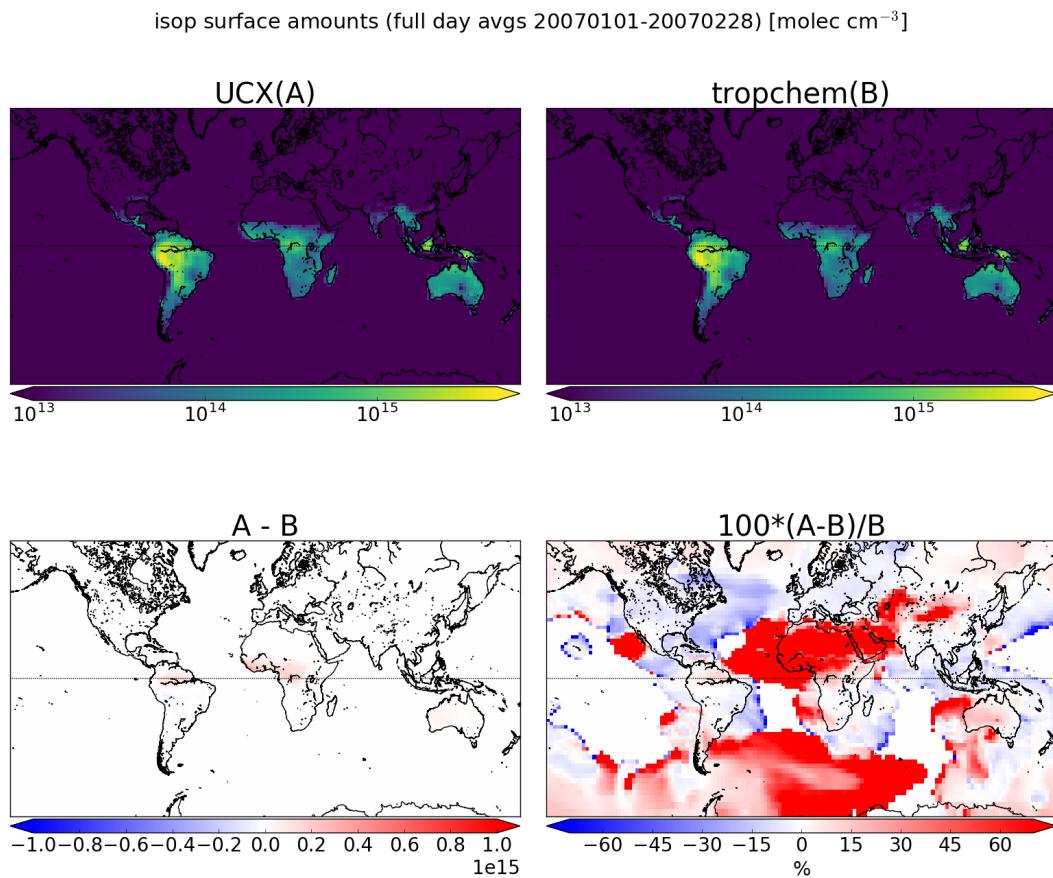


FIGURE 2.23: As figure 2.20, except looking at isoprene.

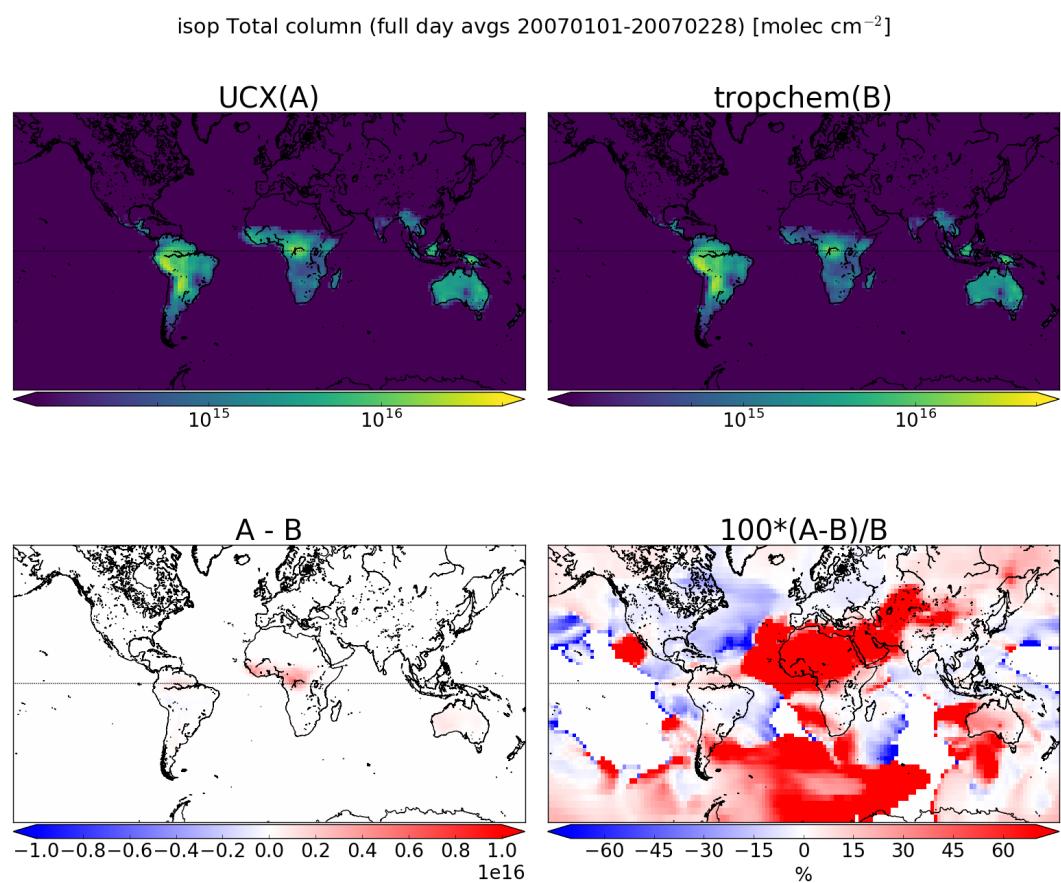


FIGURE 2.24: As figure 2.23, except looking at isoprene total column amounts.

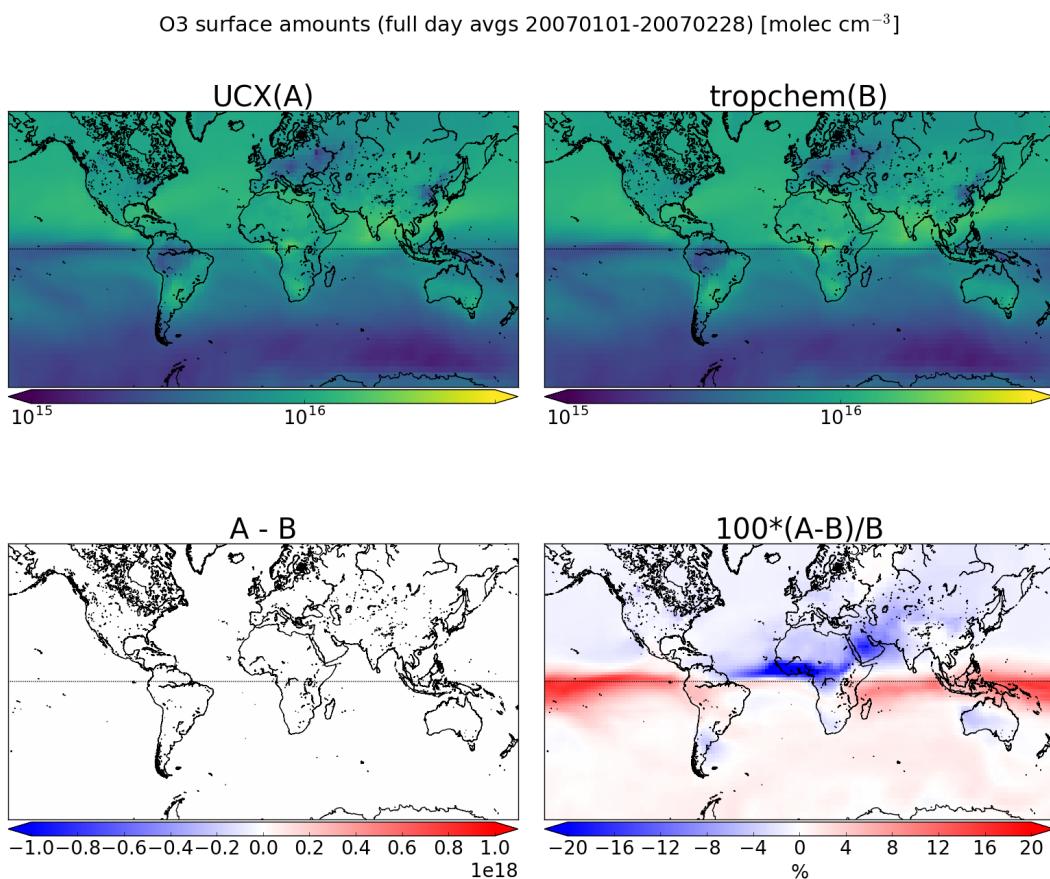


FIGURE 2.25: As figure 2.20, except looking at ozone.



The difference in isoprene between UCX and tropchem is likely caused by differences in the modelled radiation reaching the troposphere due to differences in simulated ozone in the stratosphere. With higher stratospheric ozone levels, less radiation would reach the troposphere, changing the photochemistry. Figure 2.22 shows how OH at midday is changed between runs. This figure shows a marked striping which is due to how OH is handled in GEOS-Chem. The most notable difference is again over northern Africa into eastern Europe, with some small decrease in the UCX OH over Australia everywhere except the east coast. Figure ?? shows the total column ozone between UCX and non-UCX runs, using full day averages. This shows that UCX ozone is lower everywhere except for a thin band just north and south of the equator. Decreased O<sub>3</sub> could lead to lower OH and the other differences between the runs over Australia.

The differences are minimal compared to other uncertainties in both AMF calculation and emissions estimation.



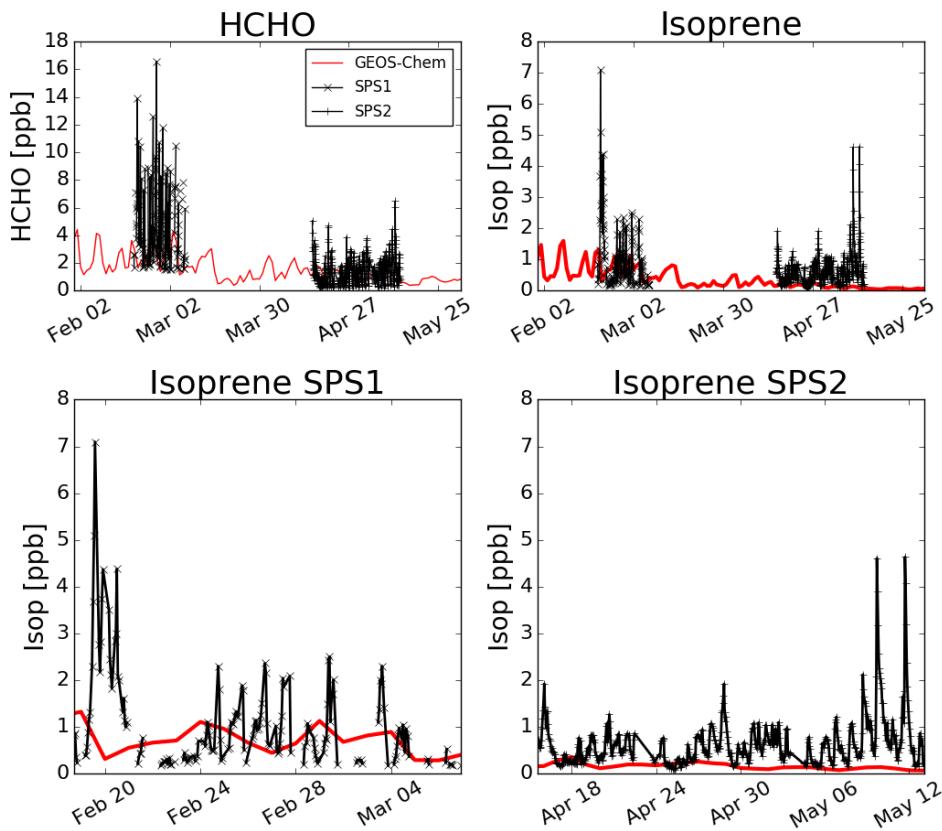


FIGURE 2.26: Comparison between GEOS-Chem HCHO concentrations in the gridsquare containing Sydney for the duration of the SPS 1 and 2 campaigns

#### 2.4.8 Comparisons with campaign data



Figure 2.26 shows GEOS-Chem output in the gridsquare containing Sydney overlaid on SPS measurement data. Superficially the comparison is not too bad between these two datasets, however GEOS-Chem output is the daily average over  $2 \times 2.5^\circ$  (latitude by longitude). 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.

## 2.5 Calculating an AMF

The AMF is the ratio of the slant column ( $\Omega_s$ ) to the vertical column ( $\Omega_v$ )

$$AMF = \frac{\Omega_s}{\Omega_v} = \frac{\tau_s}{\tau_v} \quad (2.6)$$

with  $\tau$  being the optical depth or thickness of the absorber through the slant ( $s$ ) or vertical ( $v$ ) path of light.

The OMI instrument records spectra of light which enters the viewing lens on board the Aura satellite. The spectra provide backscattered intensity ( $I_B$ ) at various wavelengths (see section ??), with the light source ( $I_{B_0}$ ) being the sun. Using the log of Beers law (equation 2.1) we get

$$\tau_s = \ln I_{B_0} - \ln I_B$$

which can be subbed into equation 2.6 to give an expression for the AMF which includes scattering:

$$AMF = \frac{\ln I_{B_0} - \ln I_B}{\tau_v} \quad (2.7)$$

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

$$AMF = \frac{\ln \left( 1 - \frac{\nabla I}{I_B} \right)}{\tau_v} \approx -\frac{\nabla I}{I_B \tau_v} \quad (2.8)$$

This is due to the logarithmic property  $\ln(1-x) \approx -x$  for  $x \ll 1$ .  $\nabla I$  can also be expressed as the integral of the absorption slices over optical depth increments:

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

which can be placed into equation 2.8 leading to

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

We can then convert  $d\tau$  to our path using equation 2.3 leading to

$$AMF = \frac{-1}{\tau_v} \int_0^{\infty} \frac{\partial \ln I_B}{\partial \tau} \alpha(z) \eta(z) dz \quad (2.9)$$

where  $\alpha(z)$  and  $\eta(z)$  represent absorption cross section in  $\text{m}^2 \text{ molecule}^{-1}$ , and number density in molecules  $\text{m}^{-3}$  respectively. This uses the attenuation cross section relationship to optical depth (see section 2.3.2).

