

Chapter 1

Data and Modelling

1.1 Introduction

Models of ozone in the atmosphere are used broadly for international assessments of ozone related emissions ([Young2018](#)). [Young2018](#) summarise current global ozone modelling standards and the metrics and processes used to evaluate these models. They show how models can be used to improve measurements, estimate concentrations in regions not sampled, and allow analysis of other processes which involve ozone (such as radiation).

Satellite datasets give us access to lots of datapoints. When looking at trace gases such as HCHO, satellite measurements can have high amounts of uncertainty. Many datapoints can be averaged in order to reduce uncertainty. Once the satellite data are quality filtered and gridded, I use additional data to account for anthropogenic and pyrogenic influences, which provides us an estimate of biogenic HCHO. MODIS fire counts are used in conjunction with smoke AAOD enhancements (also measured by satellite) to remove data points which may be affected by fires. This has a negative affect on uncertainty, as fewer measurements are available to be averaged. The affect of the filtering on uncertainty, and how many points are removed is shown in section [1.7](#).

1.2 Datasets

Here I will describe the various datasets I've used to analyse GEOS-Chem output. I will also give details on filtering and groupings which are undertaken when reading data, as each dataset has its own resolution.

These datasets are used my thesis to determine isoprene emissions estimations in chapter ?? and ozone transport extrapolations in chapter ??.

Figure [1.1](#) shows the locations of each of the campaigns I mention in this text. These took place over disparate times, and are in-situ datapoints which may not nicely compare with GEOS-Chem output which is averaged over a large horizontal space.

TODO: these summaries.

1.2.1 Daintree

Daintree summary (P. Nelson)



FIGURE 1.1: Locations of Australian campaigns which are analysed within this thesis

1.2.2 Marine and Urban MBA ? (MUMBA)

1.2.3 Sydney Particle Studies (SPS1, SPS2)

Two VOC and other trace gas measurement campaigns took place at the Westmead Air Quality Station scientists from CSIRO, OEH, and ANSTO. Stage 1 (SPS1) was from 5 February to 7 March in 2011, while stage 2 (SPS2) ranged from 16 April to 14 May 2012.

Two instruments measured VOC concentrations: one was a Proton transfer reaction mass spectrometer (PTR-MS), the other a gas chromatographer (GC) with an equipped 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, α -pinene, and several more VOCs. Further details can be found in [Dunne2012](#), [Dunne2017](#) (TODO: Check papers). 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's 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 usage starts to become frequent (TODO cite something). The GC-FID analysed samples collected in multi-absorbent tubes, with lower temporal resolution but no interferences. GC-FID data is averaged from 0500-1000 LT, and 1100-1900 LT. Further details for this method can be found in TODO: cite Min et al 2016.

Figure 1.2 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 a daily average and an overpass time (1200-1300 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 (TODO: doi:10.5194/acp-15-223-2015, 2015) which should minimise any introduced bias.

Figure 1.3 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 daily averaged 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.

1.2.4 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 using an AMF calculated through radiative transfer models. Level 3 data is a temporally aggregated version of the this, for instance monthly or yearly averages.

One satellite is NASA's Earth Observing System's "Aura", which provides several useful datasets, often called products. Aura orbits the earth in a polar sun-synchronous pattern, circling the earth on a plane coincident with the sun and the poles. OMI measurements are used to map several atmospheric trace gases, including NO₂, SO₂, BrO,

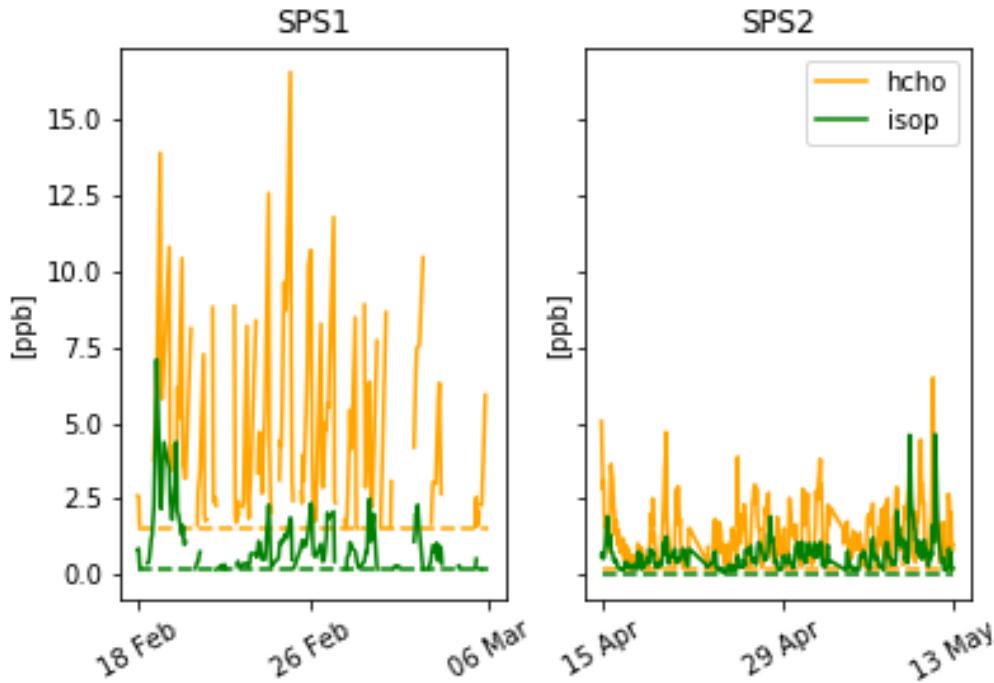


FIGURE 1.2: SPS 1 and 2 HCHO (yellow) and isoprene (green) time series, along with detection limits (dashed).

HCHO, O₃, 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 km² at nadir to 26x135 km² at the swath edges (**Abad2015**). The swaths cover Earth daily, although half of these are at night time and contain no useful near-UV/Visible information.

Aura houses the Ozone Monitoring Instrument (OMI), a near-UV/Visible Charged Coupled Device (CCD) spectrometer. 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. OMI spectra are used in several products used in this thesis, including OMNO2d, OMHCHO, and OMAERUVd.

OMNO2

NO₂ measured by OMI is used to check whether NO₂ is well represented by GEOS-Chem. OMNO2d is a gridded daily level three product with good satellite pixels averaged into 0.25x0.25° horizontally resolved bins. An example figure from Jan 29, 2005 is shown in figure 1.4.

See section ?? for the comparison between this product and GEOS-Chem calculations.

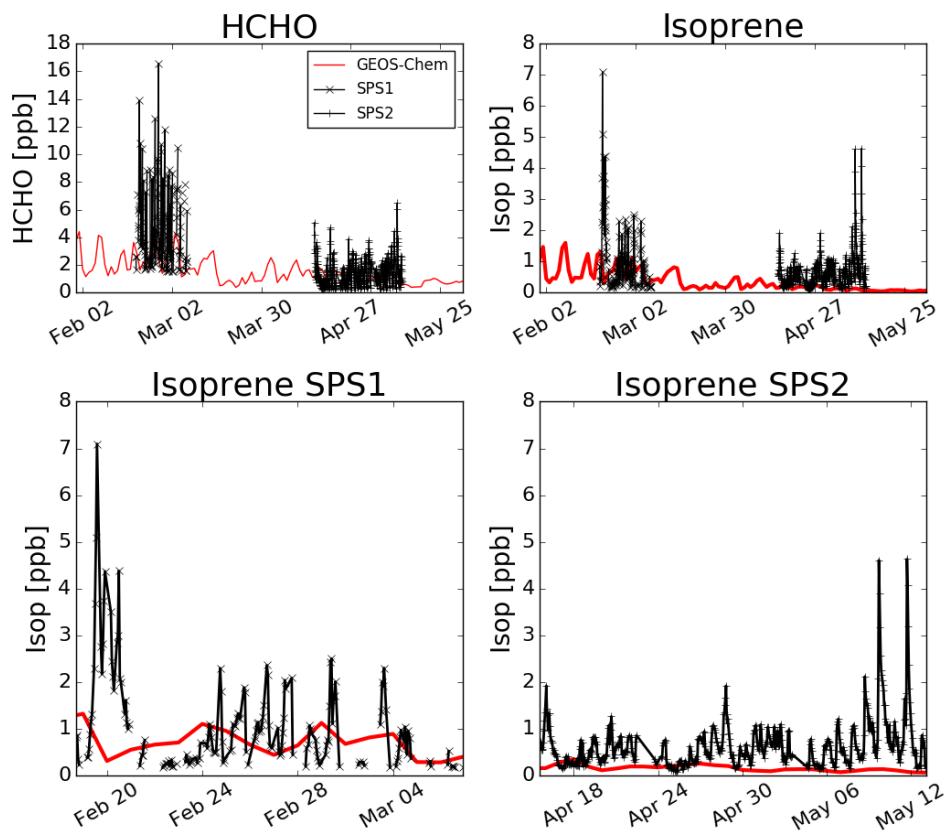


FIGURE 1.3: Comparison between GEOS-Chem HCHO concentrations in the gridsquare containing Sydney for the duration of the SPS 1 and 2 campaigns

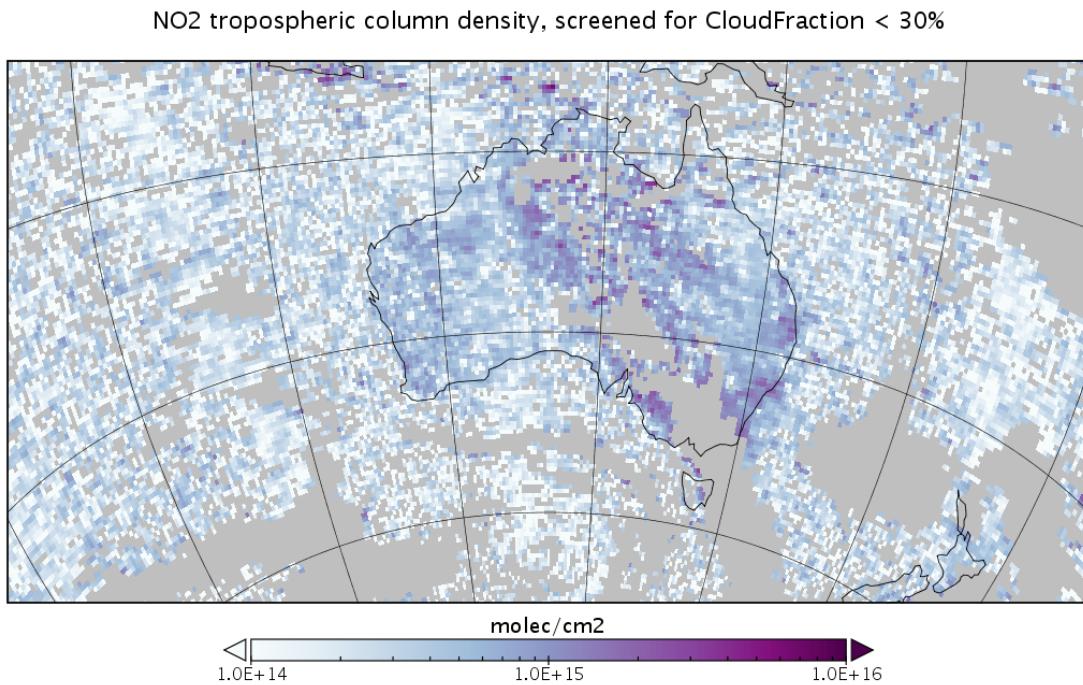


FIGURE 1.4: Example of NO₂ tropospheric columns taken from the OMNO2d product.

OMHCHO

Atmospheric HCHO can be measured using Differential Optical Absorption Spectroscopy (DOAS), as long as trace gases with similar features near the same wavelength 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 OMHCHOv003 data product (Abad2015), 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 (Ω_S) are actually the difference with respect to the radiance reference column (Ω_{S_0}). The full method details for slant column retrieval by OMI are outlined in section ???. Slant columns range from $\sim 4 \times 10^{15}$ to $\sim 6 \times 10^{16}$ molec cm⁻², with uncertainties from 30% (larger columns) to over 100% (smaller columns) (Abad2015).

OMHCHO level two data includes 14-15 daily swaths of measurements provided by NASA. Each swath contains roughly 9×10^4 pixels, each of which contains various data including latitude, longitude vertical column HCHO, etc. 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 1.1 which is taken from Kurosu2014 Filtering bad or missing measurement pixels is preformed prior to any other filtering, this includes the datapoints affected by the row anomaly. This anomaly (<http://projects.knmi.nl/omi/research/product/rowanomaly-background>.

TABLE 1.1: OMI quality flag values table from **Kurosu2014**

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"> • Fit convergence flag is < 300 but > 0: Convergence at noise level • Column $+2\sigma$ uncertainty $< 0 <$ Column $+3\sigma$ uncertainty • Absolute column value $>$ Maximum column amount ($1e19$ molec cm^{-2})
2	Bad	Avoid using as one of the following conditions are present: <ul style="list-style-type: none"> • Fit convergence flag is < 0 : No convergence, abnormal termination • Column $+3\sigma$ uncertainty < 0
< 0	Missing	No column values have been computed; entries are missing

[php](#)) 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 easy to remove before further processing.

The cloud fraction with each pixel is provided with the OMHCHO dataset, however its source is the OMI cloud product, OMCLDO2. To give an idea of how much data is filtered out, around 30% of the pixels which remain after filtering out the bad or missing data are subsequently removed due to cloudiness.

Each ~ 90 minutes the AURA satellite sweeps over the sunny side of the planet, with OMI recording roughly 90 k pixels, of which around 50 k – 80 k 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 (molecules 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 apriori shape factor rather than NASA's included apriori shape factor.

Recalculated OMI formaldehyde columns are used as a basis for estimating isoprene emissions in Chapter ??.

OMAERUVd

Aerosols in the atmosphere can be seen through their affects on light. Smoke and dust can be seen as an increase in AAOD (see section 1.5.1). This is due these particles scattering and absorbing UV radiation (**Ahn2008**).

OMAERUVd (DOI: 10.5067/Aura/OMI/DATA3003) provides a useful dataset allowing us to filter gridsquares which may be smoke affected. OMI aerosol extinction and absorption optical depths (AOD, AAOD respectively) at three wavelengths (354, 388, and 500 nm), along with UV aerosol index (UVAI), are available publicly from Earthdata: https://disc.gsfc.nasa.gov/datasets/OMAERUVd_V003/summary. 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 subpixel scale cloud interference, so I select AAOD as the basis for my smoke filter as it is least affected by clouds (**Ahn2008**).

Gridded smoke AAOD is read from OMAERUVd at 1x1 °resolution daily, and mapped to finer resolution using the nearest value for each gridsquare, before being saved along with the MOD14A1d fire product as seen in Figure 1.13. I use the AAOD at 500 nm wavelength, which should be sufficiently blocked by fire smoke plumes. This daily AAOD is used to mask fire smoke plume influence, by masking gridsquares with higher than TODO AAOD values.

MOD14A1

Daily gridded fire counts compiled from Terra and Aqua satellite into 1x1 km² resolution.

1.2.5 Drought Index

The S Precipitation Evapotranspiration Index (SPEI) is a measure of drought using various parameters such as TODO. (**Wang2017**). SPEI will be compared against the difference between top-down estimated emissions and MEGAN bottom up estimated emissions. This is used to determine whether there are biases in the MEGAN calculations due to the GEOS-Chem implementation ignoring soil moisture. It is downloaded from TODO and holds monthly averaged values at 0.5° horizontal resolution. When comparing against the emissions estimates this is interpolated linearly onto the same grid as that of GEOS-Chem output at 2x2.5°.

1.2.6 NPI

One possible solution to anthropogenic filtering is the national pollution index (TODO:cite:<http://www.npi.gov>), which contains industrial HCHO and NO_X emissions from 2003 to 2014.

1.2.7 GEOS-Chem output

There are various outputs available when running GEOS-Chem, which require understanding in order to compare with observations.

Satellite overpass is output from averaging over a window of local time for each gridbox. This output allows comparison with satellite measurements, which overpass at the same local time every day. The output is in bitpunch.

HEMCO diagnostics are the emissions TODO: averaged or instantaneous? in each gridbox, which I've stored for each 3 hours. This output is netcdf.

Tracer averages are daily or monthly averaged gridbox concentrations. This output is bitpunch.

1.3 Modelling

1.3.1 GEOS-Chem simulation

The GEOS-Chem global atmospheric chemistry model (V10.01) simulates and records 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). GEOS-5 meteorological fields from NASA's ... (TODO: ref and note) are used to drive transport and coupled with the chemical module of GEOS-Chem. MEGAN is used to determine biogenic emissions for our default GEOS-Chem simulation, with subsequent modifications based on top-down estimates made herein.

Output for an area averaged over 1300 - 1400 local time is saved for comparison and recalculation with satellite overpass records. These averages are used to calculate both the GEOS-Chem based AMF, and the modelled background HCHO over the remote pacific which is used in the reference sector correction for OMI column retrievals. They are also used to determine isoprene to HCHO yield, after removing days with high biomass burning emissions.

1.3.2 List of runs and outputs used in my work TODO: good place for this?

TODO: Go through work process and clarify these items

1.3.3 GEOS-Chem

1. UCX
 - (a) Satellite output (1300-1400LT)
 - (b) Create shape factors for AMF recalculation in OMI
2. Tropchem (standard)
 - (a) satellite output, daily tracer averages
 - (b) Recreate the AMFs for OMI when running code from Dr. Paul Palmer, modified by Dr. Luke Surl.
 - (c) Combined with an identical run where isoprene emissions are halved in order to determine smearing.
 - (d) TODO: Compare total yearly isoprene emissions before and after new estimate.