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

$$\begin{aligned} \alpha_e &= \frac{1}{\Omega_v} \int_0^{\infty} \alpha(z) \eta(z) dz \\ &= \frac{\tau_v}{\Omega_v} \end{aligned}$$

Then replacing the  $\tau_v$  in equation 2.9 we obtain:

$$AMF = - \int_0^\infty \frac{\partial \ln I_B}{\partial \tau} \frac{\alpha(z)}{\alpha_e} \frac{\eta(z)}{\Omega_v} dz \quad (2.10)$$

Often the integrand of this AMF formula (equation 2.10) is broken apart into two defining terms: the scattering weights  $\omega(z)$  and the shape factor  $S(z)$ , described here:

- ω The scattering weights describing sensitivity of the backscattered spectrum to the abundance of an absorber at altitude z:

$$\omega(z) = - \frac{1}{AMF_G} \frac{\alpha(z)}{\alpha_e} \frac{\partial \ln I_B}{\partial \tau} \quad (2.11)$$

It is worth noting that in the OMI satellite product, the provided  $\omega(z)$  term does not include the  $\frac{1}{AMF_G}$  term and the calculations which follow therefor do not include this term when utilising the provided  $\omega$ . This is not noted in any of the papers which recalculate the AMF from the OMI product, due to them recalculating the  $\omega$  term themselves with a radiative transfer model such as LIDORT.

- η the shape factor describes the profile of an absorber ( $\eta(z)$ ) normalised by its total vertical column amount ( $\Omega_v$ ):

$$S(z) = \frac{\eta(z)}{\Omega_v} \quad (2.12)$$

Plugging equations 2.11 and 2.12 into equation 2.10 gives us:

$$AMF = AMF_G \int_0^\infty \omega(z) S(z) dz \quad (2.13)$$

Since we are using the  $\omega$  provided by OMI, the  $AMF_G$  term is removed from this calculation as it is not part equation 2.11 leading to

$$AMF = \int_0^\infty \omega(z) S(z) dz \quad (2.14)$$

Additionally the AMF can be determined using the sigma ( $\sigma$ ) coordinate system. A conversion to the  $\sigma$  vertical coordinate is performed using  $P = \sigma(P_S - P_T) + P_T$ , where  $P_T$  is pressure at the top of the atmosphere and  $P_S$  is surface pressure. This can be useful when running global atmospheric models as the ground altitude is always at  $\sigma = 1$  and we need not worry about topography.  $S_\sigma$  is a dimensionless normalised shape factor on the  $\sigma$  coordinate system. In the sigma coordinate system we calculated the shape factor as defined in Palmer et al. (2001):

$$S_\sigma(\sigma) = \frac{\Omega_a}{\Omega_v} C_{HCHO}(\sigma) \quad (2.15)$$

where  $\Omega_a$  is the vertical column of air from the surface to the top of the atmosphere and  $C_{HCHO}(\sigma)$  is the mixing ratio of HCHO. The hydrostatic relation  $P = -\rho_a g z$ , with  $\rho_a$ ,  $g$ , being density of air, and gravity respectively lets us integrate over the sigma

coordinates:

$$\rho_a g z = \sigma (P_S - P_T) + P_T$$

$$d\sigma = -\frac{\rho_a g}{P_S - P_T} dz$$

Substitution into 2.14 gives AMF using the sigma coordinates:

$$AMF = \int_0^1 w(\sigma) S_\sigma(\sigma) d\sigma \quad (2.16)$$

## 2.6 Recalculation of OMI HCHO

The AMF is needed to transform the slant column (SC) viewed by the satellite into a vertical column ( $\Omega$ ):

$$AMF = \frac{SC}{\Omega} \quad (2.17)$$

A slant or vertical column is expressed in molecules  $\text{cm}^{-2}$ . OMI HCHO vertical columns are calculated using modelled a priori HCHO profiles (see section ??). When comparing satellite measurements against models it is important to recognise the impact of this a priori on the total column values. This is complicated by how OMI is differently sensitive to HCHO (and other trace gases) vertically throughout the atmosphere. When comparing OMI vertical columns ( $\Omega_O$ ) to GEOS-Chem ( $\Omega_G$ ), the satellite AMF needs to be recalculated using GEOS-Chem modelled vertical gas profiles as the a prioris. Without performing this step a bias between modelled and measured total column values may be due to the a priori rather than chemistry or measurements (Palmer et al. 2001; Lamsal et al. 2014).

Here, two new AMFs are calculated, both using GEOS-Chem HCHO profiles as the new a priori. The first ( $AMF_{GC}$ ) uses the original satellite scattering weights while recalculating the shape factor, whereas the second ( $AMF_{PP}$ ) also recalculates scattering weights.  $AMF_{PP}$  is created using code initially written by Professor P. Palmer (see sections ?? and 2.6.5 for more details). A reference sector correction is determined using the method described in Abad et al. (2016), (see section 2.6.7). This correction is unique for each of the 60 *measurement tracks* used by OMI. Finally the correction is applied to each pixel to create the corrected vertical column. The end product is three sets of corrected vertical columns (VCC): the original ( $\Omega_{OC}$ ), one using GEOS-Chem shape factors ( $\Omega_{GC}$ ), and one from Palmer's code ( $\Omega_{PC}$ ).

### 2.6.1 Outline

An outline in computational order of what takes place when recalculating the  $\Omega$  from OMI follows:

1. GEOS-Chem satellite overpass output (see section 2.4.7.1) is used to create new shape factors ( $S_z$  and  $S_\sigma$ ).
  - (a) Pressure edges and geometric midpoints are determined, along with altitudes ( $z$ ), and box heights ( $H$ ).

- (b) Number density and mixing ratio of HCHO ( $n_{HCHO}$ ,  $C_{HCHO}$  respectively) are taken or created from model outputs HCHO(ppb), air density (molec  $\text{cm}^{-3}$ ), and box heights ( $H(z)$ )
  - (c) Total column HCHO from GEOS-Chem ( $\Omega_G$ ) is calculated  $\Omega_G = \Sigma_z (n_{HCHO}(z) \times H(z))$ , along with total column air ( $\Omega_A$ , calculated similarly)
  - (d) The shape factor  $S_z(z)$  is calculated on each altitude  $S_z(z) = n_{HCHO}/\Omega_{HCHO}$ .
  - (e) Pressures ( $p$ ) are used to create sigma coordinates  $\sigma(z) = (p(z) - p_{TOA})/(p(0) - p_{TOA})$
  - (f)  $S_\sigma(z)$  is calculated on each altitude:  $S_\sigma(z) = C_{HCHO}(z) \times \Omega_A/\Omega_{HCHO}$
2. Satellite pixels (SC, scattering weights ( $\omega(z)$ )), pressure levels, latitude and longitude) are read from the OMHCHO dataset
  3. For each pixel, a new AMF ( $\text{AMF}_{GC}$ ) is created using the GEOS-Chem shape factors and satellite scattering weights:
    - (a) scattering weights ( $\omega$ ) are interpolated onto the same vertical dimensions ( $z$  and  $\sigma$ ) as the shape factors.
    - (b) Integration (approximated using rectangular method) is performed along the vertical dimension to calculate the new AMF on both coordinate systems:

$$\text{AMF}_z = \Sigma_z (\omega(z) \times S_z(z) \times H(z)) \quad (2.18)$$

$$\text{AMF}_s = \Sigma_\sigma (\omega(\sigma) \times S_\sigma(\sigma) \times d\sigma) \quad (2.19)$$

4. These two AMFs represent the same thing using different coordinates, and either one can be used as the  $\text{AMF}_{GC}$
5. The  $\text{AMF}_{PP}$  (created separately) and  $\text{AMF}_{GC}$  are used to determine the new vertical columns ( $\Omega_P$ ,  $\Omega_G$  respectively):  $\Omega = SC/\text{AMF}$
6. A reference sector correction (RSC) is defined each day using these AMFs along with modelled HCHO over the remote pacific:
  - (a) GEOS-Chem satellite overpass output ( $\Omega_{\text{GEOS-Chem}}$  from  $140^\circ\text{W}$  to  $160^\circ\text{W}$ ) are averaged monthly and longitudinally to provide a modelled reference sector  $\Omega_0[\text{lat}]$
  - (b) The modelled reference slant columns ( $MRSC$ ) are calculated using  $MRSC = \Omega_0 \times \text{AMF}$  for each AMF
  - (c) For each satellite pixel between  $140^\circ\text{W}$  and  $160^\circ\text{W}$ , the correction is calculated as the measured SC minus the  $MRSC$  at the nearest latitude:

$$\text{corr}[\text{lat}, \text{track}] = SC[\text{lat}, \text{track}] - MRSC[\text{lat}]$$

- (d) These corrections are binned by satellite detector (track: 1-60), and latitude ( $0.36^\circ$ ; 500 latitudes from  $90^\circ\text{S}$  to  $90^\circ\text{N}$ )

- (e) The median entry of each bin is determined and this forms the RSC[lat,track] (eg. figure 2.30)
- 7. VCC are determined using  $VCC = (SC - RSC[lat, track]) / AMF$  for each measured SC and using each AMF, with the RSC linearly interpolated to the latitude of the satellite pixel
- 8. The VCC (along with most of the pixel and GEOS-Chem data) are binned onto a  $0.25^\circ$ by  $0.3125^\circ$ grid along with how many pixels have been binned, and the average pixel uncertainty in product OMHCHORP

Figure 2.27 shows an overview of how these profiles are used in this thesis. Output from GEOS-Chem is combined with OMHCHO swath data to produce a gridded HCHO file which contains HCHO vertical columns recalculated with GEOS-Chem a priors. PP code is run on a subset of the globe covering Australia and the pacific ocean, producing  $AMF_{PP}$ . The output keeps the original AMF as well as those recalculated using GEOS-Chem ( $AMF_{OMI}$ ,  $AMF_{GC}$ , and  $AMF_{PP}$ ). Additionally, MOD14A1, OMAERUVd and OMNO2d data are used to create masks which are also stored in the OMHCHORP dataset. The creation of fire, smoke, and anthropogenic influence masks is described in section 2.7.

## 2.6.2 Creating new shape factors

In order to visualise and analyse OMI HCHO columns, slant columns are transformed into vertical columns using the AMF. The shape factor ( $S$ ) is one of the key components in creation of the AMF (see section 2.5, equation 2.14). The shape factor is calculated using GEOS-Chem satellite output (see section 2.4.7.1) which provide simulated HCHO concentration profiles ( $\eta(z)$ ) and total columns ( $\Omega$ ) at  $2 \times 2.5^\circ$ horizontal resolution. Using equation 2.12 to determine the shape factor is straightforwards  $S(z) = \frac{\eta(z)}{\Omega}$ . The associated OMI per-pixel scattering weights are not changed in this calculation (unlike in section 2.6.5).

Model output is in molecules per billion molecules of air (ppb), and is converted before being used in the shape factor calculation. The following equation converts model profile output from ppb into number densities:

$$\eta_{HCHO} = ppb_{HCHO} \times \eta_a \times 10^{-9} \quad (2.20)$$

where  $\eta_{HCHO}$  is the number density of a HCHO,  $\eta_a$  is the number density of air (from model output), and  $ppb_{HCHO}$  is the molecules of that species per billion molecules of air. The modelled total vertical column  $\Omega_{HCHO}$  is determined by:

$$\Omega_{HCHO} = \sum_z (\eta_{HCHO} \times H(z))$$

where  $H(z)$  is the box height for level  $z$ . In effect this equation sums over the molecules per  $\text{cm}^2$  in each altitude level.

As a sanity check  $S_\sigma$  is calculated (through equation 2.15) to confirm that these shape factors are equivalent. Comparing the resulting AMFs created by equations 2.16 and 2.14 for each pixel provides confidence in the unit conversions (and other factors)

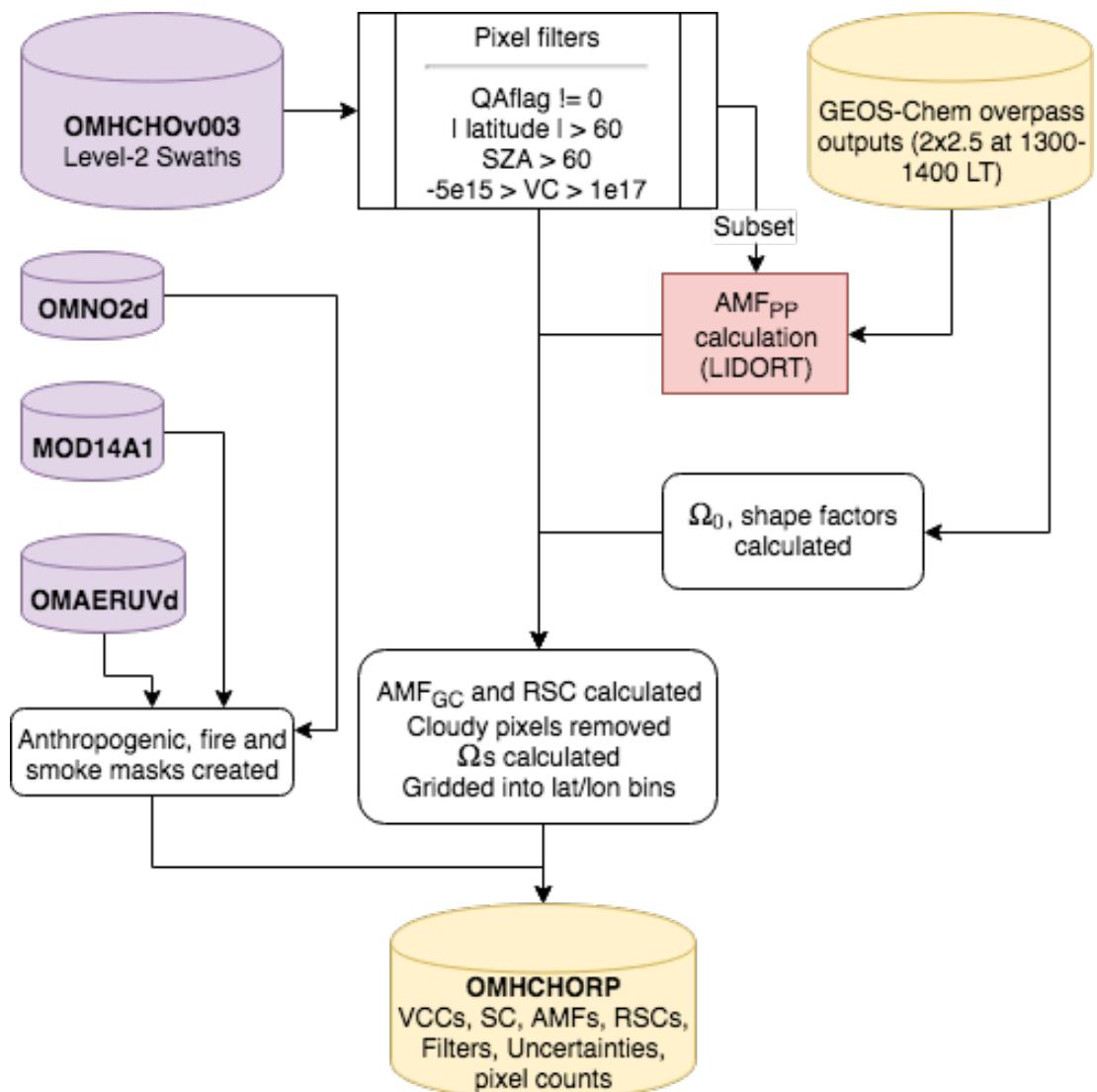


FIGURE 2.27: Flow diagram showing how OMHCHO level two swath data is read, processed, and gridded in this thesis

applied. For example see figure TODO: plot showing correlation between AMFz and AMFs - or state that these are exactly one to one if that is the case..

### 2.6.3 Reading satellite data

First satellite slant columns of formaldehyde for the years January 1st, 2005 - April 1st, 2013 are downloaded (see section ??). The data set used is from the Ozone Monitoring Instrument (OMI) on board the Aura satellite, as it has data for the entire time line and sufficiently covers the southern hemisphere. When reading OMHCHO level 2 swath files, several factors are taken into account in order to filter uncertain and erroneous pixels. The process is outlined in figure 2.27 for a single day. First all *good* pixels (those with QA flag equal to 0) are read into a long list (roughly 1 million per day). These are then filtered by solar zenith angle (SZA) and latitude, similarly to other works (Marais et al. 2012; Barkley et al. 2013; Bauwens et al. 2016; Zhu et al. 2016, eg.). This filtering removes highly uncertain pixels, along with those for which instrument problems such as the OMI row anomaly (see section ??) may have affected.

Satellite measurements polewards of  $60^{\circ}$  north or south are removed as well as measurements with SZA greater than  $60^{\circ}$ . Pixels with cloud fraction greater than 40% are removed after being used in determining the reference sector correction (see section 2.6.7), as is done in Gonzalez Abad et al. (2015) and De Smedt et al. (2015). Further filtering is performed to remove the measurements which are most likely to be unrealistic: those with column density outside the range  $-0.5 \times 10^{16}$  to  $10^{17}$  molec  $\text{cm}^{-2}$ , as is performed by Zhu et al. (2016). This filter is required due to currently unexplained large negative values which occur in the OMI HCHO product increasingly over time. Figure 2.28 shows how unfiltered HCHO columns are affected by a small set of highly negative values which heavily affect the mean column amount over any region. The histograms here show the negative (left) and positive (right) total column HCHO measurements from a subset of swaths over Australia, on the 18th of March 2013. The highly negative values can be seen around the  $-10^{19}$  molecules  $\text{cm}^{-2}$  region.

Additional information is added to each pixel, including the new AMFs (see the following subsections). Each pixel and its relevant data are saved in a long list for subsequent gridding. The shape factors and scattering weights for each pixel lie along a z-axis which is vertically resolved to 47 layers.

### 2.6.4 Creating the new AMF

From equation 2.14 we have:

$$\text{AMF} = \int_0^{\infty} \omega(z) S(z) dz$$

then using the  $\omega(z)$  from satellite output, along with our calculated  $S_z$  interpolated linearly onto the same vertical grid as  $\omega(z)$ , the AMF can be determined through integration. The integration is performed using a simple rectangular method, which multiplies the integrand midpoints by the change in height, and then takes the sum for each vertical box. This assumes that the provided scattering weights and shape factors are linear between the 47 resolved values.

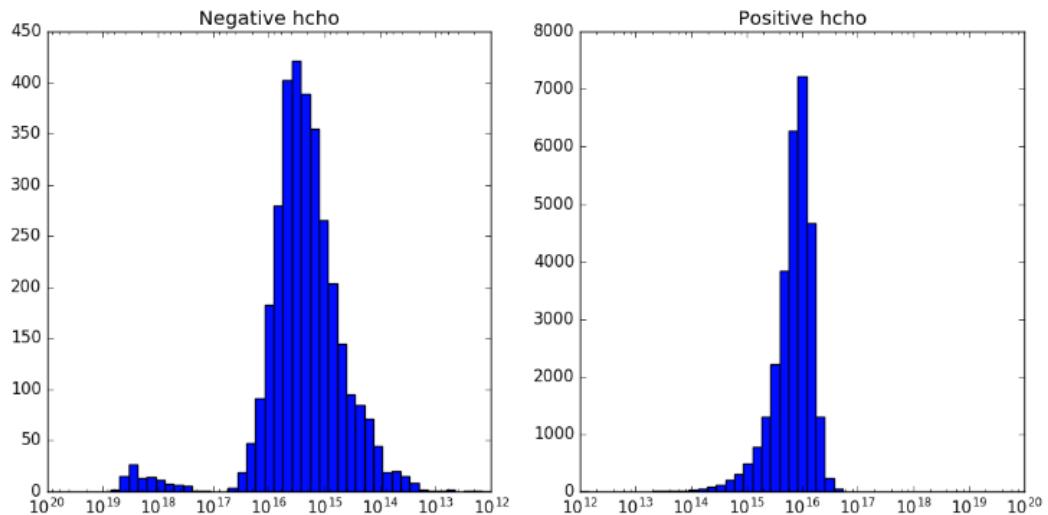


FIGURE 2.28: Column density histograms for a subset of OMI swaths over Australia on the 18th of March 2013. Negative entries are shown in the left panel, positive in the right, note the different scale between negative and positive panels.

### 2.6.5 Recalculating the AMF using PP code

Some of the pixels (those covering Australia and most of the zonal band) have their AMFs recalculated using Fortran code written by Paul Palmer, and Randal Martin, subsequently updated by Luke Surl. I will refer to this as the PP code, and subscript the VCs and AMFs with PP when referring to those calculated through this method. Running this code is computationally expensive, and is subset to within the region (50-10° S, 160° W-160° E). The code uses a combination of GEOS-Chem a priori profile information and satellite measurement data to calculate the AMF by using LIDORT radiative transfer calculations to determine scattering. The instrument sensitivity (or scattering weights;  $\omega$ ) and shape factors for each pixel are calculated within the PP code, which outputs  $\text{AMF}_{PP}$ .

Code for recalculating AMFs using satellite swaths and modelled aerosol optical depths and gas profiles can be found at [http://fizz.phys.dal.ca/~atmos/martin/?page\\_id=129](http://fizz.phys.dal.ca/~atmos/martin/?page_id=129). The original method for HCHO is layed out in Palmer et al. (2001), with modifications for clouds and use of the LIDORT RTM (Spurr 2002) as described by Martin (2003). This code does not work as is when using OMI satellite data, and requires modifications performed by Luke Surl at Edinburgh University. Additionally the tropopause heights averaged within satellite overpass times are required, and provided by modifying the GEOS-Chem diagnostic output.

Mie scattering and clouds can complicate the calculation of  $\omega(z)$ , however tables of values for this function at various parameter inputs can be used with modeled vertical shape factors for local AMF calculations. This has been done in the PP code and the AMF look-up-table (LUT) can be found in the source code at (<https://github.com/LukeSurl/amf581g>). In order to run, output is required from GEOS-Chem: the HCHO concentration profile averaged between 1300 and 1400 LT, including optical depths

at 550 nm, and dust concentrations. A subset of the OMI pixel information is also required, which is coalesced from OMHCCHO daily swath files into csv files (one per day), and read by the PP code in conjunction with the GEOS-Chem outputs for each day. The PP code then produces a list of recalculated AMFs which is read by my python code and associated with the corresponding pixel (outlined in 2.27).

### 2.6.6 Vertical columns from AMF

All that remains for recalculating the total vertical column using our new a priori shape factor is to apply the new AMF to the slant columns and grid them onto our chosen resolution. Each satellite pixel at this stage has an associated SC along with three AMFs: the original ( $\text{AMF}_O$ ), one with recalculated shape factors ( $\text{AMF}_G$ ), and one completely recalculated using PP code ( $\text{AMF}_{PP}$ ). These are used to create vertical columns ( $\Omega$ ) through equation 2.6:  $\Omega = \text{SC} / \text{AMF}$ .

Figure TODO shows a comparison between the three  $\Omega$  over Australia for 2005. The first row shows the yearly average column amounts, as well as coloured rectangles describing regions which are averaged over time in row 3. Row two from left to right shows the relative differences between  $\Omega_P$  and  $\Omega_G$ ,  $\Omega_O$  and  $\Omega_G$ , and  $\Omega_O$  and  $\Omega_P$ .  
TODO: Add this plot here.

### 2.6.7 Reference sector correction

Each satellite slant column measurement is corrected by its divergence from a modelled reference sector. HCHO products from OMI and SCIAMACHY both use a median daily remote pacific ocean radiance reference spectrum, over 15°S-15°N, 140°-160°W where it is assumed that the only significant source of HCHO is methane oxidation (De Smedt et al. 2008; Barkley et al. 2013; Kurosu and Chance 2014).

The RSC method corrects for several problems, however it introduces some a priori model influence. One of the problems removed through this correction method is instrument degradation, which can introduce bias over time. Another is the possible influence of varying dead/hot pixel masks across 2-D detector arrays such as OMI (De Smedt et al. 2015). This method also corrects for the errors introduced through correlations between BrO and HCHO absorption cross sections, which are especially significant at high latitudes (Gonzalez Abad et al. 2015).

Here a new reference sector correction (RSC) is created using modelled and measured HCHO columns over the remote pacific, to produce corrected vertical columns (VCC). This follows Abad et al. (2016), and defines the remote pacific as the band between 140°W to 160°W. It uses the difference between slant columns ( $\Omega_S$ ) and reference slant columns ( $\Omega_{S_0}$ ) divided by the AMF, plus the modelled reference sector column ( $\Omega_{V_B}$ ):

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

This method is used in several works, including De Smedt et al. (eg. 2008), De Smedt et al. (2012), De Smedt et al. (2015), Barkley et al. (2013), and Bauwens et al. (2016). Recently this correction was expanded (for OMI data) to include latitudinal and instrument track influence by Gonzalez Abad et al. (2015).

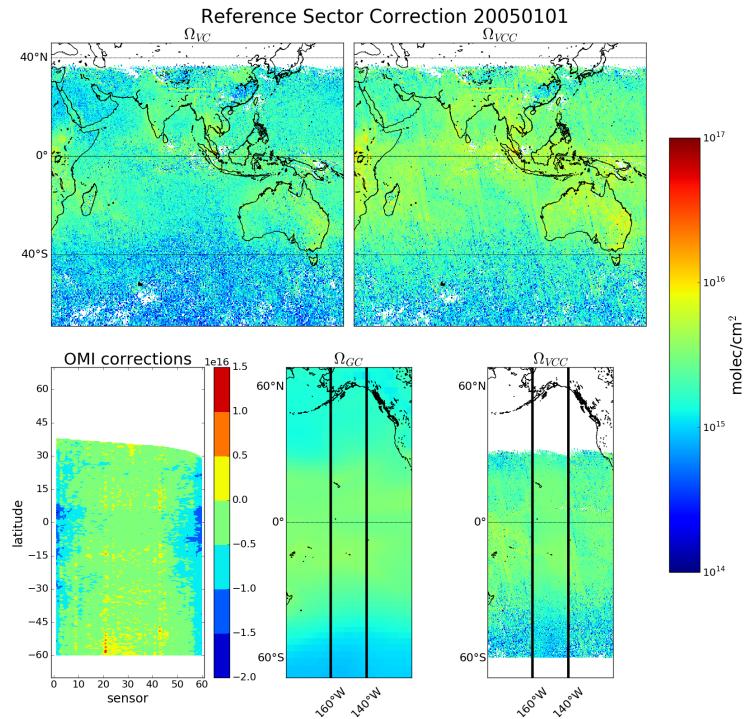


FIGURE 2.29: Example of remote pacific RSC using 8-day average measurements and one month modelled data.  $\Omega_{VC}$  shows the uncorrected vertical columns, while  $\Omega_{VCC}$  shows the corrected vertical columns. OMI corrections shows the correction applied globally based on latitude and OMI track number(sensor).  $\Omega_{GC}$  shows the GEOS-Chem modelled HCHO VC over the RSC, with  $\Omega_{VCC}$  showing the corrected VC over the same area.

To get the modelled slant columns, each of the AMFs (calculated in prior sections) is applied to the modelled vertical columns using equation 2.17. The longitudinal average is taken within the remote pacific, as corrections are assumed to be longitudinally invariant. The modelled reference sector is averaged over the month and interpolated latitudinally to 500 equidistant bins. Figure 2.29 shows the simulated reference sector VCs as an example, calculated for January 1st 2005. In this figure the vertical resolution is increased from  $2^\circ$  to  $0.36^\circ$ , through linear interpolation, in order to form 500 vertical bins which are used in correcting the satellite data. Each day, satellite measurements (pixels) over the remote pacific are used to determine a correction array. The model does not produce slant columns associated with each measurement, however one is created by multiplying the vertical column with the associated slant columns AMF.

For OMI swaths, each row of measured data contains 60 “Across track”(track) measurements. The track index ( $i$ ) relates the measurement to one of the 60 columns of data. Corrections ( $\text{molecules cm}^{-2}$ ) for each measurement are calculated by taking the difference between the measured slant column and the a priori slant column as

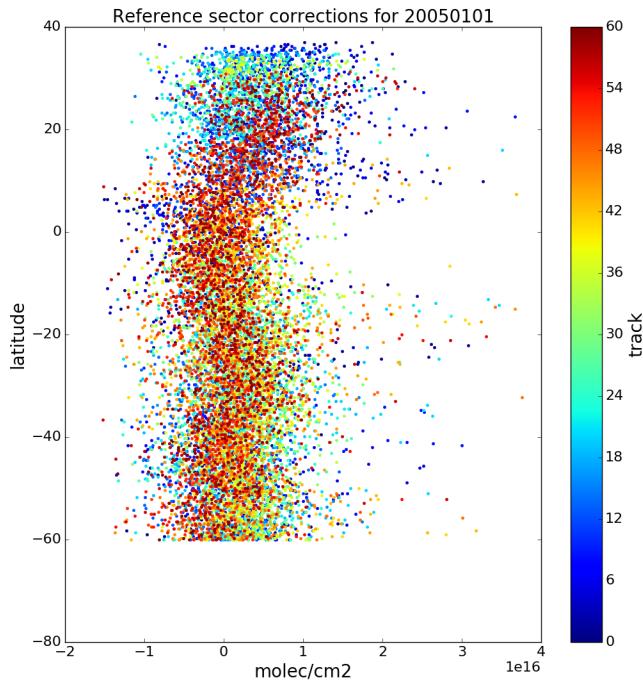


FIGURE 2.30: Example of track correction interpolations for January 1st 2005, points represent the difference between satellite slant column measurements and modelled slant columns over the remote pacific.

follows:

$$\text{Correction}(i, j) = SC_{HCHO}(i, j) - VC_{GEOS-Chem}(lat(j)) \times AMF(i, j) \quad (2.21)$$

where  $j$  represents a latitude index and  $VC_{GEOS-Chem}(lat)$  represents the a priori reference sector vertical column HCHO at the latitude corresponding to  $j$ . The RSC is independently calculated for each of the 60 tracks, at each latitude in the 500 0.36°bins. This provides a different RSC for each of the three AMFs.

Due to incomplete latitudinal coverage, the correction for each track is interpolated linearly between measurements, with corrections outside of the highest measured latitudes being equal to the corrections at the highest measured latitudes. Figure 2.30 shows an example of the 60 track corrections for January 1st 2005, the points are satellite measurements longitudinally averaged over the remote pacific, coloured by track number. Another way to look at this correction is given in the OMI corrections panel of figure 2.29, which has the sensors along the x axis, and latitude on the y axis, and shows how for this example 8-day period, the corrections are distributed with more negative values towards the left or right sensors, especially in the tropics.