- 3. Tropchem(isoprene emissions halved)
 - (a) Satellite output used to determine smearing.
- 4. Tropchem(biogenic emissions only, all other inventories turned off)
 - (a) Satellite output, hourly biogenic emissions from MEGAN
 - (b) Used to determine yield for new emissions estimates
 - (c) TODO: compared to run with updated emissions

NB: for non-UCX runs, satellite output was modified to include tropopause height

1.3.4 CAABA/MECCA

1.3.5 Reading Data

CAABA/MECCA outputs

The box model can output in netcdf or text format, TODO: which way am I better off ? Text output from CAABA/MECCA was read using tailored python scripts modified from code written by dr. Luke Surl. Dr. Luke Surl also wrote the code which implements calculations of yield from runs using isoprene injections as described in Section ?? TODO: update to more specific reference.

GEOS-Chem Satellite output

HEMCO diagnostics

In order to get hourly MEGAN modelled isoprene emissions, HEMCO (the module of GEOS-Chem dealing with emissions inventories) diagnostic output was created. When working with globally gridded data, handling local time offsets becomes more important. The hourly output emissions of isoprene is saved using GMT, which needs to be offset based on longitude in order to retrieve local time. I do this by setting up a latitude by longitude array which matches the horizontal resolution of the data, filling each gridbox with it's local time offset. This offset is determined as one hour per 15 degrees (since 360 degrees 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.

1.4 GEOS-Chem

GEOS-Chem is an atmospheric chemical model (ACM), using a 3-D grid of boxes with transport driven by the GEOS meteorological model and chemistry calculated in each box independently. Many of these terms are described in Section ??.

GEOS-Chem uses many boxes covering the globe, each with chemistry and dynamic meteorological conditions. Different meteorological conditions such as wind

and air pressure need to be handled within each box. GEOS-Chem has a meteorological model coupled to a chemical model, which simulates the world in a three dimensional grid of connected boxes.

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). 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. Several driving meteorological fields exist with different resolutions, the finest at 0.25 by 0.3125° horizontally at 5 minute time steps with 72 vertical levels.

GEOS-Chem in this work runs with a 15 minute time step for both chemistry and transport, at 2x2.5 °horizontal resolution over 47 vertical levels. Generally output is the average of these time steps either over an entire month, or else per day. Additionally output averaged between 1300-1400 LT is saved to allow comparisons with Aura satellite measurements, as Aura overpasses at ~1330 LT each day.

Global CTMs are often run using one or several emission models (or the output from them) to determine boundary conditions for many gridboxes. TODO: is this the case? Doesn't GEOS-Chem have coupled chemistry and meteorology? Check the wiki. GEOS-Chem has boundary conditions based on several meteorological and emissions inventories, the following are the versions of theses used by GEOS-Chem v 10.01. Meteorological fields can be driven by NASA's GEOS-5 data (0.5° x 0.666°) (**Chen2009**), which exists up to 2013, or GEOS-FP data (0.25° x 0.3125°). Fire emissions come from the GFED4 product (**Giglio2013**). Anthropogenic VOC emissions come from the EDGAR inventory, while biogenic VOC emissions are coupled to the MEGAN model TODO:cites. The estimated biogenic VOC emissions are important for accurately simulating chemistry within models, as discussed in Sections ?? and ??.

1.4.1 GEOS-Chem isoprene modelling

Outline

The isoprene reactions simulated by GEOS-Chem were originally based on **Horowitz1998**. This involved simulating NO_x, O₃, 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 **Mao2013** who change the isoprene nitrates yields and add products based on current understanding as laid out in **Paulot2009a; Paulot2009b**. Further mechanistic properties, like isomerisation rates, are based on results from four publications: citeCrounse2011,Crounse2012,Peeters2010 (TODO: check abstracts Peeters papers). **Crounse2011** examines the isomerisations associated with the oxidation of isoprene to six different isomers (ISO₂) formed in the presence of oxygen: isoprene + OH →^{O₂} ISO₂. They determine rates and uncertainties involved in these reactions, and study the rate of formation of C₅-hydroperoxyaldehydes (HPALDs) by isomerisation. **Crounse2012** examine the fate of methacrolein (MACR),

one of the products of isoprene oxidation. Prior to this work MACR oxidation chamber studies were performed in high NO or HO₂ concentrations, giving peroxy lifetimes of less than 0.1 s. In most environments this is not the case, and MACR products over various NO concentrations and peroxy radical lifetimes are determined in their work. **Peeters2010** examine photolysis of hydroperoxy-methyl-butinals (HPALDs, produced by isoprene isomerisation), which regenerates OH levels in areas with high isoprene emissions. Additionally, photolysis of photolabile peroxy-acid-aldehydes creates OH and improved model agreement with continental observations. The 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 TODO: citation.

Formation of isoprene nitrates have an effect on ozone levels through NO_X sequestration, and the yields and destinies of these nitrates is analysed in **Paulot2009a**. They use anion chemical ionization mass spectrometry (CIMS) to determine products of isoprene photooxidation. In a chamber with clean air and high NO concentrations, isoprene photooxidation is initially driven by OH addition, followed by NO_X chemistry (150 min - 600 min), and finally HO_X dominated chemistry. The yields of various positional isomers of isoprene nitrates is estimated, and pathways of their oxidation products is shown and used in the GEOS-Chem isoprene mechanism (**Paulot2009a; Mao2013**).

In low NO_X conditions, isoprene oxidises to yield 70% hydroxyhydroperoxides (ISOPOOH), which then oxidises to create dihydroxyperoxides (IEPOX) with OH recycling maintaining the OH levels in the atmosphere (**Paulot2009b**). In older models isoprene produced ISOPOOH which then titrated OH, however, the loss of OH has not been seen in measurements (**Paulot2009b; Mao2013**). The isoprene mechanism in GEOS-Chem has been updated to include OH regeneration from oxidation of epoxydiols and slow isomerisation of ISOPO₂ (**Mao2013**).

ISOPN can be oxidised (by OH) to form nitrated organic products (**Paulot2009a**). In low NO_X ISOPOO reacts with HO₂ (producing hydroxy hydroperoxides, ISOPOOH), RO₂ (producing mainly MACR, MVK, and HCHO), or isomerises (1,5-H shift producing MACR, MVK, HCHO, or 1,6-H shift producing hydroperoxyenals HPALDs). ISOPOOH can be oxidised (by OH) to produce epoxydiols (IEPOX), precursors to SOA (**Paulot2009b**). HPALDs can photolysse to regenerate OH and small VOCs (**Crounse2011; Wolfe2012; Peeters2014**) TODO: Check out crounse2011. See section ?? for more information.

Under high NO_X conditions, isoprene undergoes OH addition at the 1 and 4 positions, becoming β (71%) or δ (29%) hydroxyl peroxy radicals (ISOPO₂). The β -hydroxyl reacts with NO_X and produces HCHO (66%), methylvinylketone (40%) (MVK), methacrolein (26%), and β -hydroxyl nitrates (6.7%) (ISOPNB). The δ -hydroxyl reacts with NO to form δ -hydroxyl nitrates (24%) (ISOPND), and ISOPNB (6.7%). ISOPNB and ISOPND yield first generation isoprene at 4.7% and 7% respectively.

Under low NO_X conditions, ISOPO₂ may react with HO₂ to form ISOPOOH. In this case there is also production of HCHO (4.7%), MVK(7.3%), and MACR (12%). As stated in earlier; most ISOPOOH will form IEPOX (epoxydiols) after reacting with OH and lead to OH regeneration. The other mechanism in low NO_X environments is

unimolecular isomerisation of ISOPO₂. This leads to production of hydroperoxyaldehydes (HPALDS), which generally photolyse and have an OH yield of 100%. **Mao2013** show that a lower (factor of 50) rate constant for ISOPO₂ isomerisation leads to better organic nitrate agreements with ICARTT.

This update leads to more accurate modelling of OH concentrations, especially in low NO_x conditions common in remote forests. Prior to **Mao2012** 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. **Mao2012** examine an upgraded method of measurement, and compare these against a regional atmospheric chemistry model (RACM2), with the OH recycling updates from **Paulot2009b** as discussed in prior paragraphs.

The updates to isoprene chemistry by **Mao2013** and those shown in **Crounse2011**; **Crounse2012** are the last before version 11, which was not used in this work.

The full current mechanism is described online at http://wiki.seas.harvard.edu/geos-chem/index.php/New_isoprene_scheme.

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 (**Kefauver2014**). 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, an emission factor for isoprene within a forest would include the requirement of sunshine and suitable temperature. The schematic for MEGAN, taken from **Megan_Website** is shown in figure 1.5

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.” (**Guenther2012**). It allows parameterisation of various BVOC emissions, with descriptions given in **Guenther2012** Instructions to run version 2.1 are available at 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 (**Guenther2006**)) includes 147 species, in 19 BVOC classes, which can be lumped together to provide appropriate output for mechanisms in various chemical models.

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 didn’t account for abiotic stresses, such as drought, prior rainfall and development processes, although these influenced species specific emissions by more than an order of magnitude (**Niinemets1999**). Isoprene emissions were based on temperature, leaf area, and light, but have since been updated to include leaf age activity (**Guenther2000**), and a leaf energy balance model (**Guenther2006**) in MEGANv2.0.

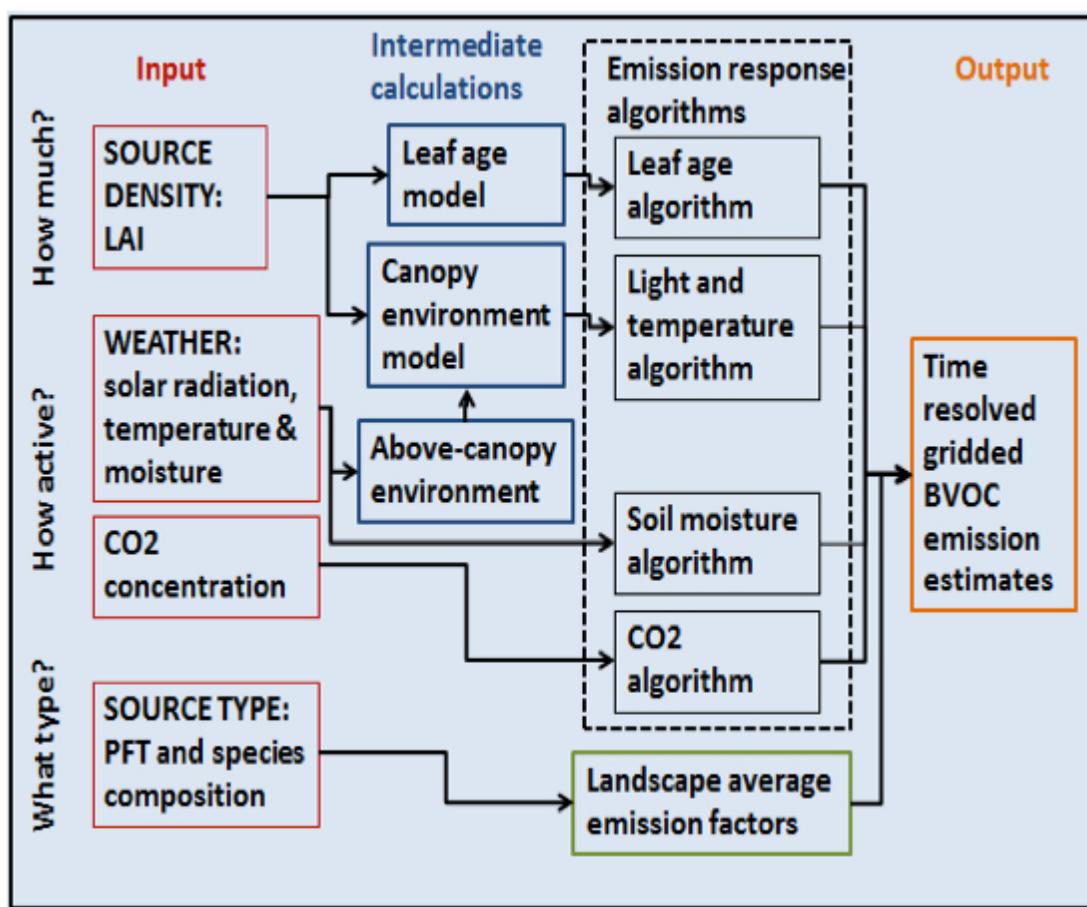


FIGURE 1.5: MEGAN schematic, copied from [Megan_Website](#)

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 (**Sindelarova2014**). Soil moisture effects on isoprene emission are very important, and can drastically affect estimates.

MEGAN has recently been analysed using 30 years of meteorological reanalysis information by **Sindelarova2014**. They estimate emissions of Biogenic VOCs (BVOCs) to be 760 Tg(C)yr^{-1} , 70% (532 Tg(C)yr^{-1}) of which is isoprene. This is similar to isoprene emission estimates from MEGAN itself, of $400\text{-}600 \text{ Tg(C)yr}^{-1}$ (**Guenther2006**). MEGAN emissions estimates are termed bottom-up, as opposed to top-down which are derived from satellite measurements of the products of various VOCs. Using GOME satellite HCHO and a Bayesian inversion technique to derive isoprene emissions, **Shim2005** estimated global isoprene emissions to be $\sim 566 \text{ TgC yr}^{-1}$. This estimate is greater than initially thought and leads to decreased ($\sim 10\%$) simulated OH concentrations to $9.5\text{e}5 \text{ molec cm}^{-3}$.

One of the important parameters in Australia is the soil moisture activity factor (γ_{SM}), which can have large regional affects on the isoprene emissions (**Sindelarova2014**; **Bauwens2016**). Generally if soil moisture is too low, isoprene emissions stop (**Pegoraro2004**; **Niinemets2010**), however in many Australian regions the plants may be more adapted to lower moisture levels. (TODO: Find cites for this - talk from K Emerson at Stanley indicated this) GEOS-Chem runs MEGANv2.1, which has three possible states for isoprene emissions based on the soil moisture (θ):

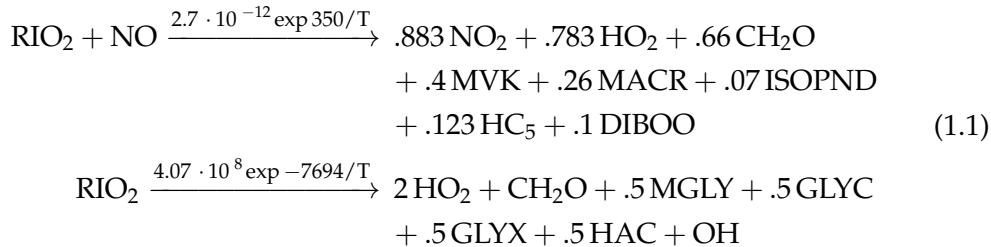
$$\begin{array}{ll} \gamma_{SM} = 1 & \theta > \theta_1 \\ \gamma_{SM} = (\theta - \theta_w) / \Delta\theta_1 & \theta_w < \theta < \theta_1 \\ \gamma_{SM} = 0 & \theta < \theta_w \end{array}$$

where θ_w is the wilting point, and θ_1 determines when plants are near the wilting point. The wilting point is set by a land based database from **Chen2001** while θ_1 is set globally based on **Pegoraro2004**

In GEOS-Chem the isoprene emissions can be globally multiplied by a constant factor. By running the model two extra times, with the biogenic isoprene emissions turned off in one run and halved in another, while other parameters remain unchanged. These modified runs allow an estimate of model sensitivity to isoprene emissions and smearing impact as described in Section ??.

1.4.2 Chemical Mechanisms

Chemical reactions are turned into systems of differential equations (DEs) to be solved by the CPU for each gridbox in GEOS-Chem. Some of the important ones involving isoprene are copied here, including reaction rates in the form $k = A \exp -ER/T$.



1.4.3 Running GEOS-Chem (before isop?)

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.

Outputs

There are several outputs or diagnostics available from GEOS-Chem. GEOS-Chem models concentrations using a 15 minute time-step, however to save space one would generally only output the daily or monthly averages for many species. In my work when estimating model yields of isoprene to HCHO, I use daily averaged HCHO columns and compare them to colocated isoprene emissions from MEGAN.

Optionally one can save high temporally resolved data for a single (or list of) column(s). I've used this diagnostic to compare modelled ozone with ozonesonde profiles at three sonde release sites discussed in Chapter ??.

One of the more frequently used outputs in my work are the satellite overpass diagnostics, which look at averages within a window of time (for instance 1200-1300) using local time (LT) for each gridbox, each day. This diagnostic allows easier analysis of model data against a satellite as one can match the output with the satellite's overpass time.

Tropospheric chemistry run

UCX run

1.4.4 Run comparisons

There are many options available when running GEOS-Chem depending on the desired chemistry, resolution, meteorology, and boundary conditions. Here we compare the model output with and without enabling the Universal tropospheric-stratospheric Chemistry eXtension (UCX). Both runs use 2° latitude by 2.5° 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.

GEOS-Chem output of HCHO does not differ much between runs with or without the Unified Chemistry eXchange (UCX). Figure 1.6 shows an example of surface HCHO amounts with and without UCX turned on. The differences do not exceed 3% over Australia for the averaged month of January, 2005.

Figure 1.7 shows the differences in surface isoprene amounts over Australia. 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 continent (4-5 orders of magnitude less), due to the short lifetime of isoprene, and lack of emissions over the oceans. Generally isoprene is 0-30% higher over Australia when the UCX mechanism is turned on. This enhancement can be seen throughout the entire tropospheric column as shown by Figure TODO fix ref ??.

Figure TODO: shows the columns for isoprene and HCHO simulated by our two mechanisms over Australia in January of 2005. The differences are minimal compared to other uncertainties in both AMF calculation and emissions estimation.

TODO: 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 1.8 shows the total column ozone between UCX and non-UCX runs, we can see that UCX has TODO: less or more ozone over Australia/USA in January.

1.4.5 Rescaling NO_x

NO_x concentrations affect HCHO yield, isoprene lifetimes, and other things due to affects on the atmospheres oxidative capacity. This means that if the model is poorly simulating NO_x, the yield (and transport, see ??) may be poorly estimated. In order to determine if rescaling the NO emissions over Australia is necessary in GEOS-Chem, I looked at modelled NO₂ amounts compared to satellite data for most of 2005.