One correction is associated with every good satellite measurement which is used to create a reference sector corrected measurement (Vertical Column Corrected or

VCC) through the following equation:

$$VCC(i,j) = \frac{SC_{HCHO}(i,j) - Correction(i, lat(j))}{AMF(i,j)} \quad (2.22)$$

For each day, good satellite measurements and associated SC, VC, VCC, and AMFs are listed.

### 2.6.8 Binning the results daily

Finally the pixels are binned into a gridded dataset named OMHCHORP, as shown in figure 2.27. The resolution is chosen to match the native resolution of GEOS-Chem ( $0.25 \times 0.3125^\circ$ ) and the GEOS met data fields. A bin entry count is used to allow easy re-binning, and can be used to check for sparse data days due to filtering or poor weather. Data averaged into this dataset are as follows:

1. satellite SC
2. satellite AMF
3. satellite VC
4. satellite RSC VC
5. GEOS-Chem recalculated AMF
6. GEOS-Chem recalculated VC
7. GEOS-Chem recalculated RSC VC
8. GEOS-Chem AMF recalculated using Paul Palmer code ( $AMF_{PP}$ )
9. GEOS-Chem RSC VC based on  $AMF_{PP}$
10. Smoke AAOD from OMAERUVd (mapped into bins from  $1 \times 1^\circ$  resolution)
11. satellite pixel counts (summed into bins)
12. fire counts (summed into bins)

This whole process requires some processing time as well as RAM and computer storage space, and has been performed on the National Computing Infrastructure (NCI) supercomputer cluster. In order to reprocess one year of swath files,  $\sim 162$  GB ( $142 + 16 + 4$  OMHCHO, MOD14A1, and OMNO2d respectively) of daily data was downloaded and then transformed into  $\sim 8$  GB (per year) of daily gridded data. This takes around 90 minutes per day, and is very parallelisable as each day is completely independent once the model has run in each required configuration. Initially parallelism was built into the python code, however simply sending separate jobs to NCI's process queue was simpler and more scalable. As much as possible, processing is done using the HDF-5 or NetCDF-4 formats, with some GEOS-Chem output being read from bitpunch. The scripts to regrid and reprocess the swath data set are available from github at TODO.

### 2.6.9 Difference between new and old OMI HCHO columns

New corrected vertical columns (VCC) of HCHO are created at global 0.25 by 0.3125° horizontal resolution. These calculations are compared over Australia in figure(s) 2.31, TODO: regression, and time series. Figure 2.31 shows how the recalculated columns compare to the original (OMI, left). Recalculation increases the January column amounts while also slightly flattening the distributions (row 3). The effect of not recalculating the  $\omega_z$  can also be seen in figure 2.32 which shows the altered satellite vertical columns using each method.

Figure 2.32 shows vertical columns of HCHO for: column 1) the original satellite swaths, column 2) recalculated without changing the provided scattering weights, and column 3) fully recalculated vertical columns. Each grid square (at 0.25 by 0.3125° resolution) has been created by binning the recalculated satellite pixels within the month. The average number of pixels per land square is inset as text, changing due to how the fire filter is applied. Each row has a stricter fire filter applied from top to bottom, with no fire filter on the first row up to filtering pixels from squares with fires up to 8 days prior. This figure looks at March 2005 with biomass burning filtered differently in each row. Active fires over the last 0, 1, 2, 4, and 8 days are filtered as the row number increases.

Figure TODO shows an analysis of the differences between running the recalculation with and without updating the  $\omega_z$ .

TODO: compare how many nans from palmers code to my code The AMF calculated using professor Palmer's produces TODO: MORE OR FEWER reasonable (within screening range of  $-5 \times 10^{15}$  to  $1 \times 10^{17}$ ) vertical column amounts. Stricter filtering must be balanced against both coverage and the sensitivity of the AMF determination to recalculating  $\omega_z$ .

Figure TODO: shows global and Australian HCHO averaged total column maps for January 2005, along with the reduced major axis (RMA) regression correlation and percentage difference. This comparison shows how reprocessing with an updated model can have a systematic influence on the total column.

## 2.7 Filtering Data

In order to examine only biogenic processes, pyrogenic and anthropogenic influences need to be removed from modelled and measured data. Biomass burning can be a large local or transported (via smoke plumes) source of HCHO, CHOCHO, glyoxal, and other compounds which influence levels of both HCHO and isoprene. Anthropogenic emissions from power generation, transport, and agriculture can influence these levels as well. Where possible these influences need to be removed so that calculations of purely biogenic emissions are not biased. In GEOS-Chem we can simply turn off pyrogenic and anthropogenic emissions, however in satellite datasets we need to mask potentially affected pixels.

Influence from biomass burning can be removed through measurements of acetonitrile and CO (eg. Wolfe et al. 2016; Miller et al. 2016), or else removal of scenes coincident with satellite detected fire counts and aerosol absorption optical depth as done in Marais et al. (2014). Marais et al. (2012) remove pixels colocated with non zero

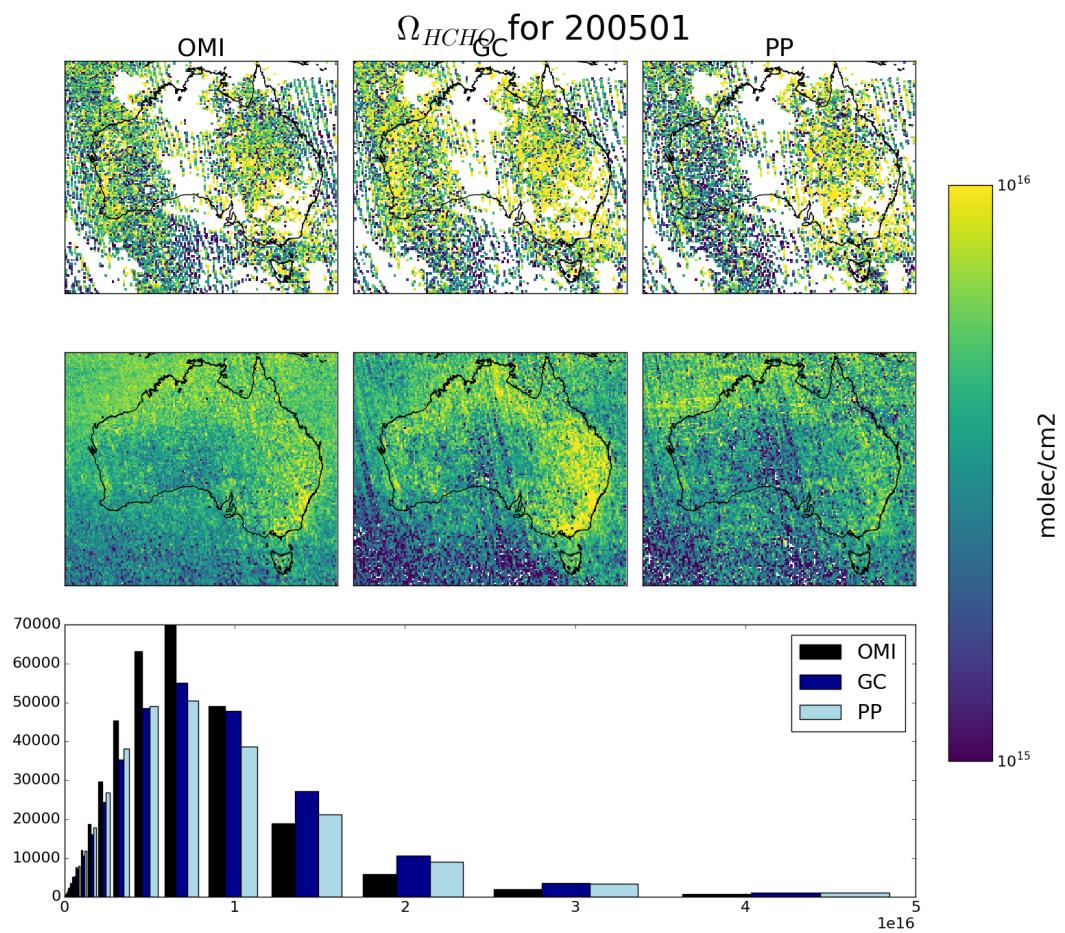


FIGURE 2.31: Row 1: regridded corrected  $\Omega_{HCHO}$  from OMHCHO on January 1, 2005: original (left), recalculated using new shape factors (middle), and additionally using updated scattering weights (right). Row 2: shows the monthly average for January 2005. Row 3: shows the distribution over the month for each of the three column amounts.

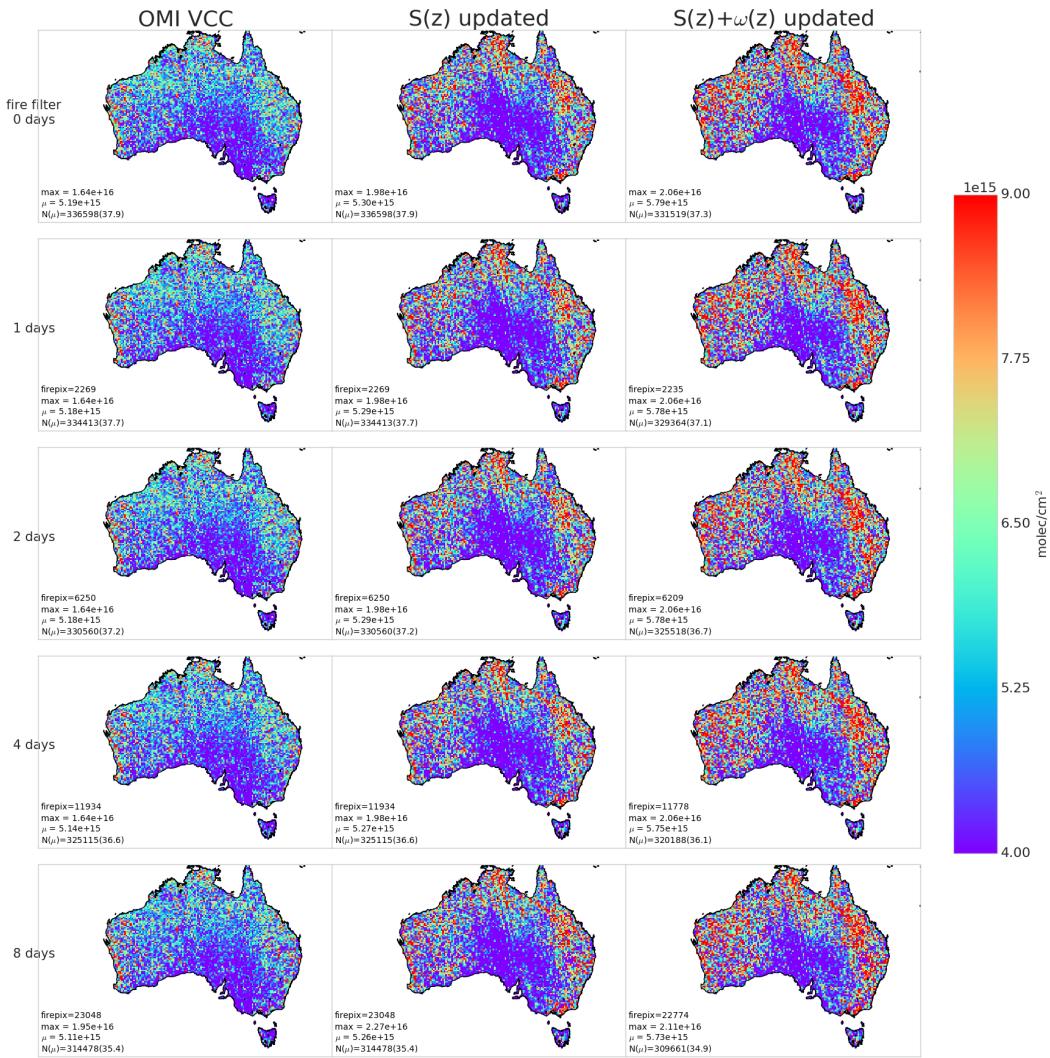


FIGURE 2.32: Column 1: Reference sector corrected HCHO vertical columns  $\Omega$  from OMHCCHOv003. Column 2:  $\Omega$  with recalculated a priori shape factors using GEOS-Chem v10.01. Column 3:  $\Omega$  with recalculated a priori shape factors and scattering weights using GEOS-Chem v10.01 and LIDORT. Row 1-5: increasing number of prior days which have active fires included when masking fire influence.

TABLE 2.3: How many satellite pixels are filtered by pyrogenic and anthropogenic masking. Left to right the columns display year, how many land pixels are read over Australia, how many of these pixels are removed by the pyrogenic filter, how many are removed by the anthropogenic filter, and how many are removed in total. In parenthesis are the portion of pixels filtered.

Year	Pixels	Pyro	Anthro	Total
2005	3.9e+06	4.1e+05(10.7%)	5.0e+04( 1.3%)	4.4e+05(11.5%)
2006	3.8e+06	5.1e+05(13.5%)	9.2e+04( 2.4%)	5.6e+05(14.7%)
2007	3.7e+06	4.4e+05(11.9%)	7.5e+04( 2.0%)	4.9e+05(13.0%)
TODO				
TODO				

fire counts in any of the prior eight days, within grid squares with  $1 \times 1^\circ$  resolution. Barkley et al. (2013) use fires from the preceding and concurrent day, within local or adjacent grid squares, with grid resolution of  $0.25 \times 0.3125^\circ$ . Wolfe et al. (2016) disregard HCHO measurements when acetonitrile  $> 210$  pptv and CO  $> 300$  ppbv, while acetonitrile  $> 200$  pptv is used to determine fire influence in Miller et al. (2016). TODO: look at yearly correlation, compare to exponential curve and look for fire outliers As seen in TODO: citation, HCHO concentrations scale exponentially with temperature. This allows another method for detecting the influence of non-biogenic HCHO emission/creation by looking for outliers above the curve at low temperature. Zhu et al. (2013) has a similar analysis over south-eastern USA showing an exponential correlation of  $HCHO = \exp(0.15 \times T - 9.07)$ .

I use satellite data to account for anthropogenic and pyrogenic influences on the OMHCHO satellite HCHO columns. MODIS fire counts are used in conjunction with smoke AAOD enhancements (from OMI) to remove data points which may be affected by fires or fire smoke plumes. OMI NO<sub>2</sub> measurements are used to mask potential anthropogenic influence. These masks negatively affect uncertainty, as fewer measurements are available to be averaged. This section describes the creation and effects of filters used on satellite data.

A quick summary of how much data is filtered over Australian land squares is provided in table 2.3, and an quick check of how many pixels are filtered in January 2006 can be seen in Figure 2.33. The anthropogenic filter completely removes grid squares over Sydney and Melbourne, and high removal rates over Brisbane. Other major cities in Australia either do not emit enough NO<sub>2</sub> or are too spread out and do not breach the threshold to be filtered as anthropogenic.

### 2.7.1 Fire and smoke

The method used in this thesis follows that of Marais et al. (2012), and Barkley et al. (2013), with active fires filtered using fire counts, and smoke filtered out using smoke AAOD. We use the MODIS fire counts, detected from space using the combined product from Terra and Aqua (Terra at 10:30, 22:30 LT; Aqua at 13:30, 01:30 LT). Smoke plumes are filtered using smoke AAOD from product OMAERUVd, with the threshold determined through analysis of Australian AAOD distributions.

Anthro and Fire filters applied on 20060101-20060131

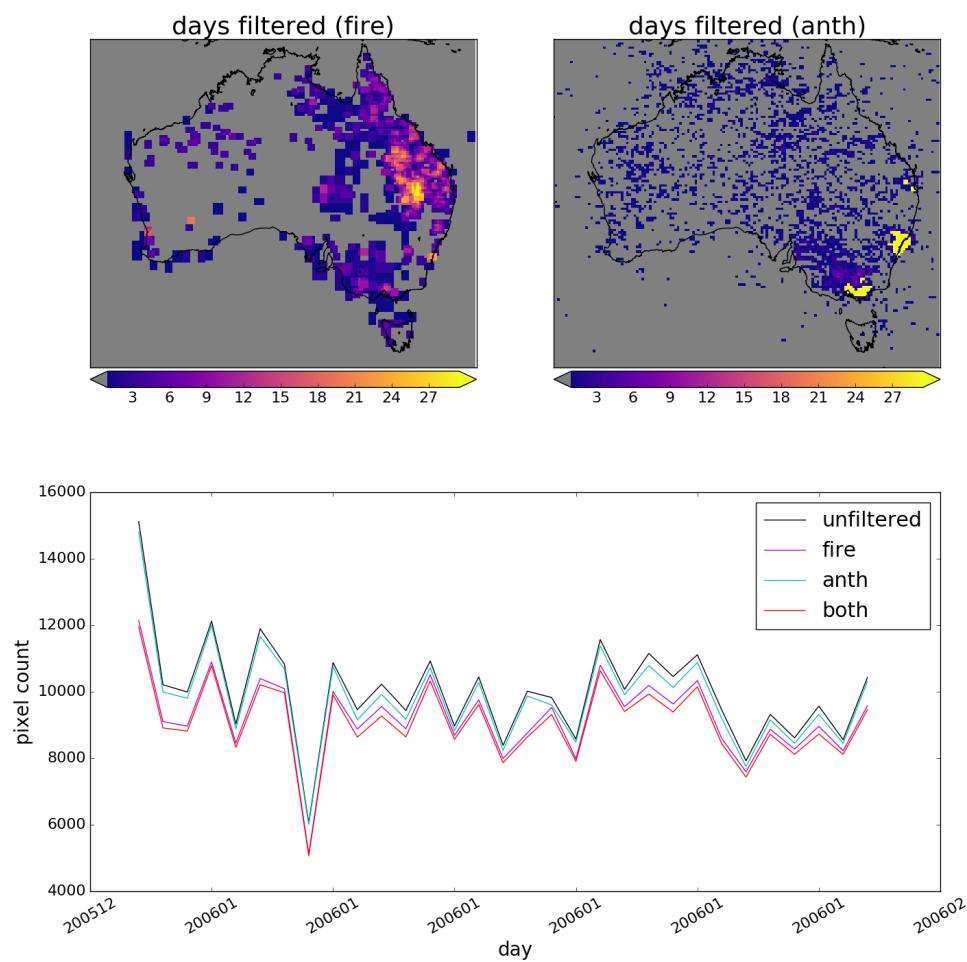


FIGURE 2.33: Top row shows grid squares filtered out by pyrogenic(left) and anthropogenic(right) influence masks during January 2006. Along the bottom is the time series of total pixels over Australian land squares with and without filtering the data.

OMHCHO total column HCHO  $\Omega$  is processed into a  $0.25 \times 0.3125^\circ$  horizontal daily grid. Pyrogenic filters are interpolated to the same horizontal resolution as  $\Omega$  to simplify application. The following steps are performed in order to create the pyrogenic influence mask:

1. MOD14A1 daily gridded Aqua/Terra combined fire counts ( $1 \times 1 \text{ km}^2$ ) are read, and binned into  $0.25 \times 0.3125^\circ$  bins (matching the resolution of binned  $\Omega$ ).
2. A rolling mask is formed which removes  $\Omega$  if one or more fires are detected in a grid square, or in the adjacent grid square, up to 2 days previously. This includes the current day, making 3 days of fires in total being filtered out on each day.
3. AAOD at 500 nm is mapped from OMAERUVd ( $1 \times 1^\circ$  resolution) onto the  $0.25 \times 0.3125^\circ$  resolution.
4. An AAOD threshold of 0.03 is determined through visual analysis of AAOD distributions over several days, including days with and without influence from active fires, dust, and transported smoke plumes (see figure 2.34).
5. Grid squares with AAOD over this threshold are considered potentially affected by transported fire smoke.

Determining the AAOD due to smoke can be difficult since both smoke and dust absorb UV radiation (Ahn2008; Marais et al. 2012). AAOD is less sensitive to cloud contamination than AOD, and I use AAOD from the daily gridded level 3 satellite product OMAERUVd (Ahn2008) described in section 2.2.1.3 to provide a filter for smoke plumes. Although removing gridsquares with dust reduces how much data is available to analyse, dust in Australia is highly episodic and should not affect more than a few days per year, especially over regions with high tree coverage (Shao et al. 2007).

Filtering fire smoke using AAOD is done by removing OMHCHO gridsquares where the AAOD is above a 0.03, after the AAOD is mapped from  $1 \times 1^\circ$  to the same  $0.25 \times 0.3125^\circ$  resolution as our OMHCHO gridded product. The threshold is determined through analysing AAOD over Australia in 4 scenarios: normal conditions, active local fires, during influence from transported fire smoke, and large scale dust storms. Figure 2.34 shows AAOD (columns 1 and 2), with AAOD distribution in column 3, along with satellite imagery on the same day in column 4 (from <https://worldview.earthdata.nasa.gov/>). The scenarios listed are shown from row 1 to 4, and AAOD = 0.03 is demarcated by a horizontal line in the density plots in column 3. Figure 2.35 shows how the smoke filter compares to the fire filter over 2005. The left panel shows how many days are filtered using the AAOD threshold, while the right panel shows how many are filtered using the combined smoke and active fire masks. The time series shows pixels available for analysis before being filtered (black), after smoke filtering (cyan) and after the full fire filter is applied (magenta). The Ayre basin is the dominant area filtered by smoke detection, likely due to dust influence in the arid region.

Figure ?? shows what portion of pixels are filtered out by the pyrogenic filter. The top panel shows the spatial distribution of fire masks, with most pixels removed along the northern and eastern coastlines. A large portion of the filtered areas appear to correspond with forested areas (see figure ??), which suggests that forest fires are being

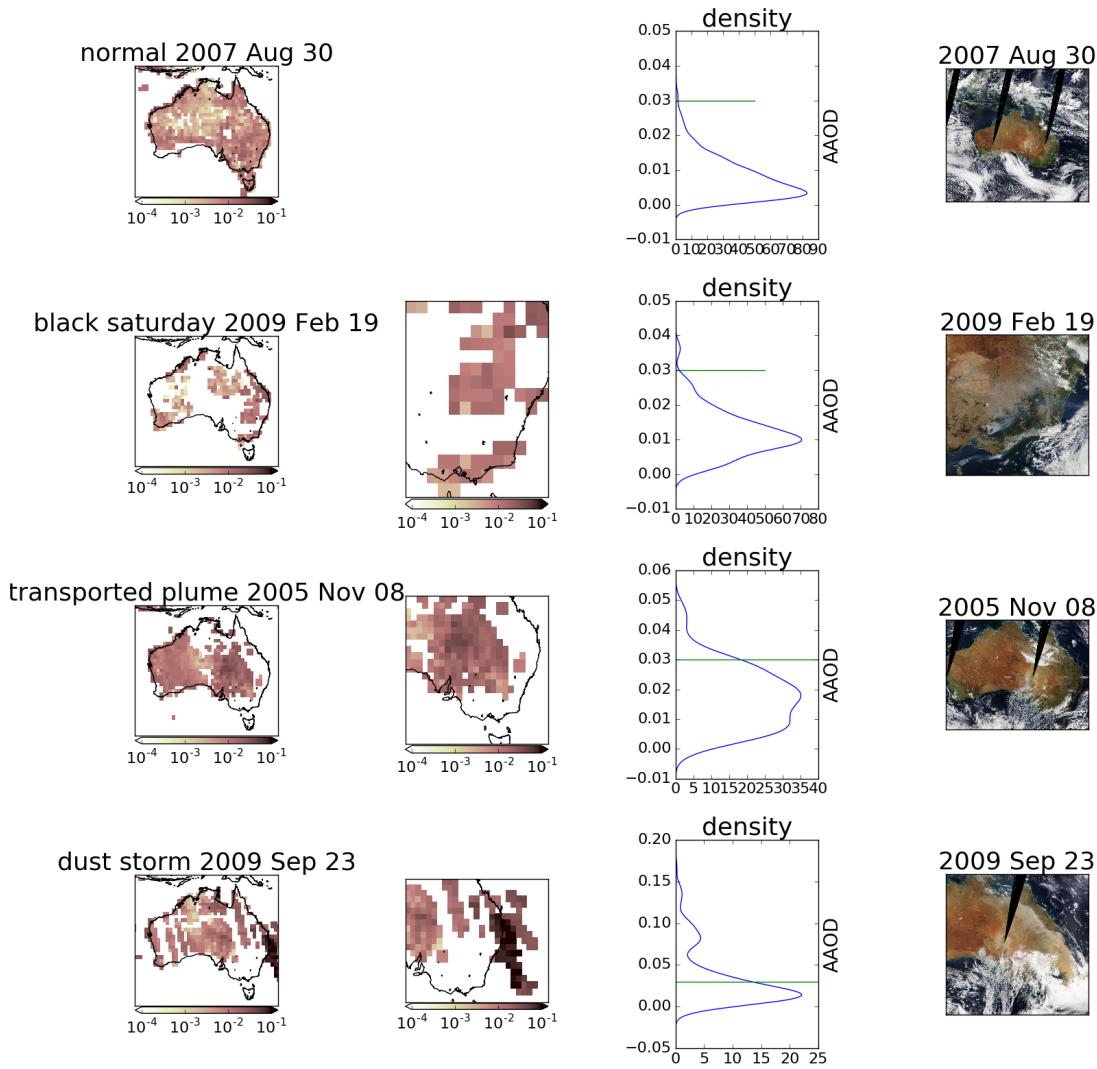


FIGURE 2.34: AAOD from OMAERUVd (columns 1, 2, 3) over Australia for four different scenarios (rows 1-4). Scenes from the same day are taken from the EOS Worldview website <https://worldview.earthdata.nasa.gov/>.

Fire and smoke filters applied on 20050101-20051231

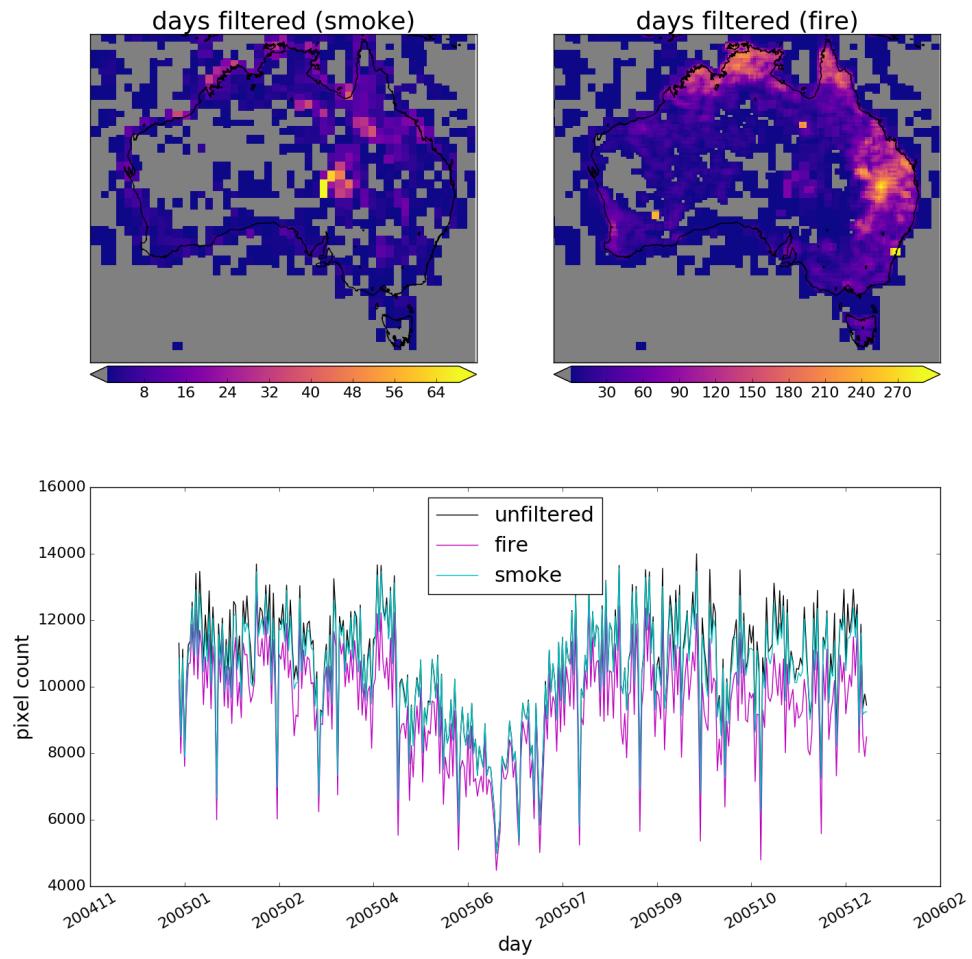


FIGURE 2.35: Smoke (top left) filtered days using AAOD $> 0.03$  compared to the combined (smoke and active fires) fire filter (top left). Time series shows how many pixels are available for analysis before and after filtering.

masked properly. Central Australia is largely unmasked, which could be due to a lack of sufficient vegetation to create a large enough fire to be seen by satellite. The other potential cause of pyrogenic filtering is the proliferation of petrol or gas wells (see figure 2.37 and figure 2.38). The filtering shown here is for 2005, when 388 gas wells existed in Queensland, however more than 2000 wells (cumulative) were approved by 2013, so this may cause more filtering over the course of this thesis' timeline (Carlisle 2012). To check this the filtering portion for 2012 is also plotted in Figure TODO: plot map of filtered squares for 2012 when it has been run by NCI. One clear hotspot is located over port Kembla (south of Sydney), most likely due to the flame which burns over the blast furnace stack throughout the year. Another hotspot can be seen in Western Australia over Kalgoorlie, where a large open cut gold mine "super pit" is always open and blasting daily. In Western Queensland over Mount Isa there is again a mining related hotspot. A large area in southern Queensland/northern NSW is also heavily filtered, potentially due to gas flaring in the Surat Basin, which has thousands of petrol and gas wells.

### 2.7.1.1 Checking that fire masks are influencing pyrogenic HCHO

Looking at temperature can provide evidence of pyrogenic HCHO. HCHO precursors are heavily tied to temperature (TODO:cite), and model output shows how higher temperature leads to an increase in HCHO levels. Figures 2.42 - 2.44 show the relationship between modelled temperature, and satellite HCHO for January 2005 within subsets of Australia. A reduced major axis regression is used to determine the correlation between surface temperature (X axis) and HCHO (Y axis). Using the natural log of HCHO we can take the linear regression and then exponentiate each side in the equation  $\ln Y = mX + b$  to get  $Y = \exp mX + b$ . This gives us the exponential fit as shown, with the correlation coefficient between  $\ln HCHO$  and temperature. The distributions of exponential correlation coefficients and  $m$  terms is shown in the embedded plot, with one datapoint available for each grid square where the regression is performed.