Simulated GEOS-Chem tropospheric NO₂ columns averaged from 1300-1400 LT are compared against OMNO2d data (Sec. 1.2.4). Figure 1.9 shows the direct comparison between these datasets averaged over January to February, 2005. It's clear that the OMNO2d product can pick out Sydney and Melbourne as NO₂ hotspots, which are underestimated by GEOS-Chem (potentially due to averaging over the

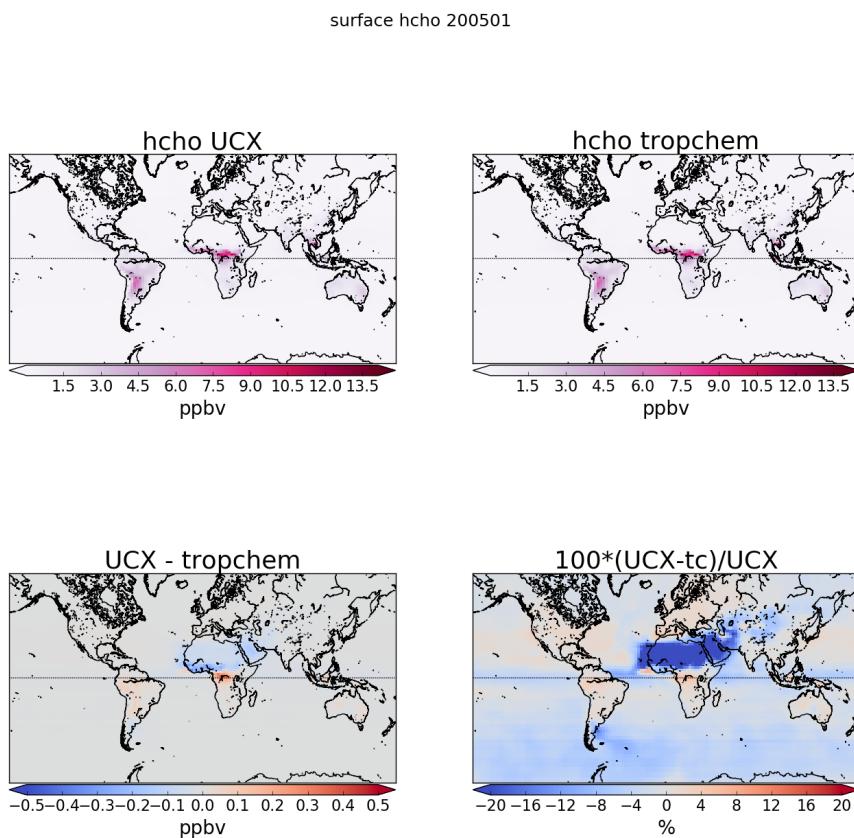


FIGURE 1.6: Surface HCHO simulated by GEOS-Chem with UCX (top left), and without UCX (top right), along with their absolute and relative differences (bottom left, right respectively). Amounts simulated by GEOS-Chem for the 1st of January, 2005.

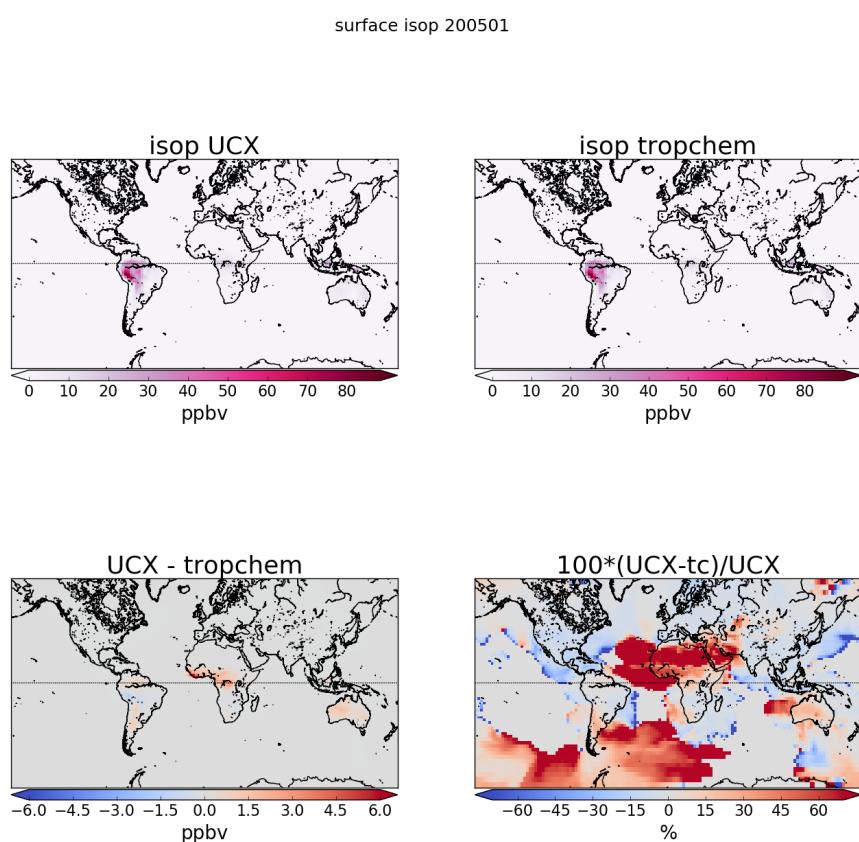


FIGURE 1.7: As figure 1.6, except looking at isoprene.

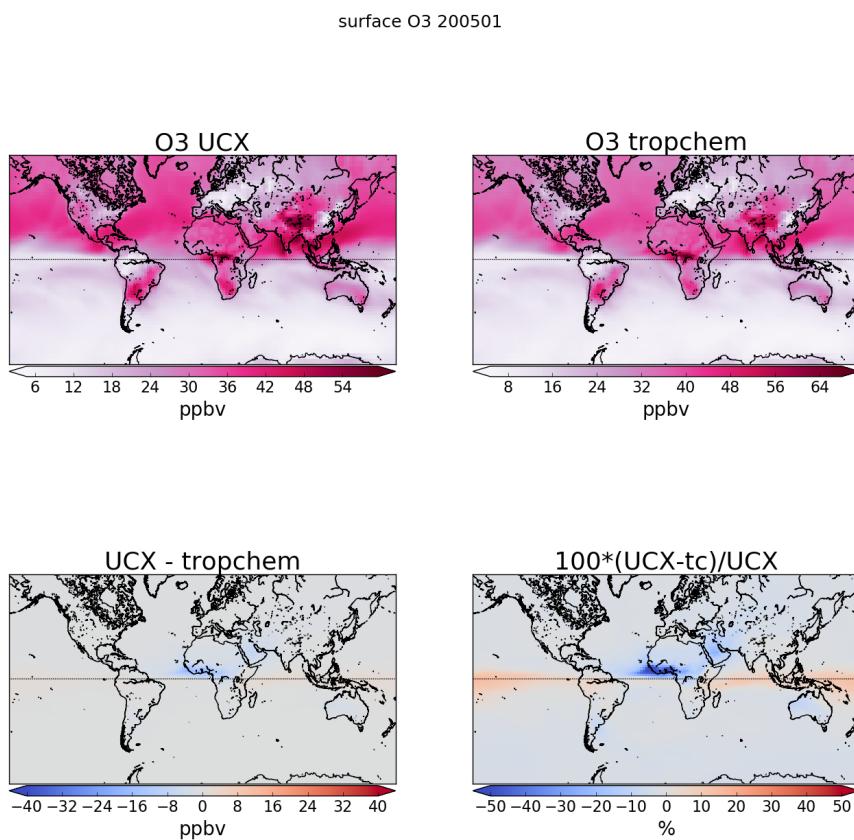


FIGURE 1.8: As figure 1.6, except looking at ozone.

2x2.5 °horizontal resolution). Over much of the country GEOS-Chem overestimates NO₂ by 10-60%, except in NA and northern Queensland where up to 50% underestimation occurs.

1.5 Measurement Techniques

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.

In-situ measurements contain errors, and depending on the device used and chemical being measured this error can be significant. **Dunne2017** analyse the uncertainty of VOC measurements (including isoprene) using three different techniques during a campaign in Sydney in 2012. The major sources of uncertainty in measurement techniques included interference from non-target compounds and under-reporting. Overall isoprene uncertainty in their measurements was a factor of 1.5 to 2. This can feed into uncertainties in modelling and satellite retrievals, as verification and correlations are affected.

1.5.1 DOAS

The DOAS technique uses solar radiation absorption spectra to measure trace gases through paths of light. Beer's law states that $T = I/I_0 = e^{-\tau}$ with T being transmittance, τ being optical depth, and I, I_0 being radiant flux received at instrument and emitted at source respectively. From $\tau_i = \int \rho_i \beta_i ds$ we get:

$$I = I_0 \exp \left(\sum_i \int \rho_i \beta_i ds \right)$$

Where i represents a chemical species index, ρ is a species density(molecules per cm³), β is the scattering and absorption cross section area (cm²), and the integral over ds represents integration over the path from light source to instrument. τ can be described using the attenuation cross section (the attenuation coefficient divided by its number density), with the following relation:

$$\tau = \int_0^l \alpha(z) \eta(z) dz$$

where $\alpha(z)$ and $\eta(z)$ represent absorption cross section in m² molecule⁻¹, and number density in molecules m⁻³ respectively, and l represents the length of the medium (in this case the path through which the light travels).

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. In the atmosphere we are interested in the optical depth of various chemical species, and we use incoming solar radiation to determine this. The

GC NO vs OMNO2d 20050101-20050228

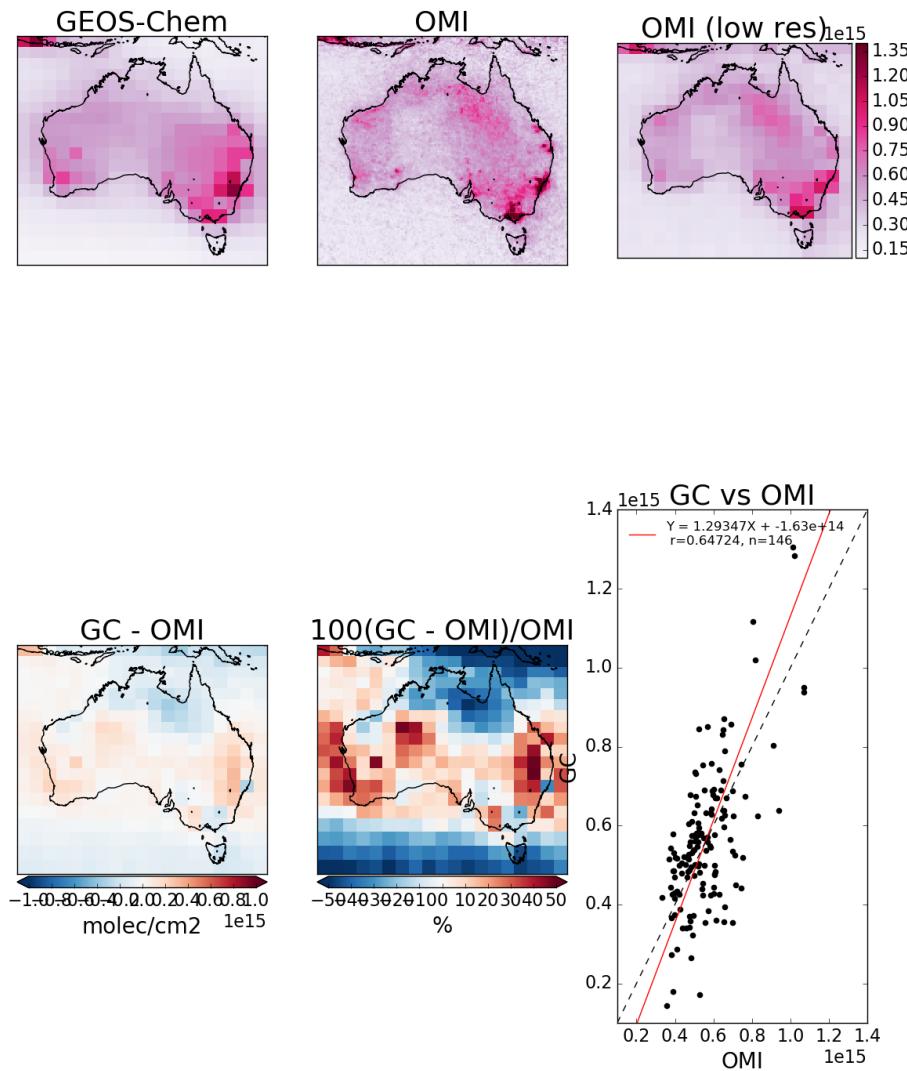


FIGURE 1.9: GEOS-Chem mid-day tropospheric column NO₂ vs OMNO2d columns (averaged to match GEOS-Chem's lower resolution). Absolute and relative differences, along with correlation shown on bottom row.

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 ϕ_e^i is radiant flux seen at the earth surface, ϕ_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.

Multiple axis DOAS (MAX-DOAS) is a remote sensing technique which uses several DOAS measurements over different viewing paths. In these retrievals, the measurements of light absorption are performed over several elevations in order to add some vertical resolution to the measurement of trace gas concentrations. An example of this is shown in figure 1.10, which was taken from Lee2015 Recently MAX-DOAS has been used to examine HCHO profiles in the clean free troposphere (Franco2015; Schreier2016) as well as in polluted city air (Lee2015). Depending on orography and atmospheric composition (ie. the influence of interfering chemicals), MAX-DOAS can be used to split the tropospheric column into two partial columns; giving a small amount of vertical resolution to HCHO measurements (Franco2015; Lee2015). In Franco2015 an FTIR spectrometer at Jungfraujoch is compared against both MAX-DOAS and satellite data, with two CTMs; GEOS-Chem and IMAGES v2 used to compare total columns and vertical resolution of each instrument.

1.5.2 Satellites

In order to detect trace gases such as HCHO, satellites use a DOAS based technique to detect concentrations along the path of light which reaches the satellite instrument. This requires chemical transport and radiative transfer models used to transform the non-vertical light path into vertical column amounts. Measurements done using DOAS often apply a forward radiative transfer model (RTM) such as LIDORT in order to determine a trace gas's radiative properties at various altitudes. The forward RTM used for satellite data products also involves functions representing extinction from Mie and Rayleigh scattering, and the efficiency of these on intensities from the trace gas under inspection, as well as accounting for various atmospheric parameters which may or may not be estimated (e.g. albedo).

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 light. Mie scattering is more dominant from 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 various properties which affect remote sensing devices such as a satellite, making them more or less sensitive at certain altitudes for detecting various species (Martin2002).

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 O₃, NO₂, and HCHO (eg: Leue2001). Several public data servers are available

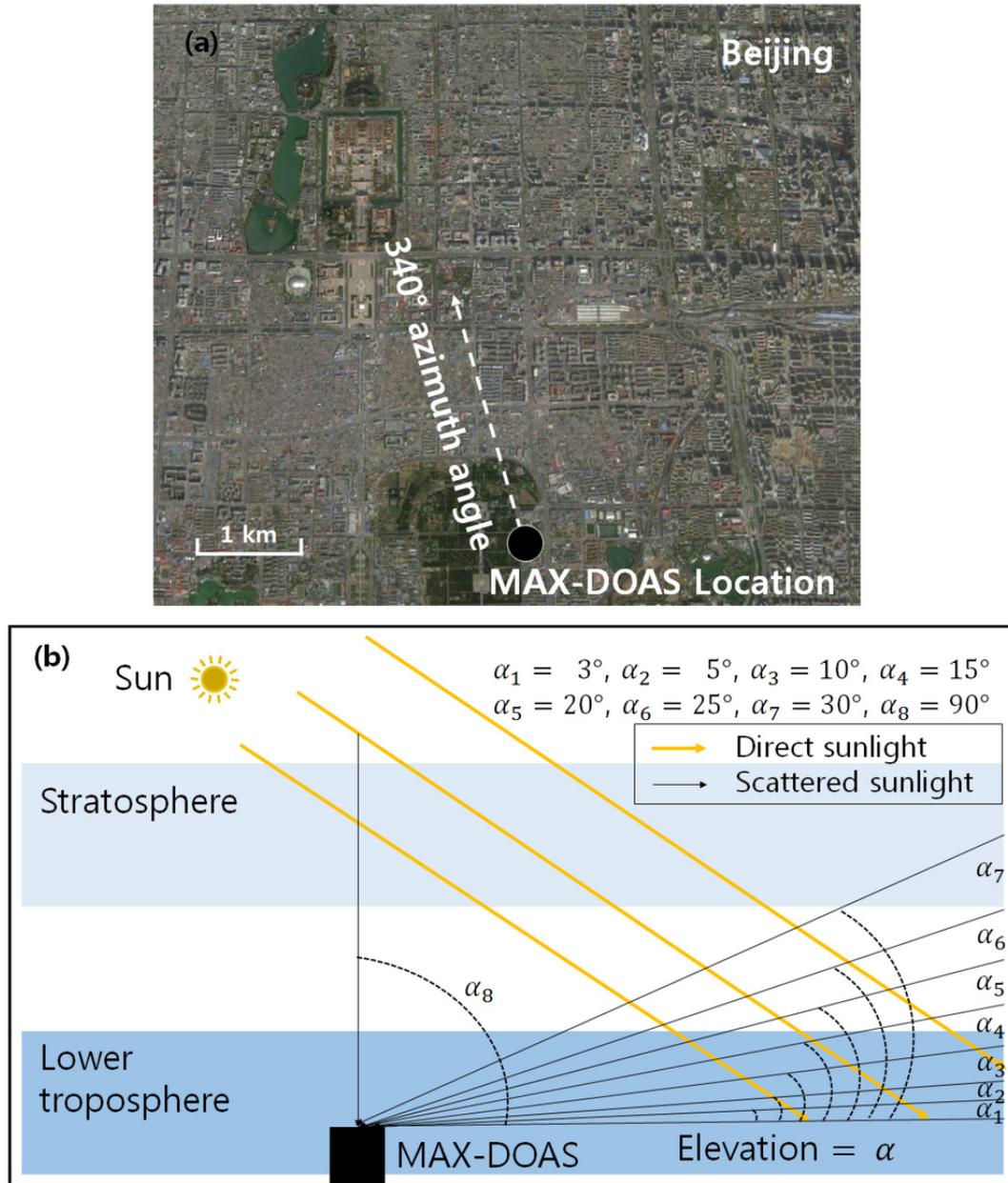


FIGURE 1.10: Image from Lee2015

which include products from satellites, including NASA's Earthdata portal (<https://earthdata.nasa.gov/>) and the Belgian Institute for Space Aeronomy (IASB-BIRA) Aeronomie site (<http://h2co.aeronomie.be/>).