Figures 2.40 and 2.41 show the regressions between OMI HCHO total columns and temperature from GEOS-Chem output and CPC daily maximum temperatures respectively. Comparing against GEOS-Chem modelled surface temperatures first requires deresolution from  $0.25 \times 0.3125$  to  $2 \times 2.5^\circ$  latitude by longitude resolution. The left column in figure 2.40 shows scatter and RMA correlation within a single gridbox over 2 months from Jan 1 to Feb 28, 2005, without having applied either the fire nor anthropogenic masks to OMI HCHO columns. The right column shows the same correlation after applying the fire filter, affected datapoints are marked in teal (matching the red marked points in the left column). The analysis is repeated for Sydney, Canberra, and three gridsquares to the north-west, west, and southwest of Sydney (w1, w2, w3 respectively). Figure 2.41 shows the same analysis at higher resolution using CPC daily maximum temperatures (see 2.2).

One cause of high HCHO at lower temperatures is direct or transported emissions and subsequent products from biomass burning. One potential problem with showing this is that days with fire enhanced HCHO are also likely to be hot. Another problem with correlating heat and HCHO is that increased temperature accelerates HCHO destruction (Zheng2015). We test the fire mask by examining the relationship between

### Pyrogenic filter: 20050101-20060101

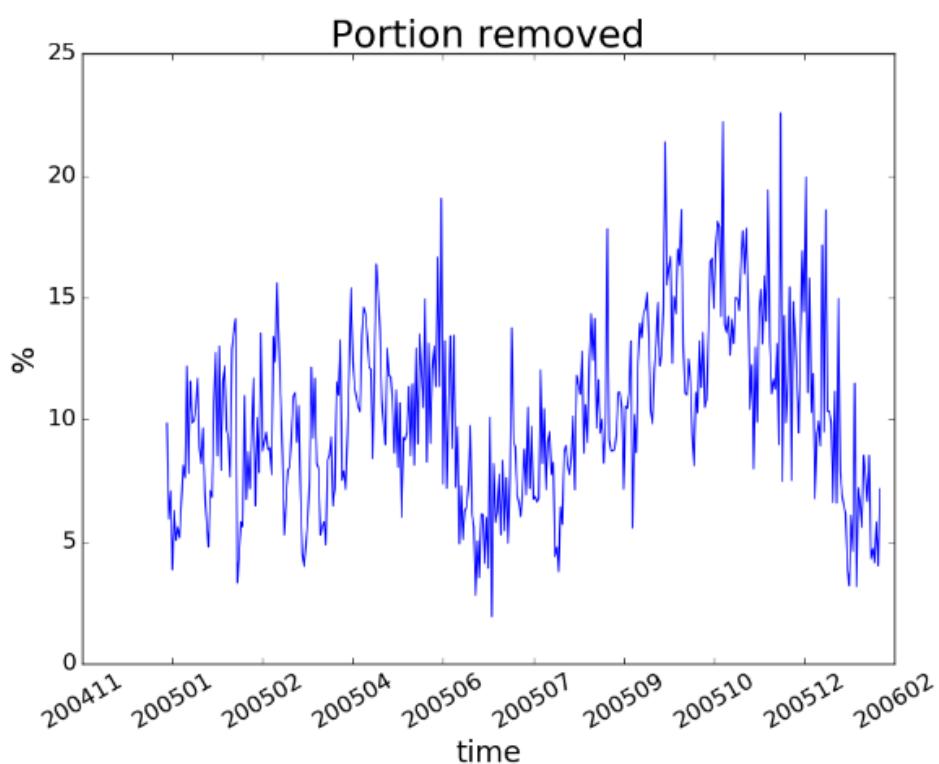
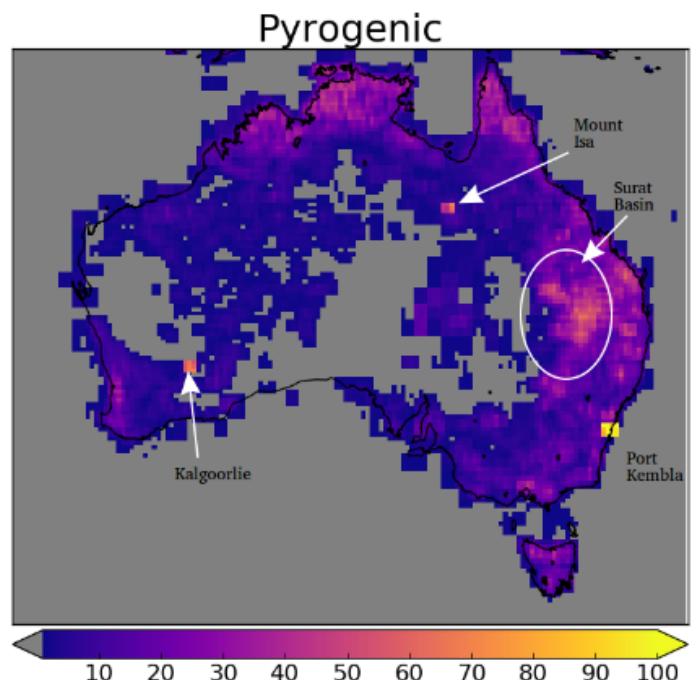


FIGURE 2.36: Top: Portion of 2005 filtered out by fire and smoke masks.  
Bottom: portion filtered out each day from land squares in Australia.

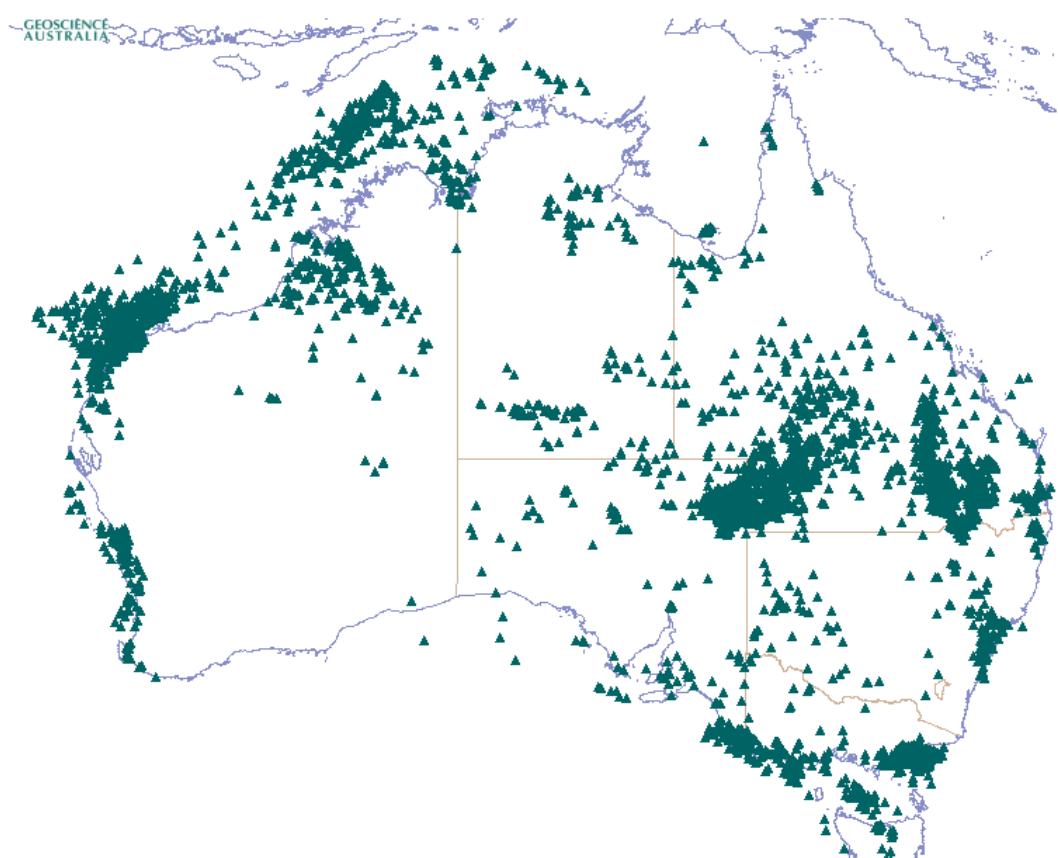


FIGURE 2.37: Petrol Well locations over Australia (current-2018) (<http://dbforms.ga.gov.au/www/npm.well.search>)

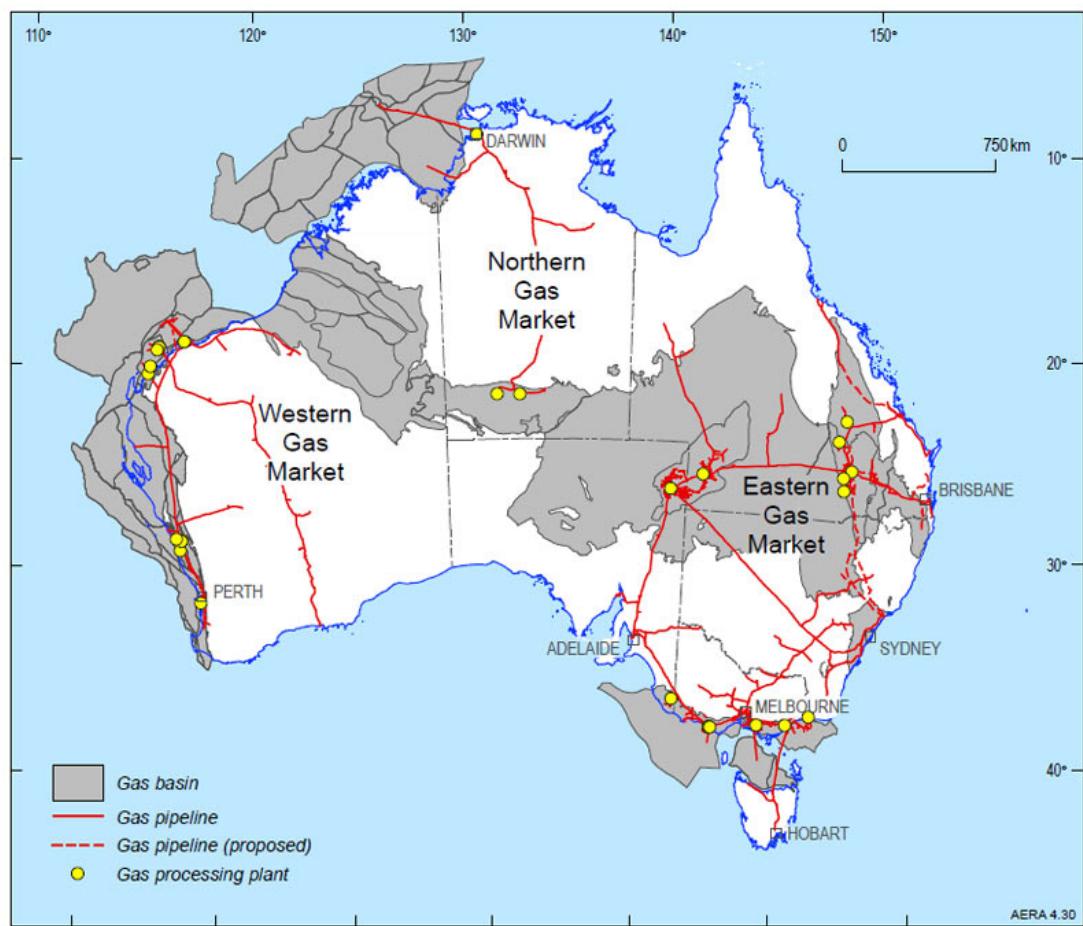


FIGURE 2.38: Gas fields and pipelines (2018) for Australia  
(<http://www.ga.gov.au/scientific-topics/energy/resources/petroleum-resources/gas>)

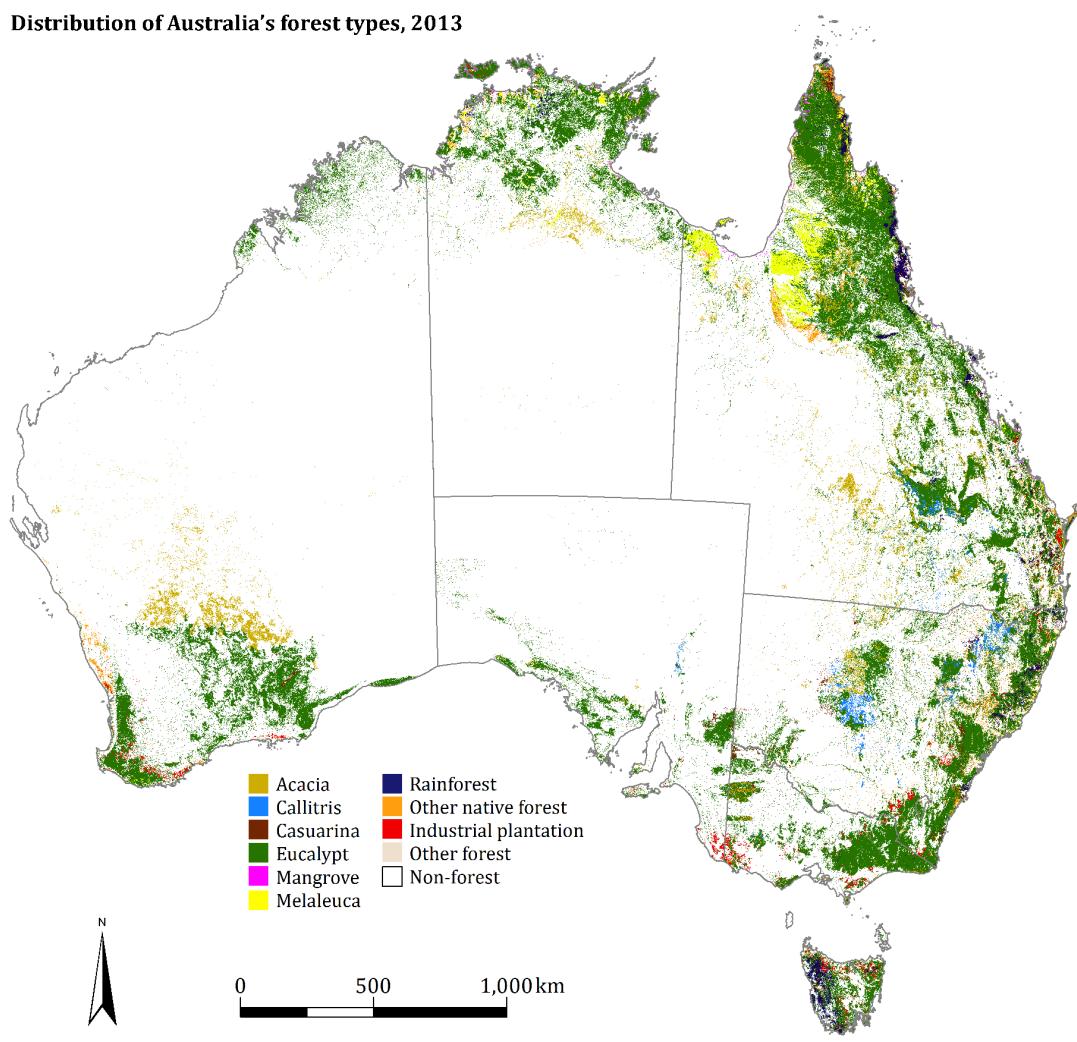


FIGURE 2.39: Forest coverage, coloured by predominant tree species.