Satellite measurements are generally performed using spectral fitting followed by conversion to vertical column densities. The use of multiple satellites can even be used to detect intradiel concentrations in trace gas columns, as shown in **Stavrakou2015** using OMI and GOME-2 measurements, which have respective overpass times of 1330 and 0930 LT. Instruments including MODIS on board the AQUA and TERRA satellites are also able to determine aerosol optical depth (AOD), a measure of atmospheric scatter and absorbance. An AOD of 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.

Uncertainty in a single pixel for OMI is quite high, roughly the same magnitude as HCHO background levels. Figure 1.11 shows uncertainty over Australia after one and eight days of averaging at 0.25° longitude by 0.3125° latitude. 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. Random error can be reduced by either temporal or spatial averaging, decreasing uncertainty by a factor of $1/\sqrt{N}$ where N is the number of observations being averaged. High resolution low detection limit estimates can be built up using "oversampling", which averages satellite measurements over time. A good example can be seen in **Zhu2014** where 0.2° by 0.2° resolution with high enough sensitivity to see anthropogenic HCHO is achieved with three summers worth of satellite data.

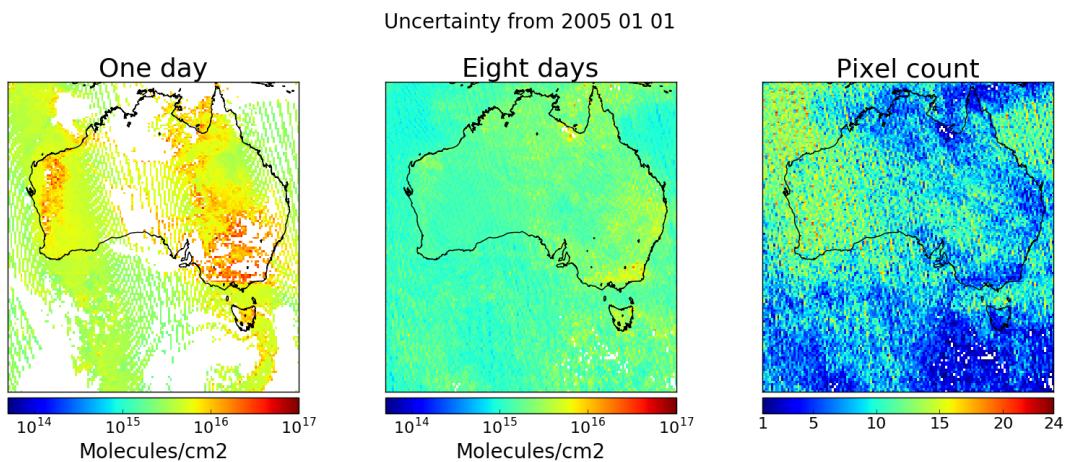


FIGURE 1.11: OMI uncertainty before and after gridding and averaging 8 days from Jan 1 2005 to Jan 8 2005. The third panel shows the number of pixels in each grid box after 8 days of averaging, before accounting for fire.

Soon even more HCHO data will be available in the form of geostationary satellite

measurements (**Kwon2017**). **Kwon2017** examine simulated geostationary measurements against GEOS-Chem column simulations to determine the most important instrument sensitivities. 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.

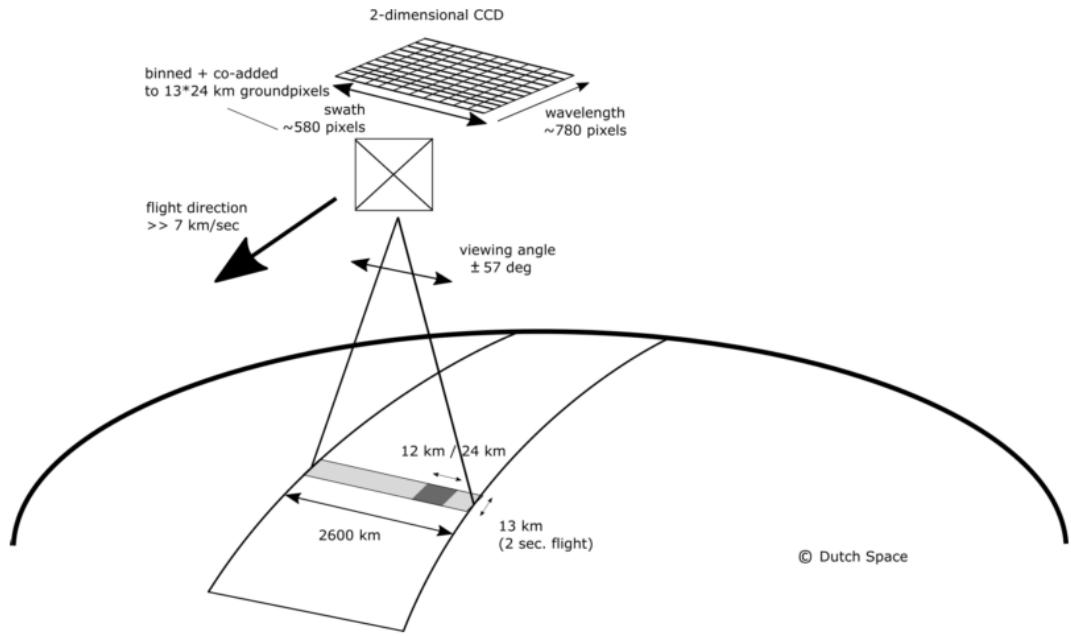
OMI

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 including O₃, NO₂, SO₂, HCHO, and various other quantities like surface UV radiation. Recently **Schenkeveld2017** 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. They also provide a very nice summary of the OMI instrument copied here in Fig. 1.12, as it shows the instruments spectral, temporal, and spatial resolutions. 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. An analysis of the row anomaly by **Huang2017** state that OMI ozone columns remain suitable for scientific use, with recommendation for further evaluation. And analysis of OMI output by **Schenkeveld2017** concludes that data is still of high quality and will deliver useful information for 5-10 more years, with radiances only changing by 1 – 2% outside of RA impacted areas. 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 (**Schenkeveld2017**).

AMF

An AMF characterises measurement sensitivity to a trace gas at various altitudes **Palmer2001** **Lorente2017** 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 (**Lorente2017**). Calculation of the AMF is important as it is multiplied against the estimated slant columns in order to give vertical column amounts. 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.

DOAS retrieval columns 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 all altitudes using an RTM. These properties are combined into a single array called the AMF. Two examples of this are GOME-2 products on the MetOp-A satellite



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

FIGURE 1.12: Figure 1 and Table 1 from Schenkeveld2017 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 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° leads to a swath width on the ground of 2600 km.” The table shows the optical properties for OMIs three channels.

(http://atmos.caf.dlr.de/gome/product_hcho.html) and OMI products which use IMAGESv2 combined with LIDORT and GEOS-Chem with LIDORT for product processing respectively (**Chance2002; Abad2015**). AMFs are unique to each trace gas and due to their complexity and the influence of cloud cover they remain one of the larger error sources in remote sensing of BVOCs (**Palmer2001; Millet2006**)).

Related to the AMF is the averaging kernal (AK), which is used to handle instrument measurements which are sensitive to concentrations at different altitudes in the atmosphere. DOAS methods can be heavily influenced by the initial estimates of a trace gas profile (the *apriori*) 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 *apriori* assumptions. One way to remove these *apriori* influences is through the satellites AK (or by using AMFs), which takes into account the vertical profile of the modelled trace gas and instrument sensitivity to the trace gas (**Eskes2003; Palmer2001**). **Lamsal2014** recommends that when comparing satellite data to models, the AMF should first be recalculated using the model as an *apriori*. This is in order to remove any *apriori* bias between model and satellite columns. Another way of removing this bias is through deconvolution ($\Omega = AK \times VC_{satellite} + \times(I - AK)VC_{apriori}$) of the averaging kernal (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:

$$AK(z) = \frac{\omega(z)}{AMF} \quad (1.2)$$

Note that this is an approximation for the OMI product, which does not include the AK but does include the ω and AMF, as explained in **Abad2015**

The latest OMI 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 (**Abad2015**). The GEOS-Chem model has been substantially updated since then, and using the more recent version V10.01 to recalculate the AMF is performed within this thesis, details are shown in section [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 to 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 (**Millet2006**). Many products require analysis of cloud and aerosol properties in order to estimate concentration or total column amounts (**Palmer2001; Palmer2003; Marais2012; Vasilkov2017**). The main source of error in satellite retrievals of HCHO are due to instrument detection sensitivities, and the vertical multiplication factor (**Millet2006**). Calculations of the AMF performed by

different groups tend to agree fairly well, as long as all the apriori and ancillary data is similar. Large differences can occur depending on the apriori vertical profile, trace gas concentrations, and cloud properties (**Lorent2017**). 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 **Lorent2017**

There are two types of measurement error, arguably the worst of these is systematic error (or bias) which 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 the other type (often reported as some function of a datasets variance, or uncertainty), and this can be reduced through averaging either spatially or temporally. By taking the average of several measurements, any random error can be reduced by a factor of one over the square root of the number of measurements. This is done frequently for satellite measurements of trace gases (which are often near to the detection limit over much of the globe). For example: **Vigouroux2009** 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.

Satellite measurements of HCHO are relatively uncertain, however this can be improved by averaging over larger grid boxes or longer time scales. An example of this can be seen in **Dufour2009** where monthly averaging is used to decrease the measurements uncertainty. They examine HCHO in Europe, which is low; near the detection limit of satellite measurements. Taking monthly averages allows enough certainty that useful inversions can be determined to estimate the source emissions of HCHO. The finer nadir resolution of OMI (13 by 24 km²) compared to other satellites reduces cloud influence (**Millet2006; Millet2008**). Although the uncertainty in each pixel is $\sim 2 \times 10^{16}$, which is 5 \times higher than GOME, there are $\sim 100 - 200 \times$ as many measurements due to the smaller footprint and better temporal resolution of OMI, which allows a greater reduction of uncertainty with averaging (**Chance2002; Millet2008**).

In cloudy, hazy or polluted areas measurements are more difficult to analyse (**Palmer2003; Marais2014**). Recent work by **Vasilkov2017** showed that updating how the surface reflectivity is incorporated into satellite measurements can change the retrievals by 50 % in polluted areas. With the satellite HCHO columns, we filter cloud fractions over 40%, which introduces a clear-sky bias. This bias has been measured as a 13% positive monthly mean bias by **Palmer2001; Surl2018**

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 artifact using modelled HCHO over the remote pacific (**Shim2005**). OMI HCHO products use a similar technique to account for sensor plate drift and changing bromine sensitivity (**Abad2015**).

For many places the tropospheric column HCHO measured by satellite is biased low, **Zhu2016** examine six available datasets and show a bias of 20 - 51% over south east USA when compared against a campaign of aircraft observations (SEAC⁴RS).

DeSmedt2015 also found OMI and GOME2 observations were 20 - 40% lower than ground based vertical profiles, and **Barkley2013** determine OMI to be 37% low compared with aircraft measurements over Guyana. These bias can be corrected by improving the assumed apriori HCHO profiles which are used to calculate the AMFs of the satellite columns. **Millet2006** examine OMI HCHO columns over North America and determine overall uncertainty to be 40%, with most of this coming from cloud interference. **Millet2008** 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.

Uncertainty in the OMI satellite instrument is calculated by the Smithsonian Astrophysical Observatory (SAO) group using the uncertainty in backscattered radiation retrievals (**Abad2015; Abad2016**). 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 (**DeSmedt2012; DeSmedt2015**).

A full analysis of the AMF uncertainty in OMI measurements, as well as the structural uncertainty (between different systems of calculations applied to the same data) is performed by **Lorente2017**. They determine the structural uncertainty using ensemble techniques on seven AMF calculation approaches used by different retrieval groups. They show that 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. The importance of apriori and ancillary data (such as surface albedo and cloud top height) is also shown, as it sharply affects the structural uncertainty.

GOME suffers from similar uncertainties to OMI, as the same general method of DOAS remote measurements are performed. The uncertainty from slant column fitting has been calculated for GOME to be 4×10^{15} molecules cm^{-2} (**Chance2000; Millet2006**). The conversion factor for slant to vertical columns (AMF) calculation also suffers from errors; primarily from surface albedo, HCHO vertical profile apriori, aerosol, and cloud influence (**Millet2006**). AMF uncertainties for GOME are calculated to be 1 to 1.3×10^{15} molecules cm^{-2} by **Shim2005**

1.6 Recalculation of OMI HCHO

When comparing satellite observations to a chemical model, one needs to recalculate the satellite AMF using their own modelled vertical gas profiles as the a-prior shape factor in order to remove any total column bias which may be due to the satellite's apriori. The method used here largely follows that of **Palmer2001**. When comparing satellite observations to a chemical model, recalculation of the satellite AMF using modelled vertical gas profiles removes any bias introduced by differences from the a-priori shape factor to the model.

The AMF is needed to transform the slant column, as viewed by the satellite, into a vertical column:

$$AMF = \frac{\Omega_s}{\Omega_v} \quad (1.3)$$

where s and v subscripts refer to slant and vertical values, while Ω represents a column of absorber in molecules cm^{-2} .

In order to visualise and analyse satellite column data it is generally transformed into vertical columns. This is done using AMF calculations as shown in the following subsections. Taking the biogenic slant columns, scattering weights, and apriori estimates of HCHO vertical profiles we determine vertical HCHO column amounts. This is an in depth process involving radiative transfer modelling in order to work out satellite sensitivities at various altitudes, as well as the effect from the local HCHO profile on those sensitivities. Several of these required data are available from the satellite data products, including the scattering weights and the zenith angles required to determine an AMF at any particular measured point. In this work the shape factor is recalculated from GOES-Chem, with the associated OMI per-pixel scattering weights unchanged. The satellite shape factor is replaced by GEOS-Chem's overpass time simulated HCHO profile, normalised and saved daily along with air density.

When comparing satellite measurements against models it is important to recognise the impact of the apriori shape factor on the total column values. This is due to the sensitivity of instruments varying vertically through the atmosphere, and how the simulated distribution of HCHO is accounted for. In order to remove a possible bias caused by systematic differences between the old model and the current model, the shape factor used by the satellite is replaced using the profile from the current model before satellite total columns are recalculated (for example using equation 1.14). Both the shape factor and scattering weights of the satellite are recalculated using a combination of GEOS-Chem apriori profile information and satellite measurement data using code initially written by dr. Paul Palmer, which calculates the AMF after running the LIDORT radiative transfer calculations to determine apriori scattering, see sections 1.6.2 and 1.6.3 for more details. Without performing this step a bias between modeled and measured total column values may be due to an apriori rather than actual chemistry or measurements.

Apriori GEOS-Chem HCHO profiles are created using output averaged between 1200 and 1300 local time (LT). Using this one hour average matches the overpass time of OMI, and is similarly performed in Jin2017

Figure 1.14 shows an overview of how this process is performed in my work. Output from GEOS-Chem is combined with both OMHCHO swath data and MOD14A1 gridded fire data in order to produce a gridded HCHO file which contains HCHO vertical columns and fire counts. The output keeps the original AMF as well as those recalculated using GEOS-Chem, and optionally those recalculated using GEOS-Chem and LIDORT.

After recalculating the AMFs for each satellite pixel using GEOS-Chem v10.01, we are left with the new vertical columns which are essentially what the satellite would show given that the modelled column was known and correct. Comparing these new vertical columns to model output removes any bias caused by the apriori vertical column in the satellite product, since now biases in the modelled column affect both data sets equally.