??

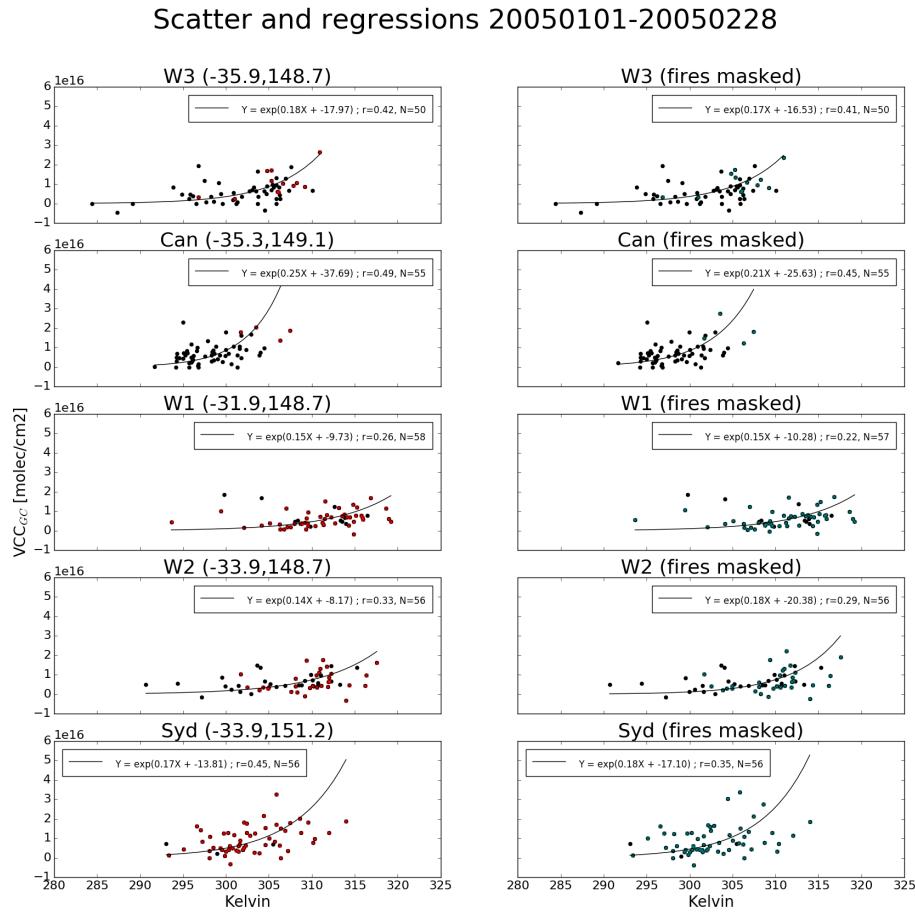


FIGURE 2.40: HCHO vs GEOS-Chem daily midday temperatures  
TODO add longer caption

## Scatter and regressions (CPC temperature) 20050101-20050228

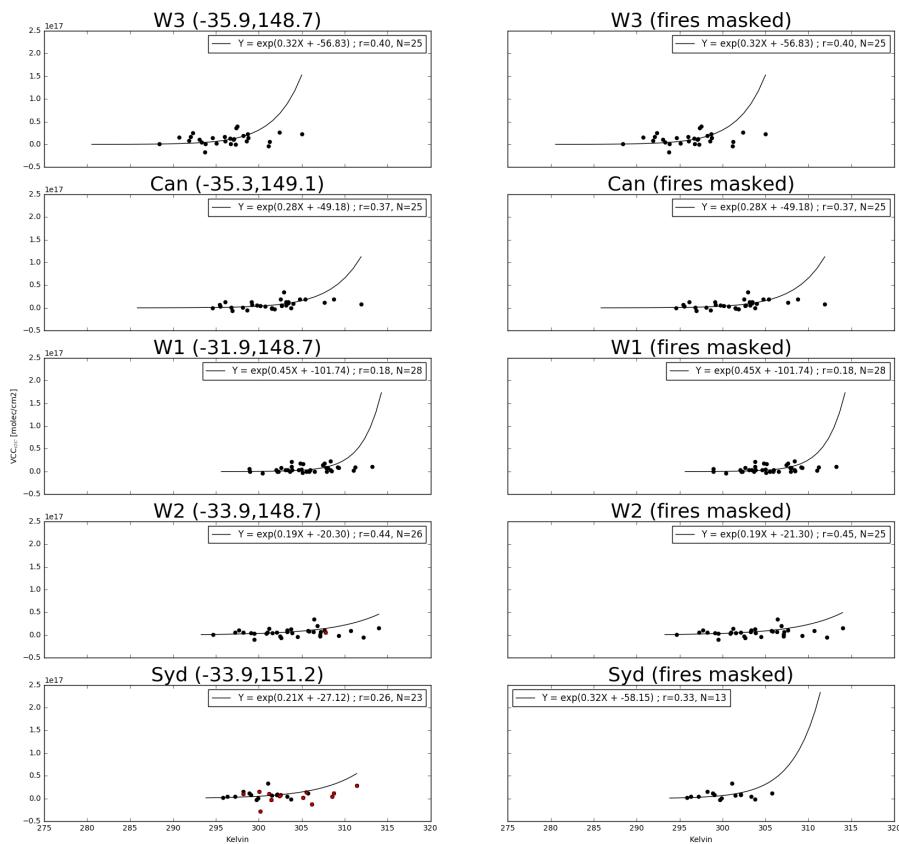


FIGURE 2.41: HCHO vs CPC daily maximum temperatures TODO add longer caption

modelled temperature and satellite HCHO with and without applying the filters for smoke and active fires. Figures TODO-TODO show the exponential fits for one month of datapoints (January 2005) in Northern, Southeastern, and Southwestern Australia respectively. Each grid square (2 by  $2.5^\circ$ ) provides one datapoint per day, with satellite HCHO initially averaged onto the lower resolution of the GEOS-Chem modelled surface (from 0 to  $\sim 100$  m) temperature. The scatter between HCHO and temperature is coloured by fire counts, and we see TODO. The TODO lines show the exponential fit before and after filtering fire and smoke. TODO plot showing how fire mask affects HCHO - Temperature relationship

### 2.7.2 NOx

Enhanced  $\text{NO}_2$  concentrations can indicate anthropogenic influence over Australia. In order to filter out these influences on satellite HCHO measurements, a filter is designed using the OMNO2d product which includes tropospheric  $\text{NO}_2$  columns.

OMNO2d from 2005 is used to determine a suitable threshold for anthropogenic influence by looking at  $\text{NO}_2$  columns near several major cities in the south eastern sector of Australia. The mean, standard deviation, and time series over Australia of tropospheric  $\text{NO}_2$  seen by Aura is shown in figure 2.45. The average tropospheric  $\text{NO}_2$  column averaged within all of Australia and then each region shown in this figure is listed in table TODO 2.4.

Anthropogenic influences on the  $\text{NO}_2$  columns are clearly visible near major cities in Australia. A filter is created each year from the OMNO2d product in two steps:

1. Daily gridsquares with  $\text{NO}_2$  greater than  $10^{15}$  molec  $\text{cm}^{-2}$  are flagged as anthropogenic.
2. After taking the yearly average over Australia, any gridsquares greater than  $1.5 \times 10^{15}$  molec  $\text{cm}^{-2}$  are flagged for the whole year.

This removes both the gridsquares close enough to cities to be affected by their emissions year round, as well as effects from transported pollution plumes. The affects of applying this filter to the OMNO2d product itself can be seen in figure 2.46

The same regions as in figure 2.45 are shown again in figure 2.47, with  $\text{NO}_2$  pixels densities for each region shown, along with the threshold of  $1 \times 10^{15}$  molec  $\text{cm}^{-2}$ . This led to a reduction of TODO gridsquares from the total available measurement space over Australia. The removal of gridsquares which went above the yearly averaged limit of  $1.5 \times 10^{15}$  molec  $\text{cm}^{-2}$  further reduced the available data by TODO gridsquares.

## 2.8 Data Access

TODO: ADD MORE HERE

**OMHCHO** Satellite swaths of HCHO slant columns downloaded from <https://search.earthdata.nasa.gov>, with DOI 10.5067/Aura/OMI/DATA2015

**OMNO2d** Daily satellite  $\text{NO}_2$  product downloaded from <https://search.earthdata.nasa.gov/search>, DOI:10.5067/Aura/OMI/DATA3007. For more information in refer to section

Temperature 20050101-20050228

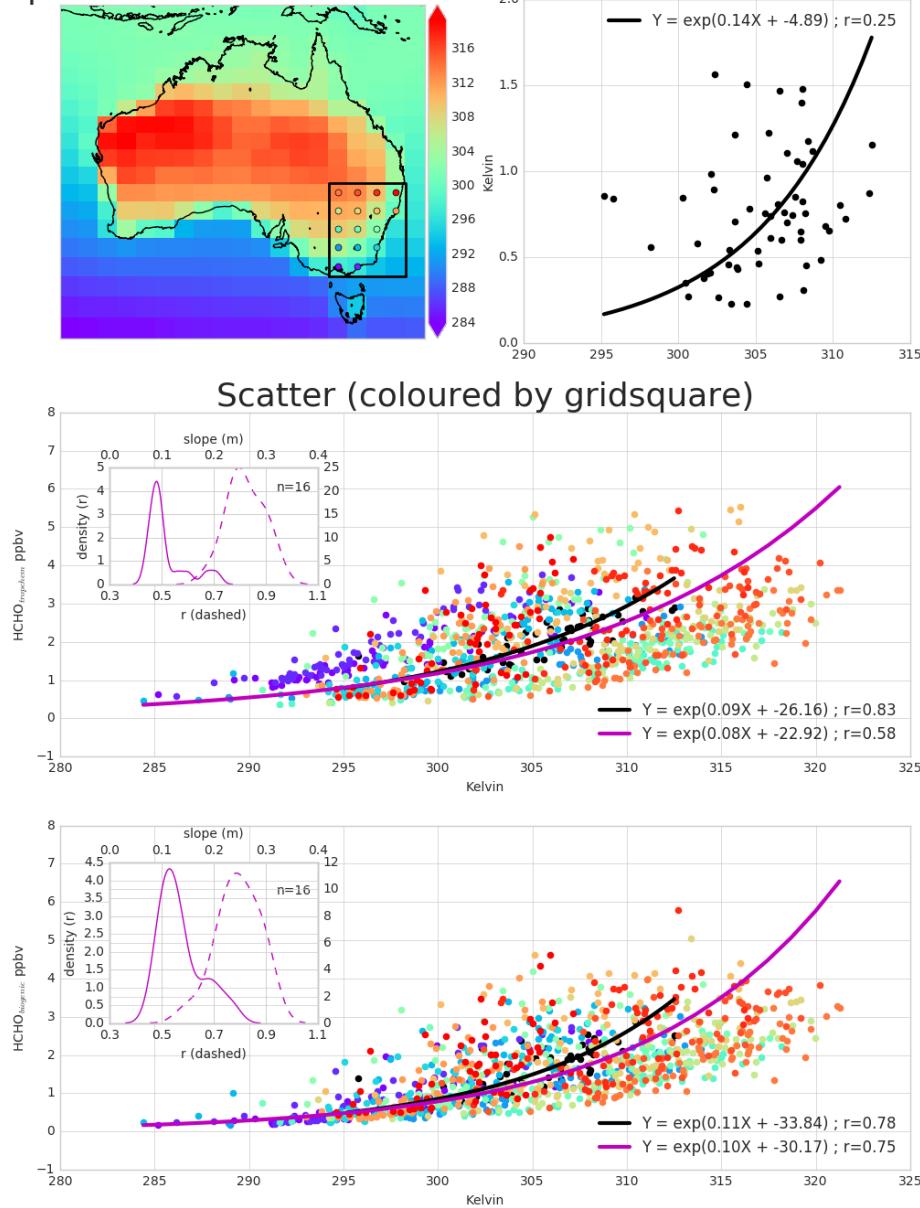
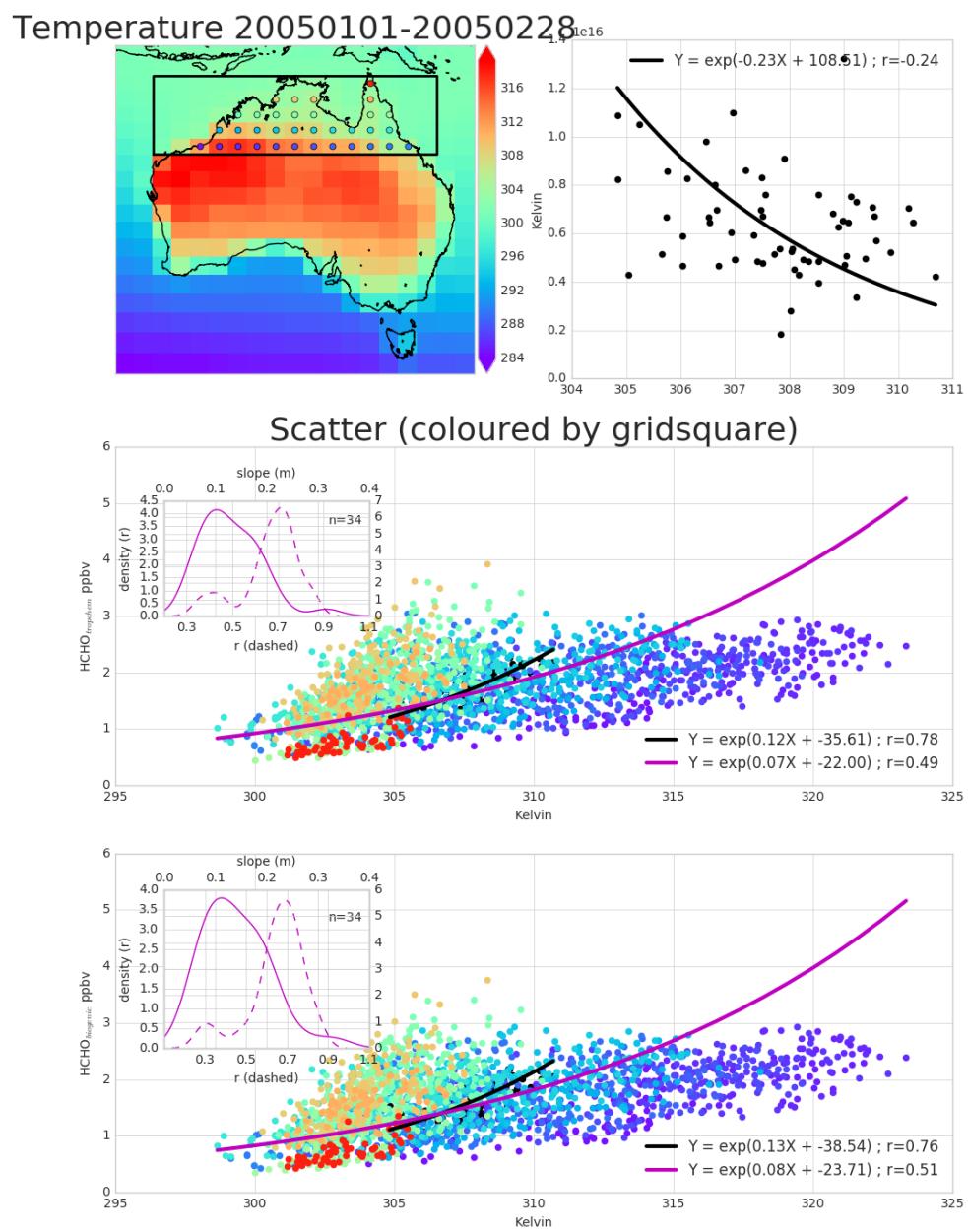


FIGURE 2.42: Top panel: surface temperature averaged over January and February 2005. Bottom panel: surface temperature correlated against temperature over, with different colours for each gridbox, and the combined correlation. A reduced major axis regression is used within each gridbox (shown in top panel) using daily overpass time surface temperature and HCHO amounts (ppbv). The distribution of slopes and regression correlation coefficients (one datapoint per gridbox) for the exponential regression is shown in the embedded plot.