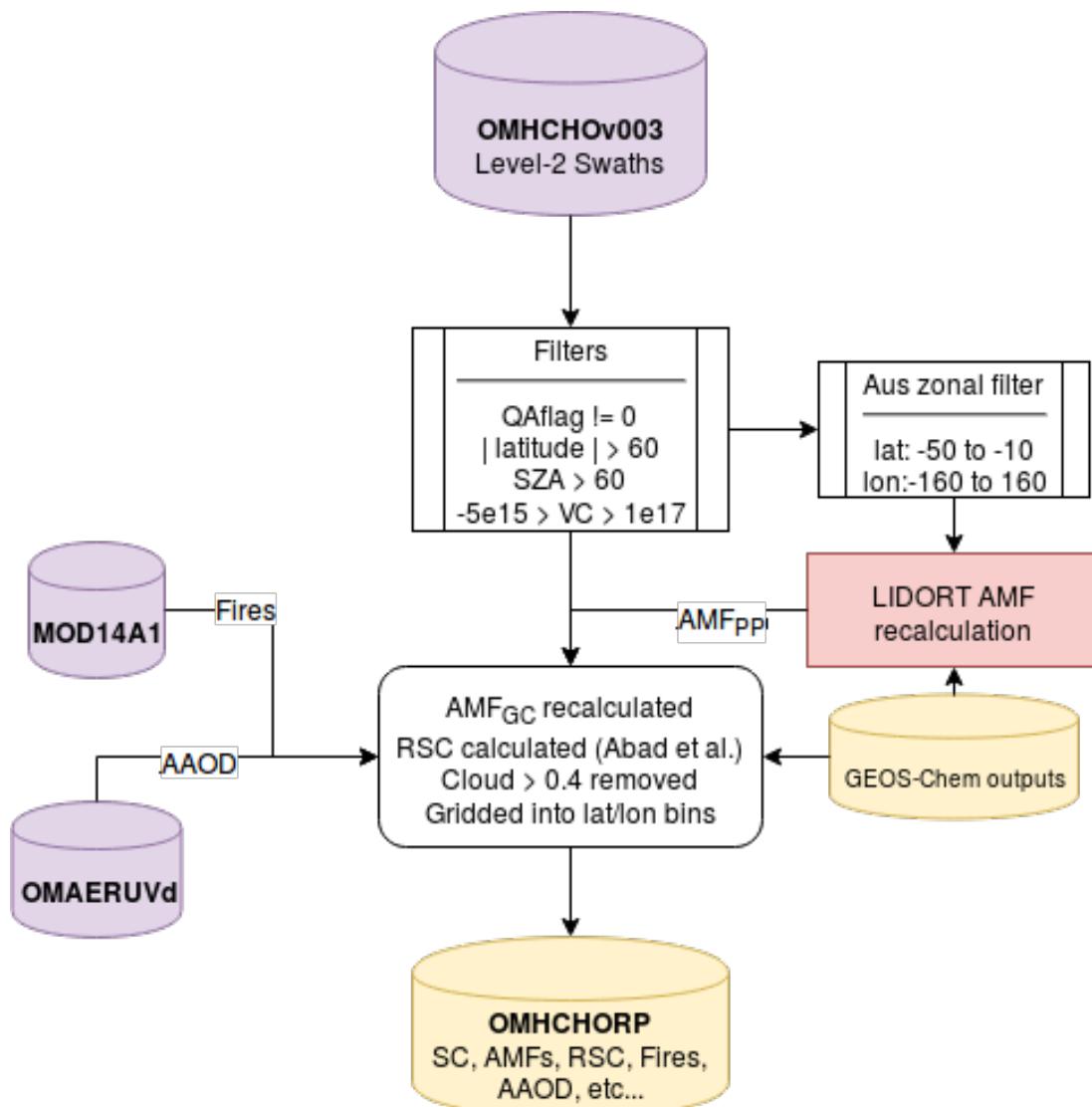


FIGURE 1.13: Flow diagram showing how OMHCHO level two swath data is read, processed, and gridded in this thesis

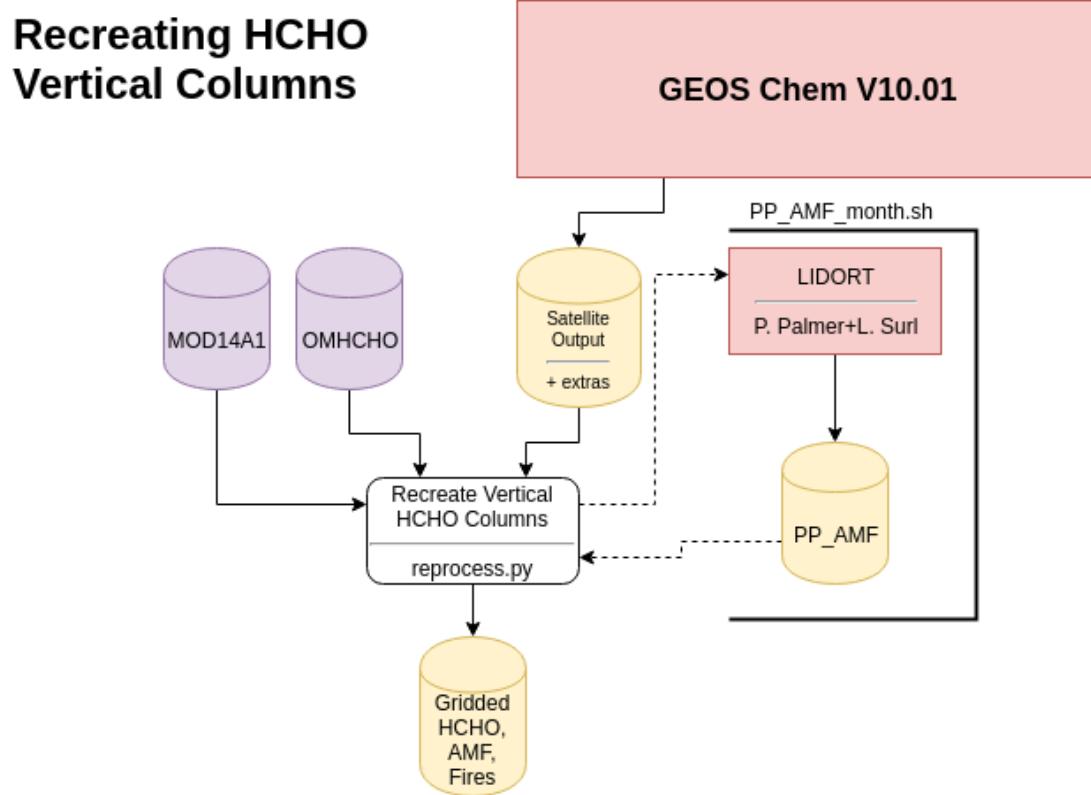


FIGURE 1.14: Depiction of processes and datasets used to recalculate OMI AMFs.

1.6.1 Reading satellite data

First satellite slant columns of formaldehyde for the years January 1st, 2005 - April 1st, 2013 are downloaded (see section 1.2.4). 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.

The satellite data comes with quality assurance flags which are used as filtering criteria as is done in several other studies (**Marais2012; Barkley2013; Bauwens2016; Zhu2016**). This filtering removes highly uncertain pixels, along with those for which instrument problems such as the OMI row anomaly (see section 1.2.4) may have affected. Clouds also have various detrimental effects on slant column uncertainty and AMF calculation, so cloud fraction is saved in the OMHCHO product.

Each pixel and its relevant data are saved in a long list, around 1.1 million pixels per day. Additional information is added to each pixel, such as the new AMF calculated through replacing the apriori vertical profile with the newer GEOS-Chem simulated profile. The shape factors and scattering weights for each pixel lie along a z-axis which is vertically resolved to 47 layers.

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 1.13 for a single day. First all the pixels for are read, filtered by the QA flag, solar zenith angle (SZA) and latitude (**Palmer2001**). Satellite measurements polewards of 60 °north or south are removed as well as measurements with SZA greater than 60 °. Further filtering is performed to remove the measurements which are most likely to be unrealistic: those with column density outside the range -0.5×10^{16} to 10^{17} molecules cm $^{-2}$. These are similar filters to those applied in **Zhu2016** (TODO: add similar justification if succinct). This final filter is required due to currently unexplained large negative values which occur in the OMI HCHO product increasingly over time. Figure 1.15 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 cm $^{-2}$ region.

Finally the pixels which may be too cloudy are dealt with after the pixel has been used in determining the reference sector correction (see section 1.6.4), as is done in **Abad2015; DeSmedt2015** Any pixel with a cloud fraction of greater than 40% is removed,

1.6.2 Calculating the apriori shape factor

OMI's apriori shape factor is based on the GEOS-Chem (v9) model, which uses 47 layers between the earth's surface and the top of the atmosphere using a pressure-eta hybrid (the actual values are shown in table ??). The equations and method of recalculating this apriori shape factor with the newer GEOS-Chem V10.01 code is layed out in this section. For each pixel the shape factor used in the AMF is recalculated using GEOS-Chem by following the method layed out initially in **Palmer2001**

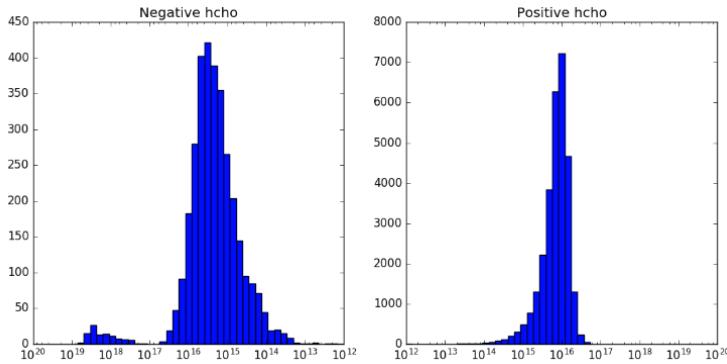


FIGURE 1.15: Column density histograms for a subset of OMI swaths over Australia on the 18th of March 2013. Negative entries are shown in the left panel, positive in the right, note the different scale between negative and positive panels.

The Beer-Lambert law of extinction allows spectroscopic measurement of absorbing chemical species (absorbers) in the atmosphere:

$$I_B = I_{B_0} e^{-\tau_s} \quad (1.4)$$

where I_B, I_{B_0} is backscattered intensity with and without the absorber respectively, and τ_s is the optical thickness of the slant column absorber (see section 1.5.1).

The AMF is needed to transform the slant column, as viewed by the satellite, into a vertical column:

$$AMF = \frac{\Omega_s}{\Omega_v} = \frac{\tau_s}{\tau_v} \quad (1.5)$$

where s and v subscripts refer to slant and vertical values, while Ω represents a column of absorber in molecules cm^{-2} .

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 (θ_s) and the satellite viewing angle (θ_v) are shown in image 1.16. However, in the UV-VIS region of the spectrum, Rayleigh and Mie scattering (see section ??) must be accounted for. Using equations 1.4 and 1.5 gives us an expression for the AMF which includes scattering:

$$AMF = \frac{\ln I_{B_0} - \ln I_B}{\tau_v} \quad (1.6)$$

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

$$AMF = \frac{\ln \left(1 - \frac{\nabla I}{I_B} \right)}{\tau_v} \approx \frac{-\frac{\nabla I}{I_B}}{\tau_v} \quad (1.7)$$

∇I can also be expressed as the integral of the absorption slices over optical depth increments:

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

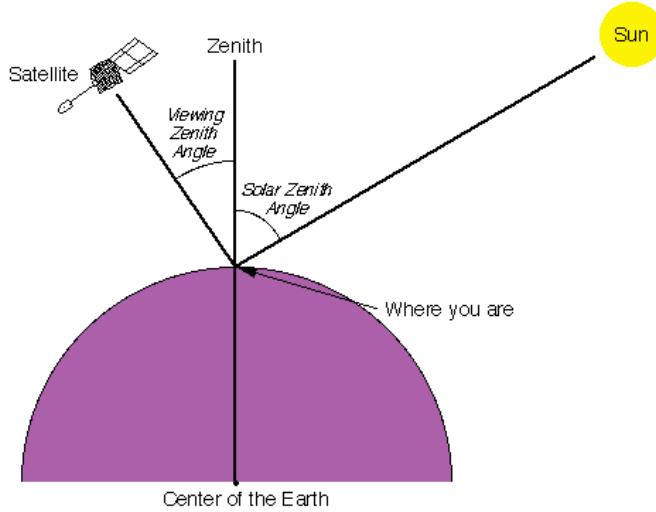


FIGURE 1.16: Solar and viewing zenith angles, image copied from [SZA_Image](#) originally from a NASA website.

which can be placed into equation 1.7:

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

and rewritten as:

$$AMF = \frac{-1}{\tau_v} \int_0^{\infty} \frac{\partial \ln I_B}{\partial \tau} \alpha(z) \eta(z) dz \quad (1.8)$$

where $\alpha(z)$ and $\eta(z)$ represent absorption cross section in m^2 molecule $^{-1}$, and number density in molecules m^{-3} respectively. This uses the attenuation cross section relationship to optical depth (see section 1.5.1).

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

$$\alpha_e = \frac{1}{\Omega_v} \int_0^{\infty} \alpha(z) \eta(z) dz$$

This is the same as $\tau_v = \Omega_v \alpha_e$, which we can place into equation 1.8 to obtain:

$$AMF = - \int_0^{\infty} \frac{\partial \ln I_B}{\partial \tau} \frac{\alpha(z)}{\alpha_e} \frac{\eta(z)}{\Omega_v} dz \quad (1.9)$$

Defining $\omega(z)$ as the scattering weights describing the sensitivity of the backscattered spectrum to the abundance of an absorber at altitude z :

$$\omega(z) = - \frac{1}{AMF_G} \frac{\alpha(z)}{\alpha_e} \frac{\partial \ln I_B}{\partial \tau} \quad (1.10)$$

It's 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 ω . This is not noted in any of the papers which recalculate the AMF from the OMI product, due to them recalculating the ω term themselves with a radiative transfer model such as LIDORT.

Plugging equation 1.10 into equation 1.9 gives us:

$$AMF = - \int_0^\infty \omega(z) \frac{\eta(z)}{\Omega_v} dz \quad (1.11)$$

which is best described as the scattering weights (ω) multiplied by the shape factor (η , normalised by its total column Ω_v) integrated over altitude z . Since we are using the ω provided by OMI, the AMF_G term is removed from this calculation when compared against **Palmer2001**

The integration is done in Python using a simple rectangular method, which multiplies the integrand midpoints by the change in height, and then takes the sum. This is identical to calculating the integral if we assume the integrand is linear between each measured point, which is true as the provided omega is linear between the 47 resolved values, and introduces no new uncertainty.

All that remains for recalculating the total vertical column using our new apriori shape factor is to apply the new AMF to the slant columns and grid them onto our chosen resolution.

Recalculation of the ω is done separately, and explained in section 1.6.3.

$S_z(z)$ can be calculated using an apriori vertical profile, which may be sourced from any atmospheric chemistry model. I recalculate $S_z(z)$ using satellite output from GEOS-Chem, keeping the provided OMI $\omega(z)$. As a sanity check I also recalculate S_σ and confirm that these shape factors are equivalent, for details see section 1.6.2.

For example see figure TODO: sigma vs z shape factor plot.

Normalised vertical shape factor

Additionally the vertical shape factor is normalised and the recalculated AMF is checked against that using the non-normalised version. The additional steps to calculate a normalised vertical shape factor S_σ are first laid out in **Palmer2001**

The following equation converts model profile output from ppb into number densities:

$$\eta_{HCHO} = ppb_{HCHO} \times \eta_a \times 10^{-9} \quad (1.12)$$

where η_{HCHO} is the number density of a HCHO, η_a is the number density of air (from model output), and ppb_{HCHO} is the molecules of that species per billion molecules of air. In order to normalize these vertical density profiles over the globe, we divide by the modelled total vertical column Ω_{HCHO} which is determined by:

$$\Omega_{HCHO} = 2.12 \times 10^{13} \sum_z (ppb_{HCHO}(z)(P(z) - P(z + 1)))$$

where $P(z)$ is the pressure (hPa) at the bottom of altitude level z , the constant 2.12e13 is determined from equation (TODO: Add this equation). In effect this equation sums over the molecules per cm^2 in each altitude level.

The normalised vertical shape factor $S_z(z)$ is defined as:

$$S_z(z) = \frac{\eta(z)}{\Omega_v} \quad (1.13)$$

where $\eta(z)$ is the number density in molecules m^{-3} .

Now the AMF can be expressed as

$$AMF = \int_0^\infty \omega(z) S_z(z) dz \quad (1.14)$$

We have $S_z(z)$ and $\omega(z)$ over the vertical pressure coordinate z at all latitude and longitude points on whatever grid we wish. A conversion to the sigma (σ) 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. In the sigma coordinate system we calculated the shape factor as follows:

$$S_\sigma(\sigma) = \frac{\Omega_a}{\Omega_v} C_{HCHO}(\sigma) \quad (1.15)$$

where Ω_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. This equation comes from **Palmer2001** and is unitless since Ω_a/Ω_v is molecules of air per molecule of HCHO; the opposite of C_{HCHO} .

Sigma coordinate conversion

Using the vertical coordinate sigma (σ), which is related to pressure (P) by $P = \sigma(P_S - P_T) + P_T$, where subscripts S and T represent earth surface and top of the atmosphere respectively. The hydrostatic relation $P = -\rho_a g z$, with ρ_a , g , being density of air, gravity, respectively lets us switch to the sigma coordinate using:

$$\begin{aligned} \rho_a g z &= \sigma (P_S - P_T) + P_T \\ d\sigma &= -\frac{\rho_a g}{P_S - P_T} dz \end{aligned}$$

Substitution into 1.14 gives AMF using the sigma coordinates:

$$AMF = \int_0^1 w(\sigma) S_\sigma(\sigma) d\sigma \quad (1.16)$$

Where S_σ is defined as a vertical shape factor representing a normalised mixing ratio:

$$S_\sigma(\sigma) = \frac{\Omega_a}{\Omega_v} C(\sigma) \quad (1.17)$$

where Ω_a is the vertical column of air and $C(\sigma)$ is the mixing ratio of the absorber. This normalised shape factor is dimensionless. 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. When

1.6.3 Recalculating the AMF using PP code

Next a subset covering Australia and most of the zonal band has its AMF recalculated with by fortran code written by Paul Palmer, Randal Martin, and updated by Luke Surl. I will refer to this as the PP code, and subscripting the vertical column and AMF with PP when it is calculated through this method. This code is computationally expensive, and is only run on pixels within a specific range including Australia ($50\text{--}10^{\circ}$ S, 160° W– 160° E). These are used to recalculate the instrument sensitivity or scattering weights for each pixel, as well as the shape factor which together are integrated to give the pixel AMF.

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. The original method for HCHO is layed out in **Palmer2001** with modifications for clouds and use of the LIDORT RTM (**Spurr2002**) as described by **Martin2003**. 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 output by GEOS-Chem is required, which is achieved by modifying the ND51 diagnostic.

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 TODO: add git repo with this code.

First special output is required from GEOS-Chem, averaged between 1300 and 1400 LT, including optical depths at several wavelengths (TODO: list), dust, and HCHO. I then pull out a subset of the OMI pixel information into a daily csv file, which can be read by the PP code as modified by Dr. Luke Surl, in conjunction with the GEOS-Chem outputs for each day. The PP code then produces a csv of recalculated AMFs which get read by my python code and associated with the corresponding pixel (outlined in [1.14](#)).

1.6.4 Reference sector correction

Each satellite slant column measurement is corrected by some amount, based on the divergence from a modeled reference sector. HCHO products from OMI and SCIA-MACHY 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 (**DeSmedt2008**; **Barkley2013**; **Kurosu2014**). This reference sector correction (RSC) is calculated following **Abad2016** and applied to the satellite AMF, the AMF recalculated by GEOS-Chem, and the AMF calculated using Paul Palmers code.

The reference sector correction method corrects for several problems, however it introduces some apriori 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 (**DeSmedt2015**). This method also corrects for the errors introduced through correlations between BrO and HCHO absorption cross sections, which are especially significant at high latitudes (**Abad2015**).

Vertical columns in OMI use this oceanic background instead of a solar irradiance spectrum In order to recalculate the vertical columns using modelled data, a RSC needs to be applied. The corrected vertical column (Ω_{VC}) is calculated as the slant column (Ω_S) minus the reference slant column (Ω_{S_0}) multiplied by the AMF, plus the modelled reference sector column (Ω_{V_B}):

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

This method is used in various papers, including **DeSmedt2008; DeSmedt2012; DeSmedt2015; Barkley2013; Bauwens2016** Recently this correction was expanded (for OMI data) to include latitudinal and instrument track influence by **Abad2015**

A correction for each instrument pixel is created based on the difference between the background HCHO measurements from OMI and the GEOS-Chem modelled HCHO columns within the reference sector. This correction is calculated daily and applied to all good pixels based on their latitude.

The longitudinal average is taken within the apriori reference sector, as corrections are assumed to be longitudinally invariant. The modeled reference sector is interpolated latitudinally in for use in the OMI measurement correction array creation. Figure 1.17 the simulated reference sector VCs as an example, calculated on January 1st 2005. In this figure the vertical resolution is increased from 2° to 0.36° , through linear interpolation, in order to form 500 vertical bins which are used in correcting the satellite data. Each day, good satellite measurements taken over the reference sector are used to determine a correction array. The correction is based on the difference between measured slant column and the modeled slant column within the reference sector. The model does not produce slant columns associated with each measurement, however one is created by multiplying the VC with the associated slant column's AMF.

For OMI swaths, each row of measured data contains 60 'Across track'(track) measurements. The track index (i) relates a the measurement to one of the 60 columns of data. Corrections for each measurement are calculated by taking the difference between the measured slant column and the apriori slant column as follows:

$$Correction(i, j) = SC_{HCHO}(i, j) - VC_{GEOS-Chem}(lat(j)) \times AMF_{OMI}(i, j) \quad (1.18)$$

where j represents a latitude index and $VC_{GEOS-Chem}(lat)$ represents the apriori reference sector vertical column HCHO at the latitude corresponding to j. Note that the correction is in molecules cm^{-2} . The reference sector correction is independently calculated for each of the 60 tracks, at each latitude where a good satellite measurement exists which used that track. The $Correction(i, lat(j))$ function is determined by binning corrections for each track into 500 equidistant latitude bands.

Due 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 1.18 shows an example of the 60 track corrections for January 1st 2005, the points are satellite measurements and the lines are the interpolations for each track.

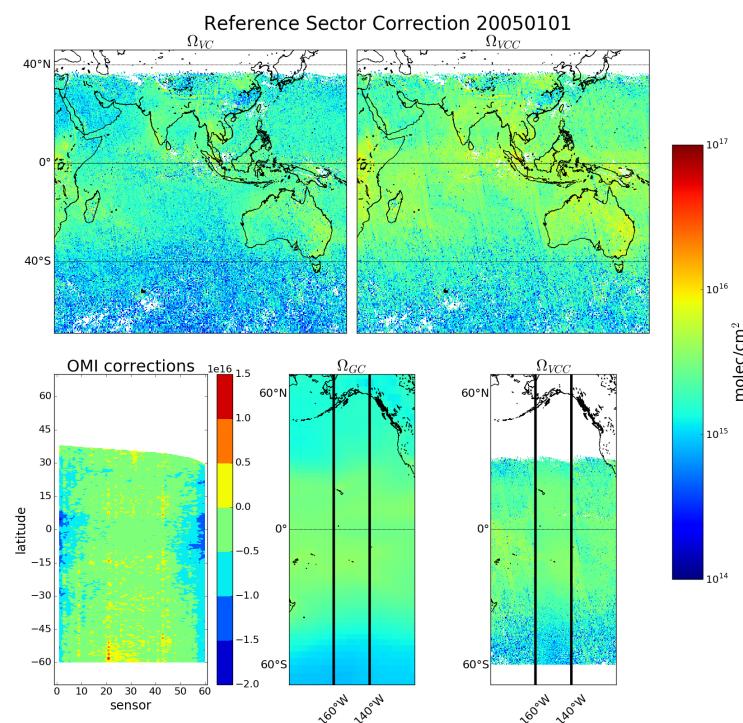


FIGURE 1.17: Example of remote pacific reference sector correction (RSC) using 8-day average measurements and one month modelled data. Ω_{VC} shows the uncorrected vertical columns, while Ω_{VCC} shows the corrected vertical columns. OMI corrections shows the correction applied globally based on latitude and OMI track number(sensor). Ω_{GC} shows the GEOS-Chem modelled HCHO VC over the RSC, with Ω_{VCC} showing the corrected VC over the same area.

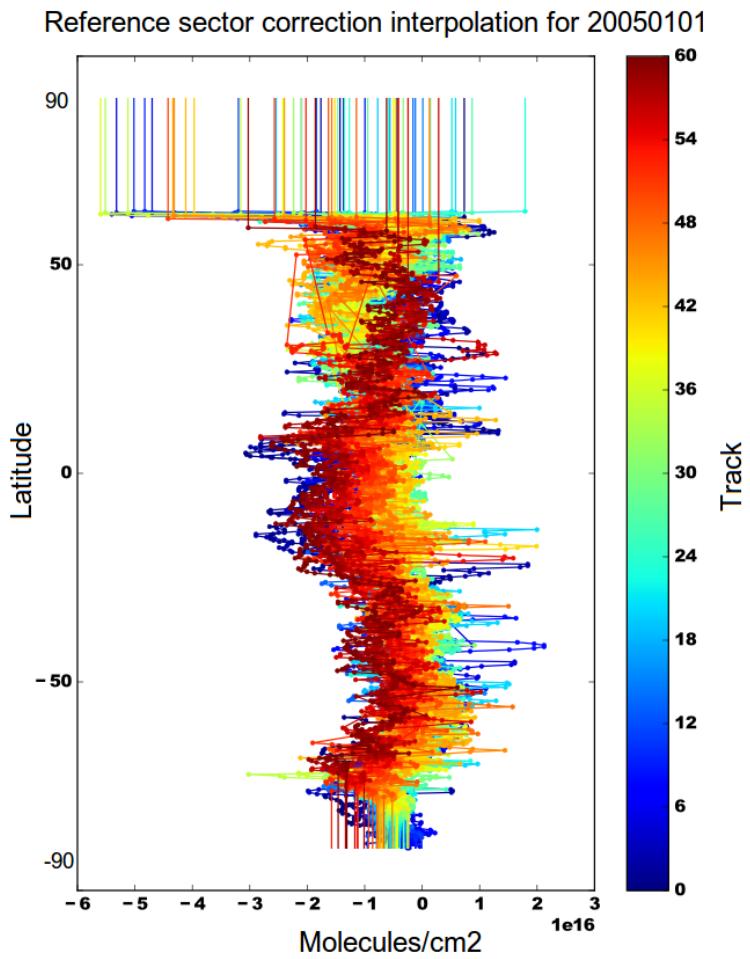


FIGURE 1.18: Example of track correction interpolations for January 1st 2005, points represent satellite slant column measurements, with lines interpolating and extrapolating along the latitudinal direction.

Another way to look at this correction is given in the OMI corrections panel of figure 1.17, 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:

$$VCC(i,j) = \frac{SC_{HCHO}(i,j) - Correction(i, lat(j))}{AMF(i,j)} \quad (1.19)$$

Finally, for each day, the good satellite measurements are averaged into our own latitude longitude resolution bins along with the associated corrected SC, VC, VCC, AMF, and bin entry count. The bin entry count is used to create an 8-day average out of the one day averages, as it is the daily mean multiplied by the daily count summed over 8 days divided by the total count for each bin.

1.6.5 Binning the results daily

Finally the pixels are binned into a gridded dataset I've called OMHCORP, as shown in figure 1.13. The resolution is chosen to match the native resolution of GEOS-Chem ($0.25 \times 0.3125^\circ$) and the GEOS met data fields. 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)

TODO: time per regridding and reprocessing: 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, X GB of daily data was downloaded and then transformed into Y GB 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 running sending separate ‘jobs’ to NCI’s process queue was simpler and more scalable. As much as possible, processing is done using the HDF-5 format, with some GEOS-Chem output being read and processed from bitpunch to HDF-EOS5 prior to reprocessing. The scripts to regrid and reprocess the swath data set are available in the supplementary (TODO).

1.6.6 Difference between new and old OMI HCHO columns

Two HCHO products are created, both using GEOS-Chem output at global 2 by 2.5 °horizontal resolution. One uses the OMI product’s ω_z and equation ?? in order to calculate an AMF. While the other uses code provided by Dr. Paul Palmer, with alterations by Dr. Randal Martin, and Dr. Luke Surl to run LIDORT on the satellite slant columns and the GEOS-Chem output in order to calculate an AMF. These two calculations are compared over Australia in figure(s) TODO: Map comparison, regression, and time series once AMFpp is working properly. The effect of not recalculating the ω_z is can be seen in figure 1.19 which looks at the altered satellite vertical columns using each method.

Figure 1.19 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 °lat lon resolution) has been created by binning the recalculated satellite pixels within the month. The average pixels per land square is overlaid and changes due to how a 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 ω_z .

TODO: Ask Luke if this is true: The AMF calculated using Dr. Palmer’s code uses a more strict series of filters, leading to fewer satellite based HCHO columns and reduced coverage over Australia. Stricter filtering must be balanced against both coverage and the sensitivity of the AMF determination to recalculating ω_z .

1.7 Filtering Data

In order to examine only biogenic processes, pyrogenic and anthropogenic influences need to be removed from modelled and measured data. As biomass burning can be a large local or transported source of HCHO, CHOCHO, glyoxal, and other compounds we would like to use to determine BVOC emissions, it is advantageous to filter out this source. Just filtering active fires does not account for transported smoke plumes, which can carry HCHO precursors. One complication when computing HCHO yield from VOC emissions is biomass burning interference, as smoke plumes can contribute

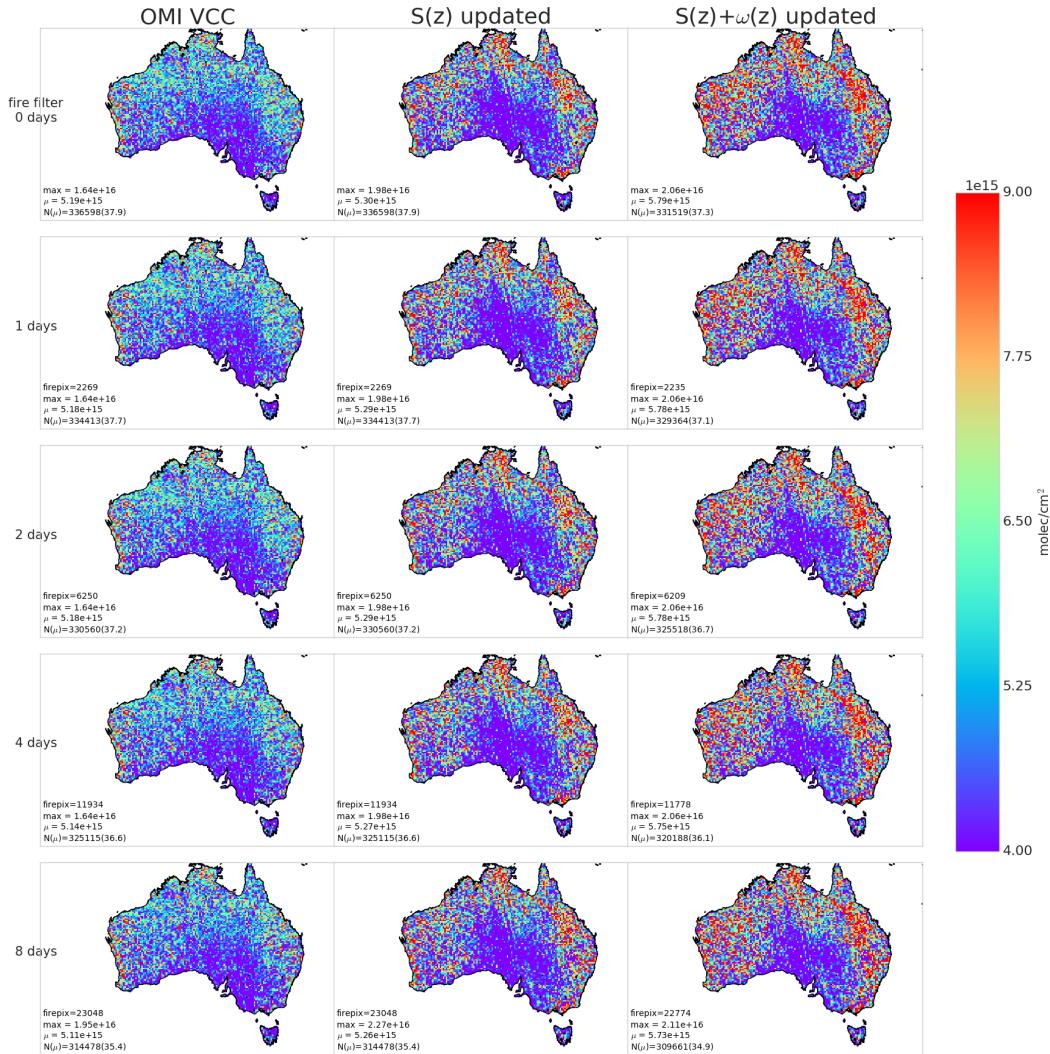


FIGURE 1.19: Column 1: Reference sector corrected HCHO vertical columns Ω from OMHCVOv003. Column 2: Ω with recalculated apriori shape factors using GEOS-Chem v10.01. Column 3: Ω with recalculated apriori shape factors and scattering weights using GEOS-Chem v10.01 and LIDORT. Row 1-5: increasing number of prior days which have active fires are included when masking fire influence.

to column HCHO. In GEOS-Chem we can simply turn off pyrogenic and anthropogenic emissions, however in satellite datasets we need to mask pixels affected by biomass burning.

Influence from biomass burning can be removed through measurements of acetonitrile and CO (eg: ([Wolfe2016](#); [Miller2017](#)), or else removal of scenes coincident with satellite detected fire counts and aerosol absorption optical depth as done in [Marais2014](#) [Wolfe2016](#) disregard HCHO measurements when acetonitrile > 210 pptv and CO > 300 ppbv, while acetonitrile > 200 pptv is used to determine fire influence in [Miller2017](#) 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. [Zhu2013_poster](#) has a similar analysis over south-eastern USA showing an exponential correlation of $HCHO = \exp(0.15 * T - 9.07)$.

1.7.1 Fire and smoke

The method used in this thesis follows that of [Marais2012](#) and [Barkley2013](#) with active fires filtered using fire counts, and smoke filtered out using smoke aerosol absorption optical depth (AAOD). [Marais2012](#) remove pixels colocated with non zero fire counts in any of the prior eight days, within grid squares with $1 \times 1^\circ$ resolution. [Barkley2013](#) use fires from the preceding and concurrent day, within local or adjacent grid squares, with grid resolution of $0.25 \times 0.3125^\circ$. 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 can be filtered using product OMAERUVd, although care needs to be taken when deciding the threshold for smoke detection ([Marais2012](#)).

When analysing satellite OMHCHO vertical columns (Ω), the following steps are performed in order to mask influence from biomass burning:

1. MOD14A1 daily gridded Aqua/Terra combined fire counts are read at $1 \times 1 \text{ km}^2$ resolution, and binned into $0.25 \times 0.3125^\circ$ bins, matching the resolution of binned Ω .
2. A rolling mask is formed which removes Ω 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.
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](#); [Marais2012](#)). AAOD is should be less sensitive to

cloud contamination than AOD, and I use AAOD from the daily gridded level 3 satellite product OMAERUVd ([Ahn2008](#)) described in section [1.2.4](#) to provide a filter for smoke plumes. Although removing gridsquares with dust reduces how much data is available to analyse, it's considered a minor problem as dust in Australia is highly episodic and should not affect more than a few days per year, especially over regions with high tree coverage ([Shao2007](#)).

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 [1.20](#) 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.

1.7.2 NO_x

Enhanced NO₂ 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 NO₂ columns.

OMNO2d from 2005 is used to determine a suitable threshold for anthropogenic influence by looking at NO₂ columns near several major cities in the south eastern sector of Australia. The mean, standard deviation, and time series over Australia of tropospheric NO₂ seen by Aura is shown in figure [1.21](#). The average tropospheric NO₂ column for averaged within all of Australia and then each region shown in this figure is listed in table [TODO 1.2](#).

Anthropogenic influences on the NO₂ 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 NO₂ greater than 10^{15} molec cm⁻² are flagged as anthropogenic.
2. After taking the yearly average over Australia, any gridsquares greater than 1.5×10^{15} molec 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 [1.22](#).

The same regions as in figure [1.21](#) are shown again in figure [1.23](#), with NO₂ pixels densities for each region shown, along with the threshold of 1×10^{15} molec 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 1.5×10^{15} molec cm⁻² further reduced the available data by TODO gridsquares.