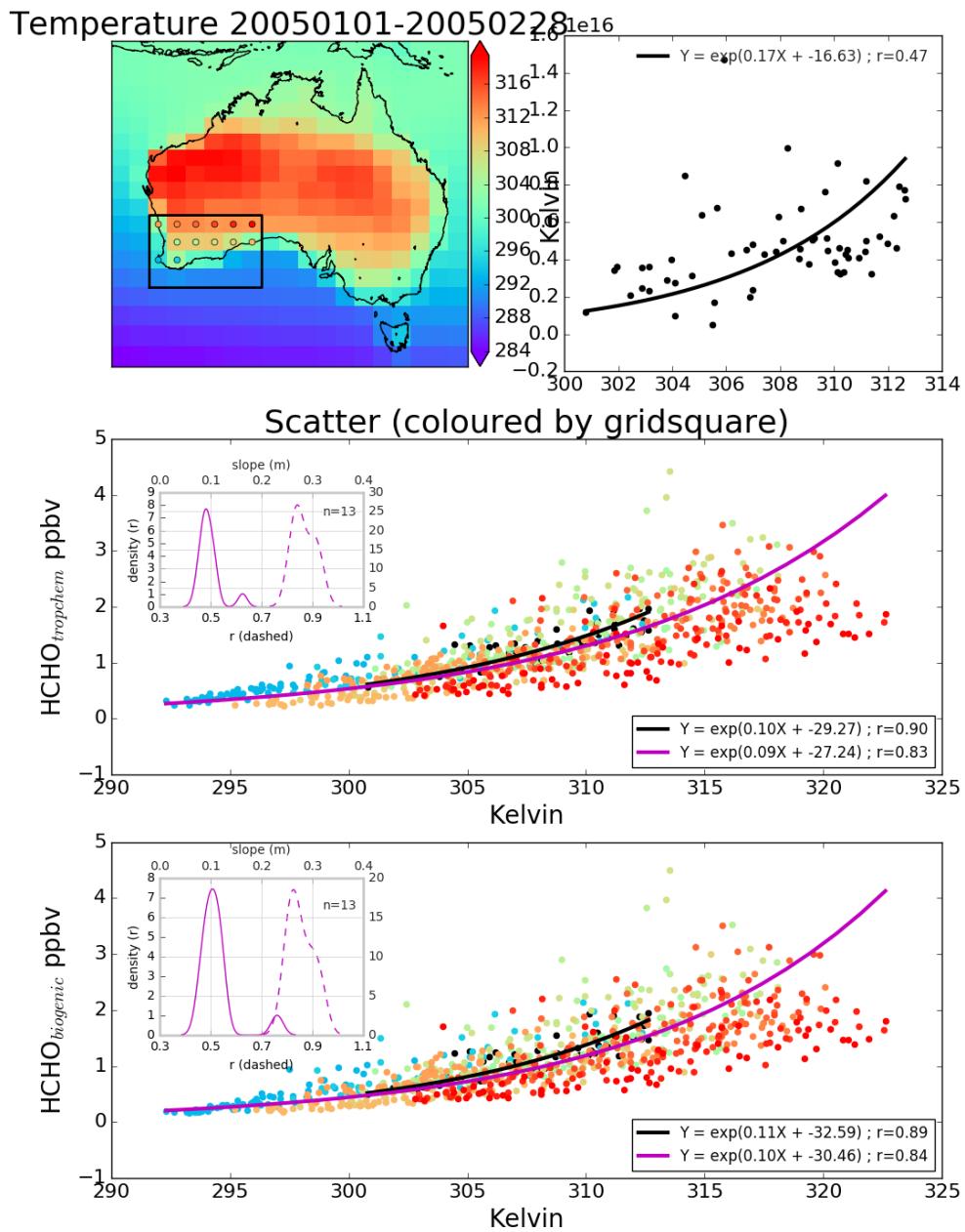


FIGURE 2.44: As figure 2.42 but for south-western Australia.

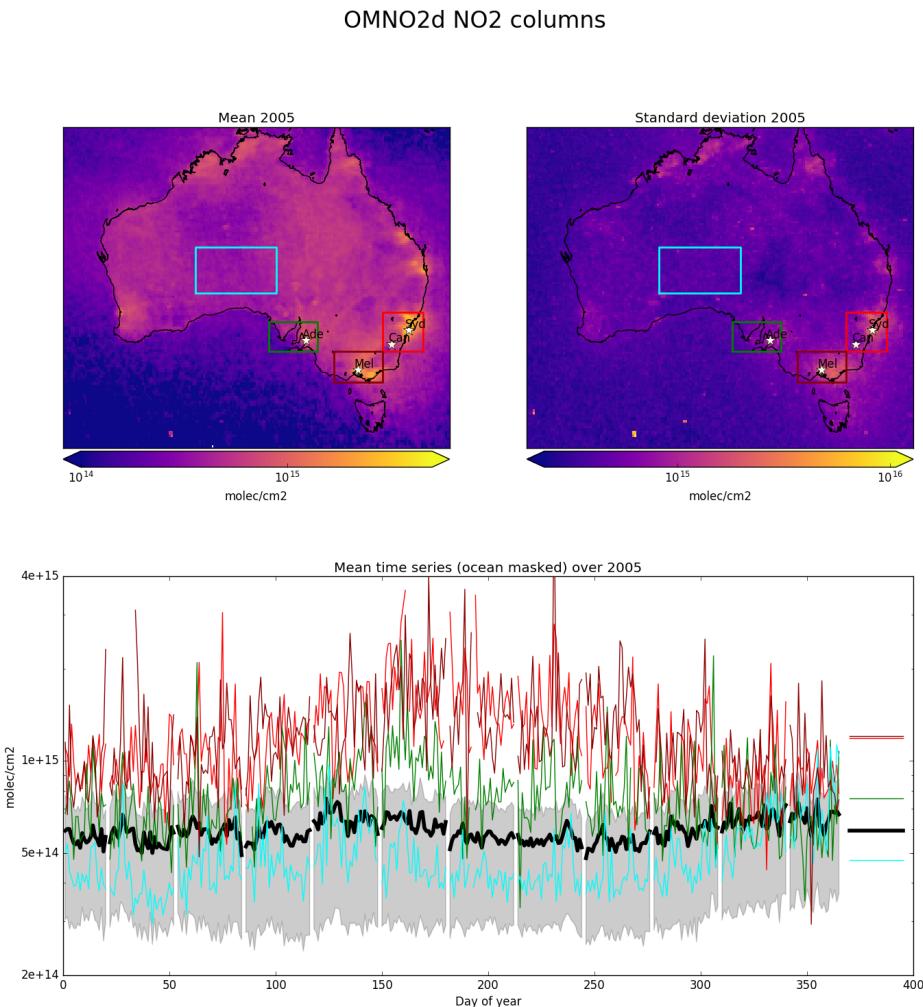


FIGURE 2.45: Mean (top left) and standard deviation (top right) of OMNO2d daily  $0.25 \times 0.25^\circ$  tropospheric cloud filtered NO<sub>2</sub> columns. Time series for Australia, and each region (by colour) shown in the bottom panel, with mean for that region shown on the right. A grey shaded area depicts the 25th to 75th percentiles of Australia averaged NO<sub>2</sub> columns for each day in the time series, with a thicker black line showing the Australia-wide mean value.

TABLE 2.4: NO<sub>2</sub> averages by region before and after filtering for anthropogenic emissions using 2005 data from the OMNO2d product.

Region	NO <sub>2</sub>	NO <sub>2</sub> after filtering	% Data lost
Aus	1	2	3
BG	1	2	3
Syd	1	2	3
Melb	1	2	3
Adel	1	2	3

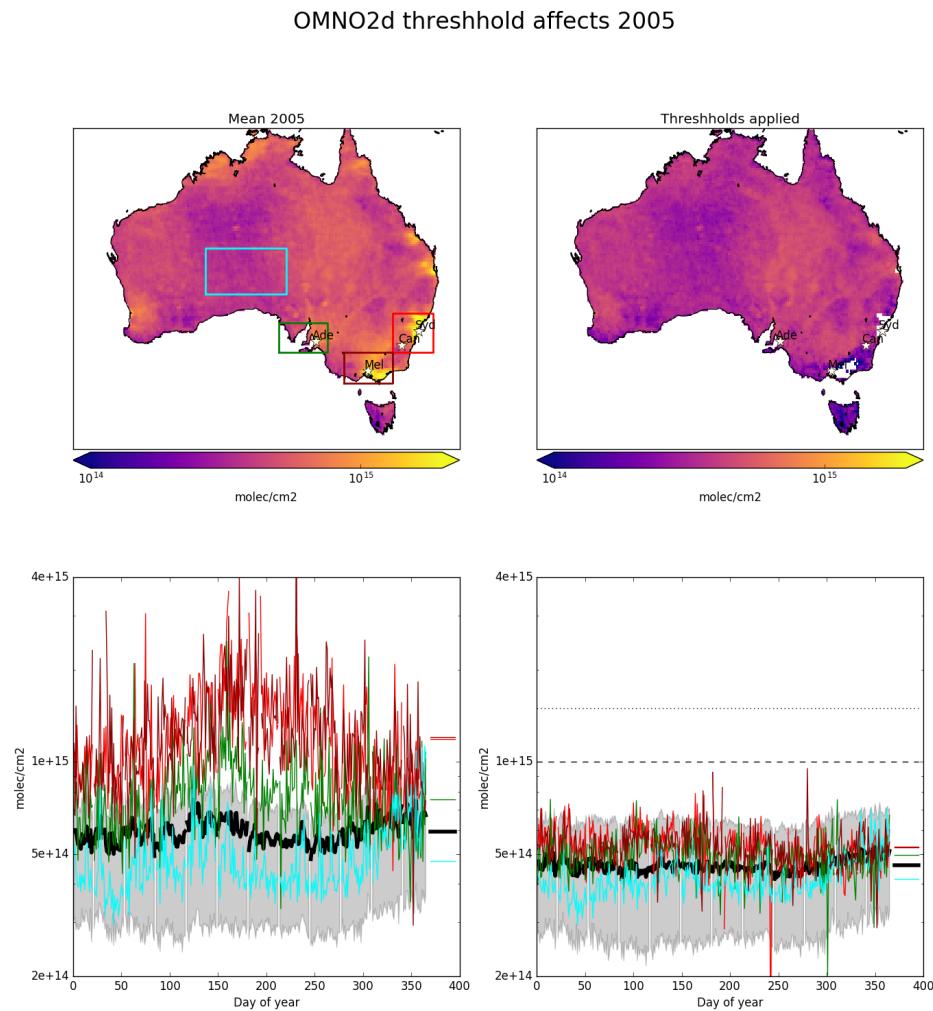


FIGURE 2.46: 2005 OMNO2d NO $_2$  column mean before (left) and after (right) applying the threshold filters as described in the text. Time series for Australia, and each region (by colour) shown in the bottom panel, with mean for that region shown on the right.

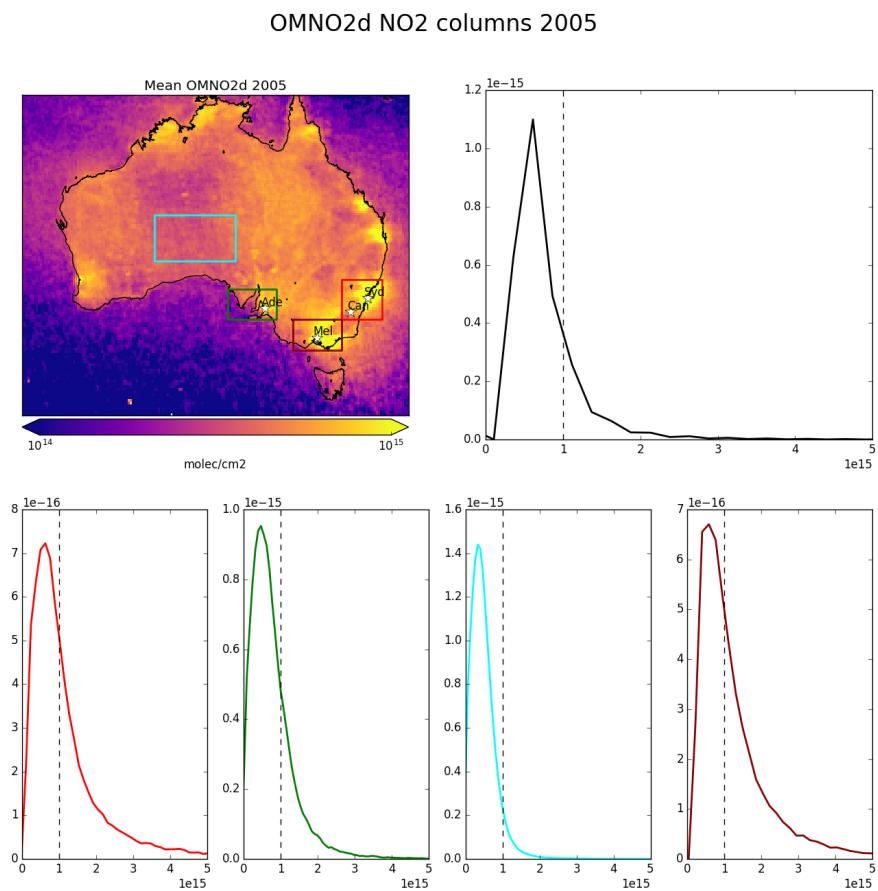


FIGURE 2.47: 2005 OMNO2d NO<sub>2</sub> column means (top left), along with column amount distributions for Australia (top right) and each region shown in the area map (by colour)

**OMAERUVd** Gridded satellite based AAOD measurements downloaded from <https://search.earthdata.nasa.gov>, DOI 10.5067/Aura/OMI/DATA3003.

**SPEI** Monthly standardised precipitation evapotranspiration index (metric to determine drought stress) downloaded from <http://hdl.handle.net/10261/153475> with DOI:10.20350/digitalCSIC/8508. See more information in section