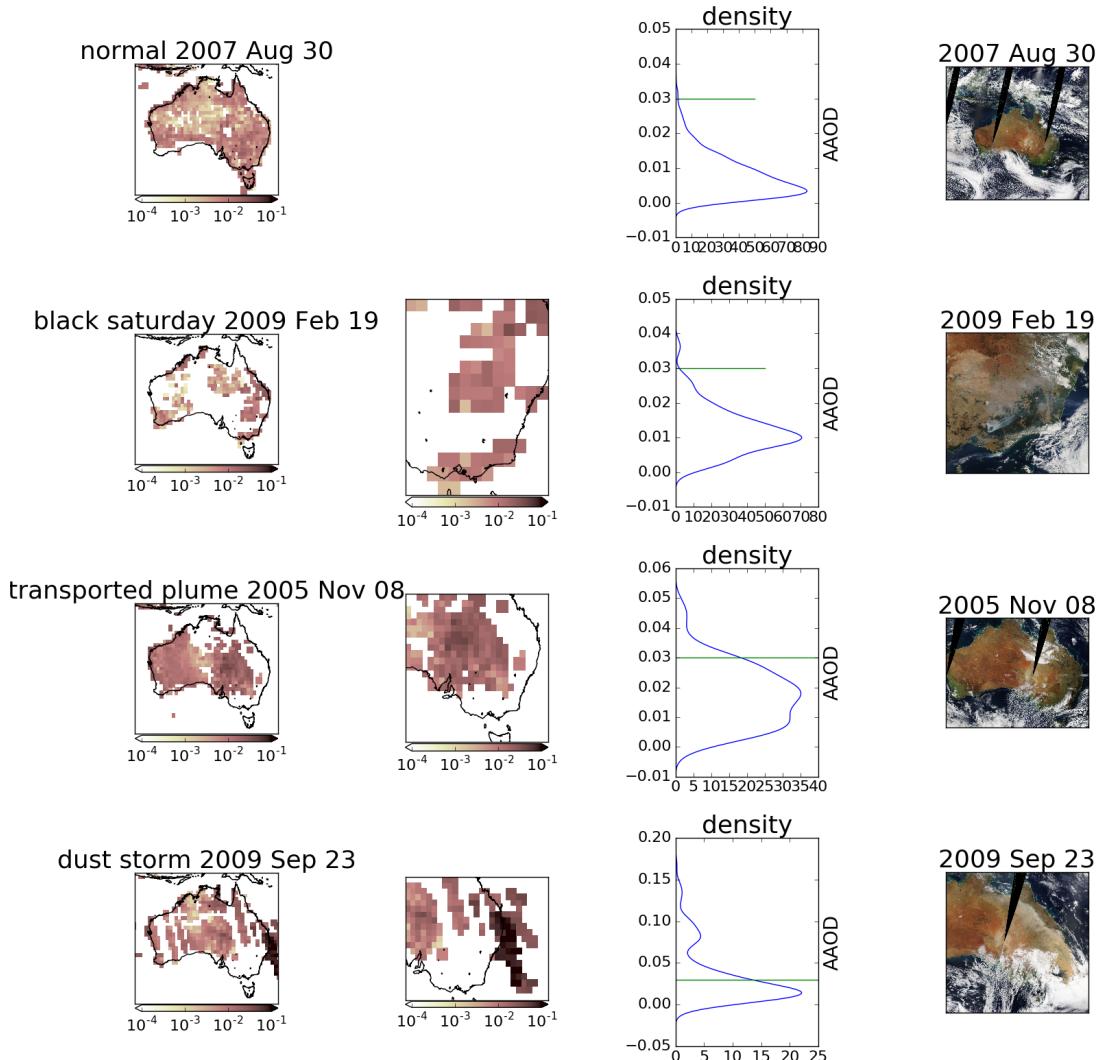


FIGURE 1.20: AAOD from OMAERUVd (columns 1, 2, 3) over Australia for four different scenarios (rows 1-4). Scenes from the same day are taken from the EOS Worldview website <https://worldview.earthdata.nasa.gov/>.

TABLE 1.2: NO₂ averages by region before and after filtering for anthropogenic emissions using 2005 data from the OMNO2d product.

Region	NO ₂	NO ₂ after filtering	% Data lost
Aus	1	2	3
BG	1	2	3
Sydney	1	2	3
Melb	1	2	3
Adelaide	1	2	3

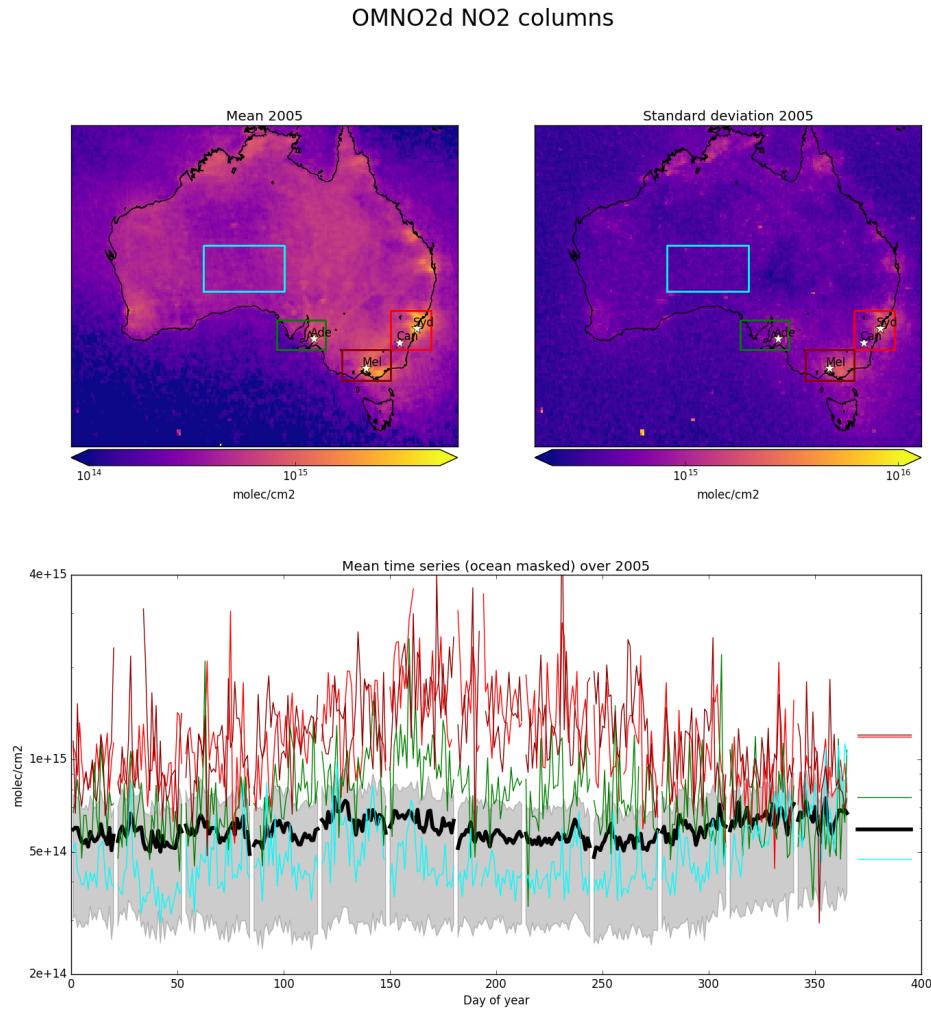


FIGURE 1.21: Mean (top left) and standard deviation (top right) of OMNO2d daily 0.25×0.25 °tropospheric cloud filtered NO₂ columns. Time series for Australia, and each region (by colour) shown in the bottom panel, with mean for that region shown on the right. A grey shaded area depicts the 25th to 75th percentiles of Australia averaged NO₂ columns for each day in the time series, with a thicker black line showing the Australia-wide mean value.

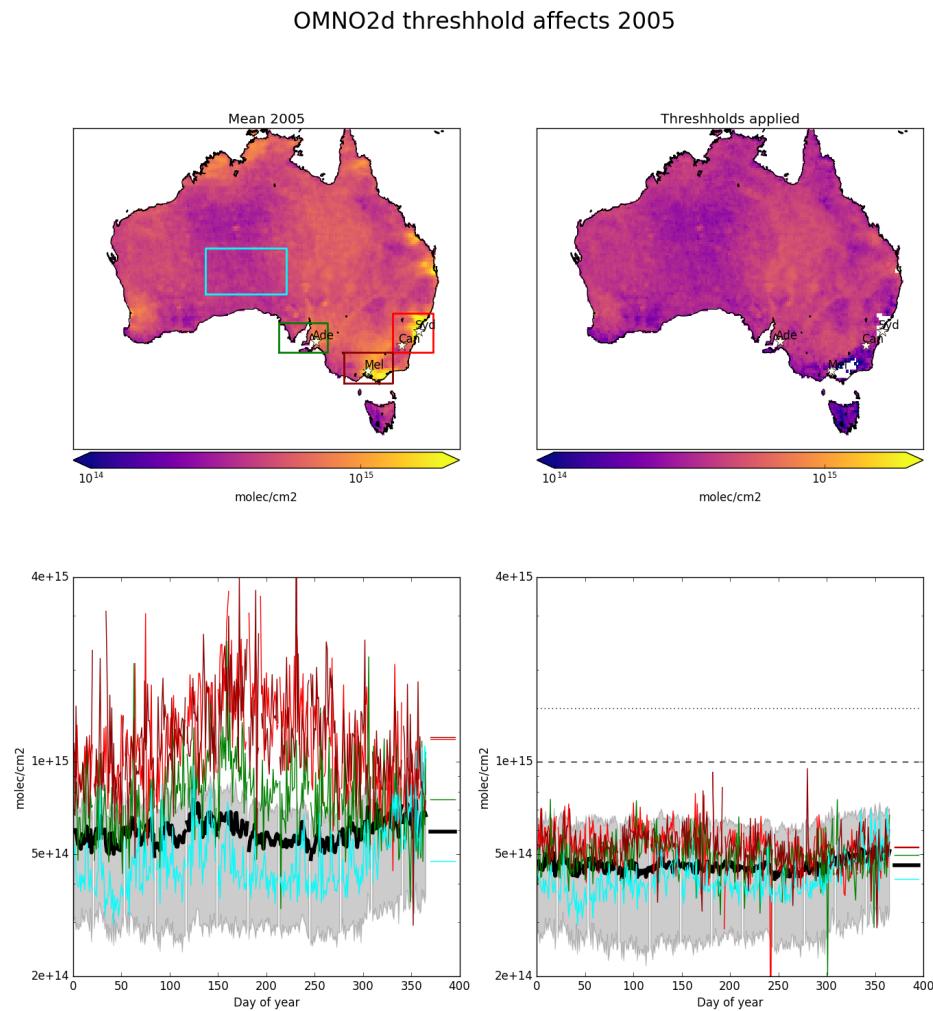


FIGURE 1.22: 2005 OMNO2d NO₂ column mean before (left) and after (right) applying the threshold filters as described in the text. Time series for Australia, and each region (by colour) shown in the bottom panel, with mean for that region shown on the right.

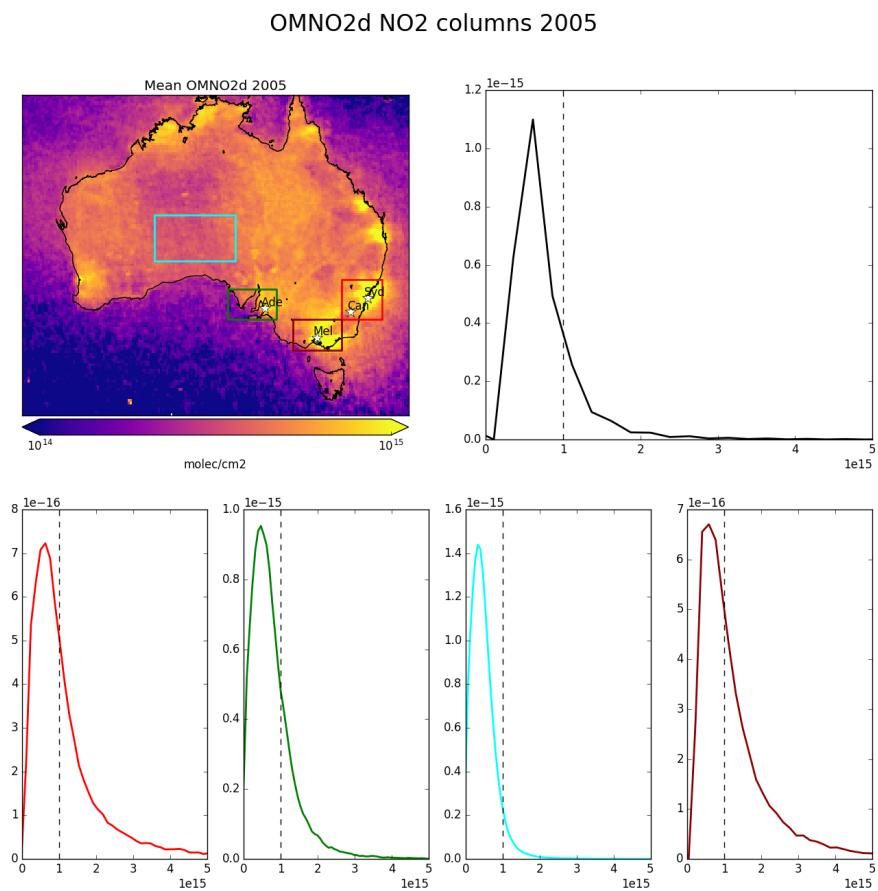


FIGURE 1.23: 2005 OMNO2d NO₂ column means (top left), along with column amount distributions for Australia (top right) and each region shown in the area map (by colour)

1.8 Analysing output

1.8.1 Circadian emissions cycle

HEMCO diagnostics provide the simulated MEGAN isoprene emissions at high temporal resolution. TODO: Figure X shows the daily emissions cycles for a few regions over each season. The regions are labelled in the top panel, and seasonally averaged emissions from grid-boxes in each region are shown below. TODO: Figure XX shows the emissions from SPS1 and 2 compared against GEOS-Chem estimates in the same grid square.

1.8.2 HCHO: Simulated vs Measured

HCHO precursors are heavily tied to temperature (TODO:cite), and model output shows how higher temperature leads to an increase in HCHO levels. Figures 1.24 - 1.26 show the relationship between temperature and HCHO, for January 2005, within subsets of Australia. A reduced major axis regression is used to determine the linear slopes between surface temperature (X axis) and HCHO (Y axis). This gives us a linear regression for each region however it's clear from the straight line and from literature that the relationship is not linear but rather exponential (TODO: cite and example studies). 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, which is not directly comparable to the linear coefficient. The distributions of exponential correlation coefficients and exponential 'm' terms is shown in the embedded plot, with one datapoint available for each grid square where the regression is performed.

Simulated GEOS-Chem tropospheric NO₂ columns averaged from 1300-1400 LT are compared against OMNO2d data (Sec. 1.2.4). Figure 1.27 shows the direct comparison between these datasets averaged over the month of January, 2005. It's clear that the OMNO2d product can pick out Sydney and Melbourne as NO₂ hotspots, which are underestimated by GEOS-Chem due to averaging over the 2x2.5 °horizontal resolution. Over much of the country GEOS-Chem overestimates NO₂ by 10-60%, except in NA and northern Queensland where up to 50% underestimation occurs. The correlation between the bias (GEOS-Chem - OMNO2d) with anthropogenic and soil emissions is shown in the bottom rows. The comparison for January and February of 2005 in Figure 1.27, and winter (JJA) of 2005 in Figure 1.27. The poor correlations for anthropogenic NO suggest that blanket alterations over Australia would not lead to improved NO₂ fit.

This comparison is expanded, including a comparison against modelled emissions, and repeated for autumn (MAM), winter (JJA), and spring (SON) in figures 1.29 to 1.36. These show an analysis of GEOS-Chem NO emissions and their correlations with the bias between GEOS-Chem NO₂ 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 model and satellite NO₂ columns is OK throughout the year over Australia, with some overestimation in the north during non-summer months.

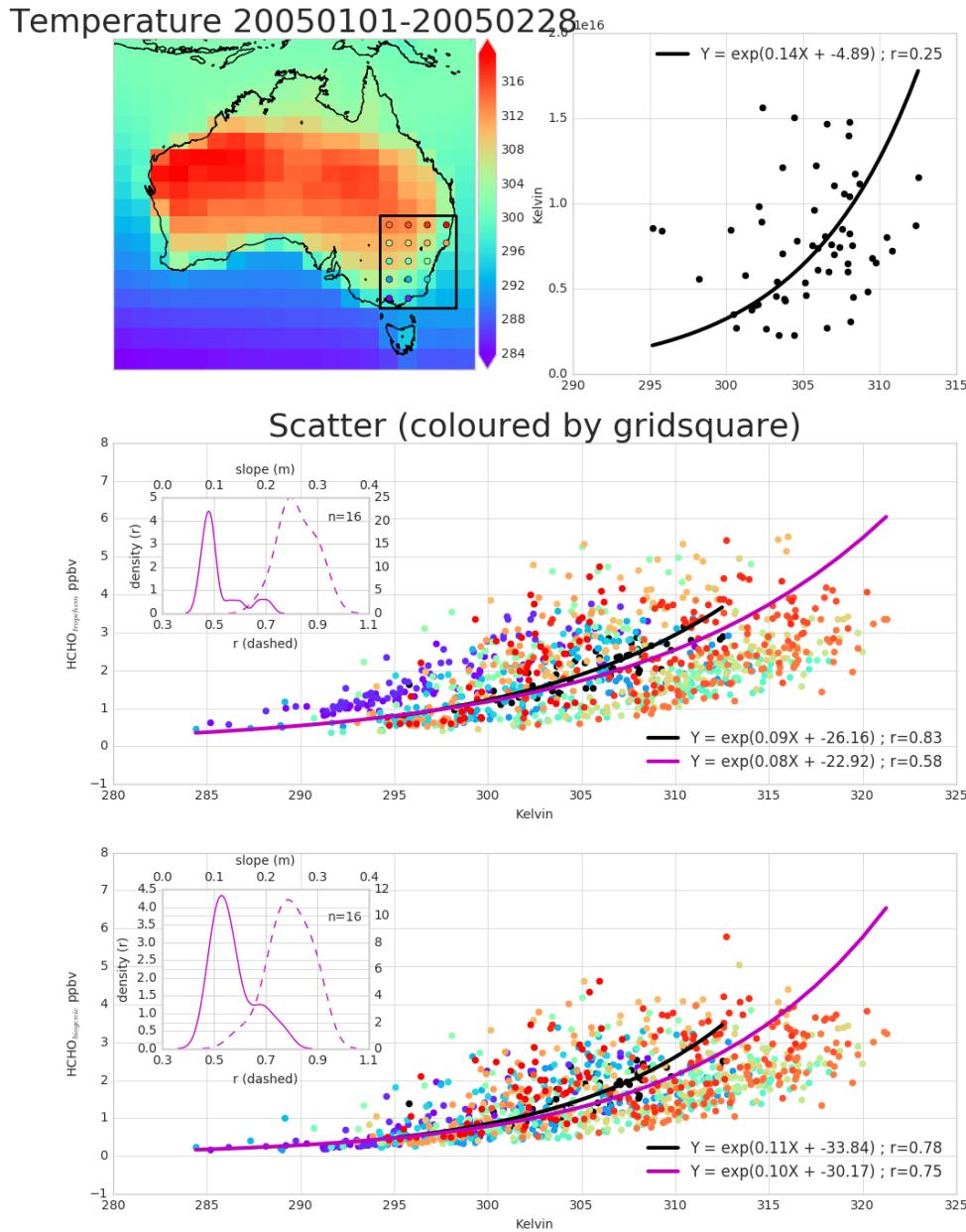


FIGURE 1.24: Top panel: surface temperature averaged over January and February 2005. Bottom panel: surface temperature correlated against temperature over, with different colours for each gridbox, and the combined correlation. A reduced major axis regression is used within each gridbox (shown in top 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.

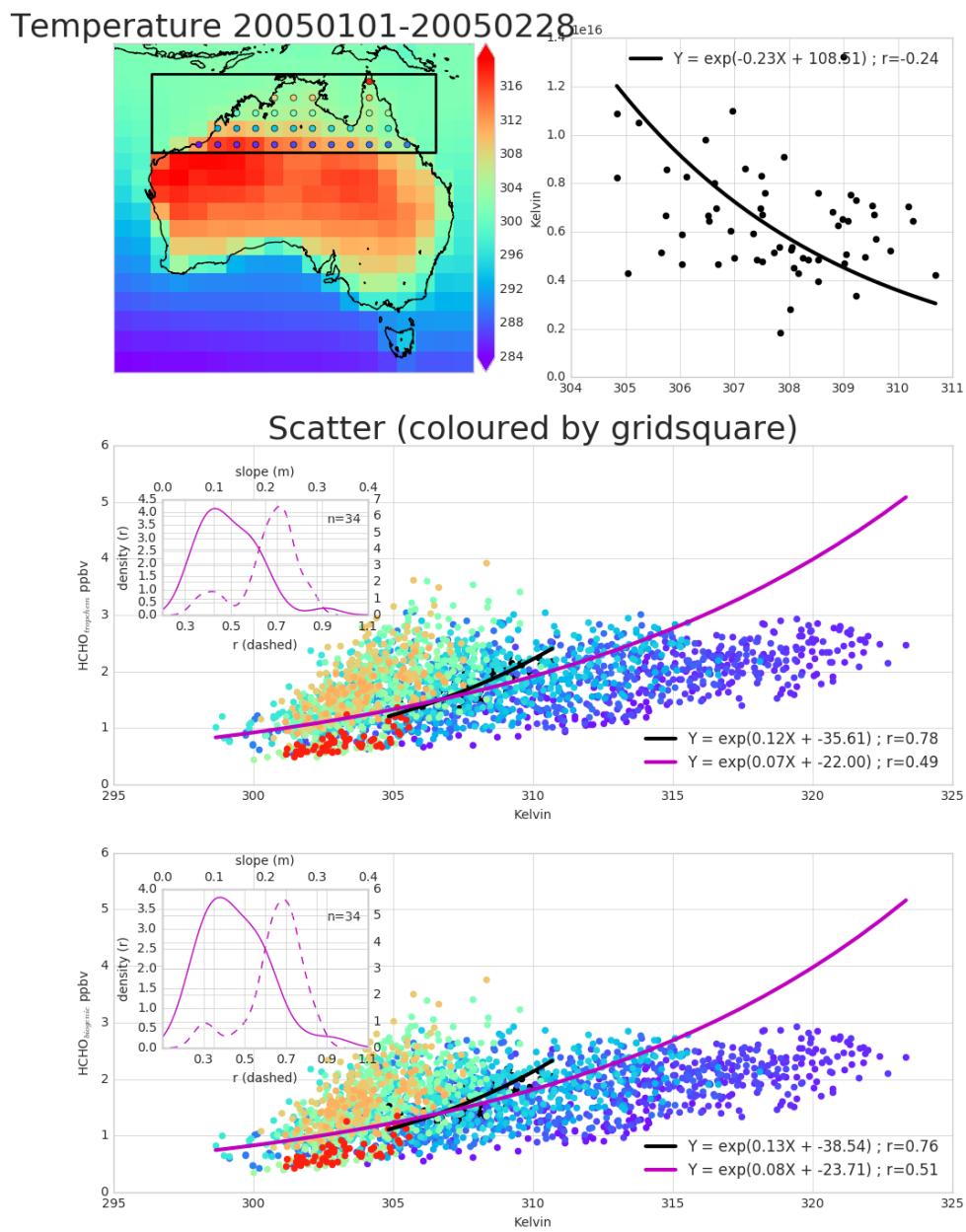


FIGURE 1.25: As figure 1.24 but for northern Australia.

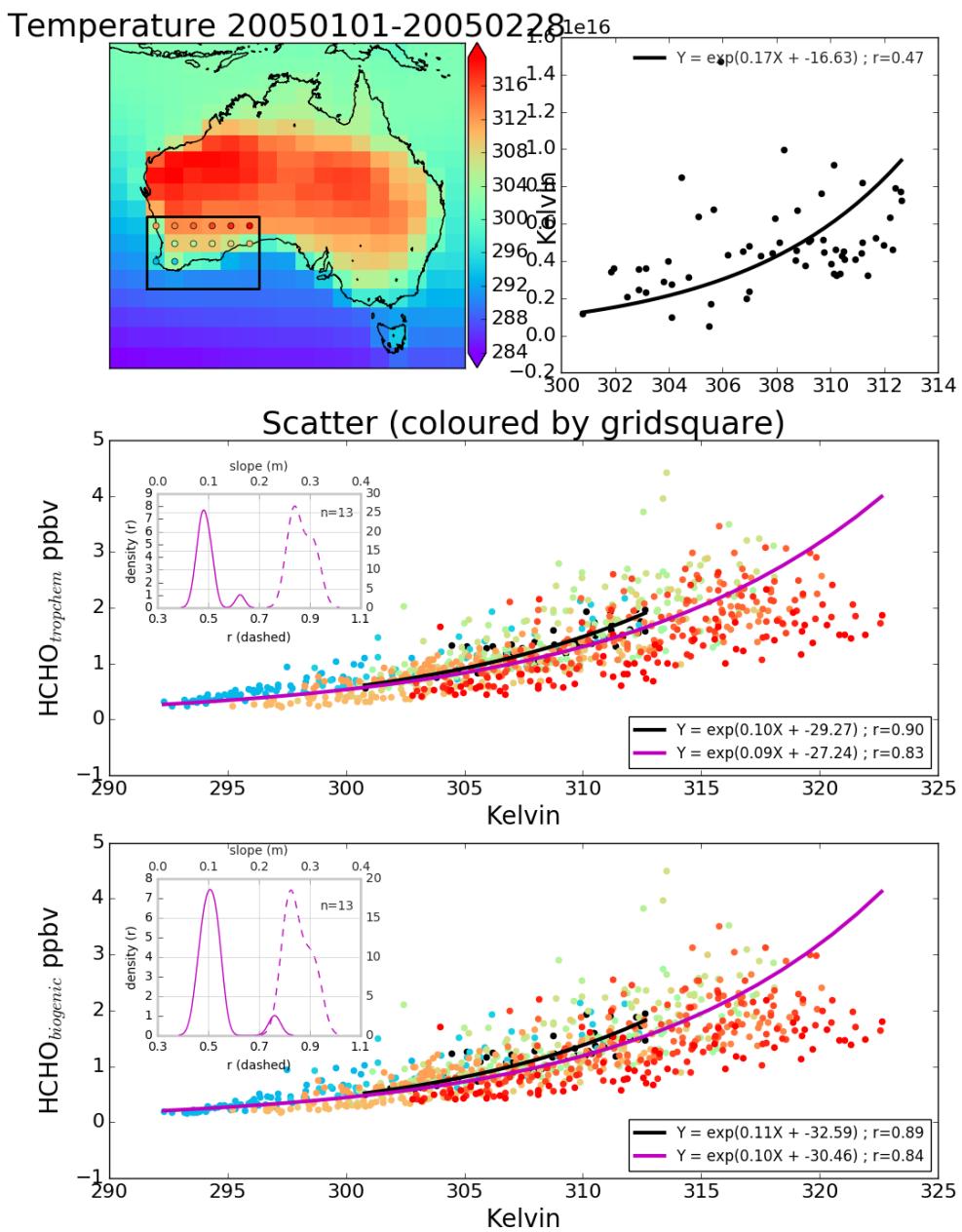


FIGURE 1.26: As figure 1.24 but for south-western Australia.

GC NO vs OMNO2d 20050101-20050228

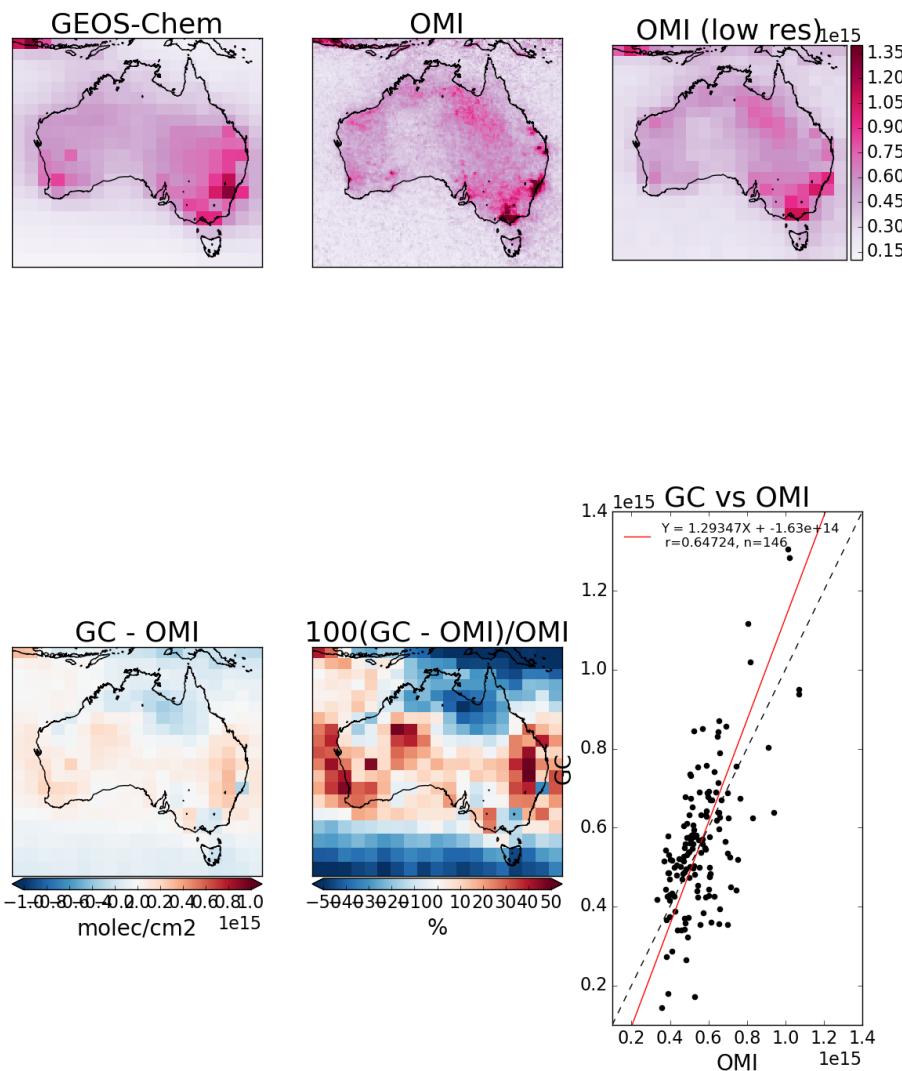


FIGURE 1.27: Row 1 shows the tropospheric columns in molec cm^{-2} , GEOS-Chem, OMNO2d, and OMNO2d averaged onto the lower resolution of GEOS-Chem from left to right. Row 2 shows the correlations of GEOS-Chem (X axes) between daily anthropogenic emissions, and mid-day OMNO2d columns. Row 3 shows the differences with OMNO2d columns averaged into the lower resolution of GEOS-Chem.

GC NO vs OMNO2d 20050101-20050228

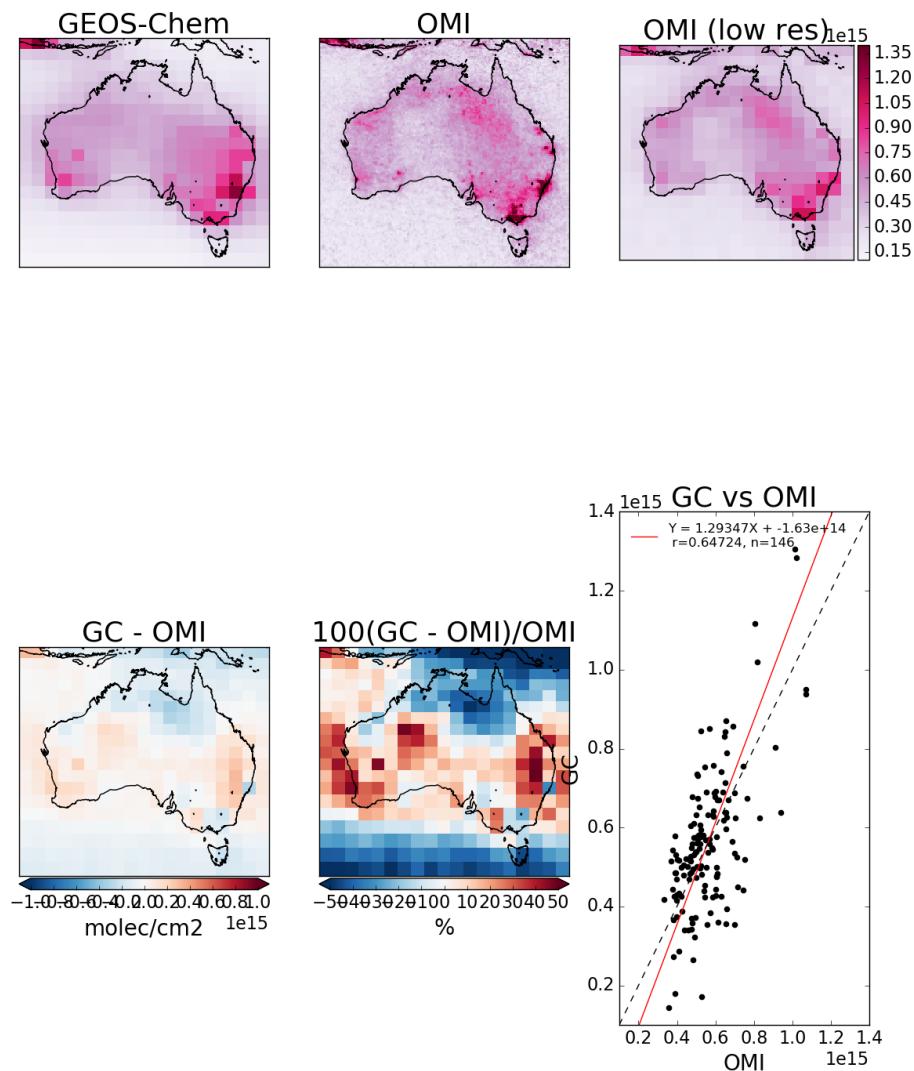


FIGURE 1.28: Row 1 shows the tropospheric columns in molec cm^{-2} , GEOS-Chem, OMNO2d, and OMNO2d averaged onto the lower resolution of GEOS-Chem from left to right. Row 2 shows the correlations of GEOS-Chem (X axes) between daily anthropogenic emissions, and mid-day OMNO2d columns. Row 3 shows the differences with OMNO2d columns averaged into the lower resolution of GEOS-Chem.

There is also slight underestimation over Sydney and Melbourne throughout the year. Figures 1.29 to 1.36 show that the visible biases are not over Australia are not driven by modelled emissions of NO. While the correlation between column NO₂ and emitted NO is clear, emissions do not appear to bias the model in either direction away from the satellite data.

The conclusion drawn is that modelled anthropogenic and soil NO emissions do not show sufficient evidence of biasing GEOS-Chem NO₂ columns away from satellite measurements over Australia. For this reason modelled NO emissions are not scaled in model runs in this thesis.

1.8.3 HCHO Comparisons

TODO: GOME2 HCHO stuff? During days with more than one HCHO column measurement we can more confidently fit the cycle. For example EOS AURA's OMI measurements from 2004 can be combined with MetOp-A's GOME2 after October 2006, with daily overpasses by OMI and GOME2 at 1345 and 0930 respectively.

1.8.4 Comparison with standard OMI product

Figure TODO: shows global and Australian HCHO eight day averaged total column maps for 1-8 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.

1.8.5 Comparison with in-situ measurements

TODO: Describe Wollongong FTIR and junk Analyse comparison of gridbox with instrument!

1.9 Data Access

TODO: ADD MORE HERE

OMNO2d Daily satellite NO₂ product downloaded from <https://search.earthdata.nasa.gov/search>, DOI:10.5067/Aura/OMI/DATA3007. See more information in section

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

OMHCHO Satellite swaths of HCHO slant columns downloaded from TODO, with DOI TODO