Project Outline

Isoprene (2-methyl-1,3-butadiene; C_5H_8) is the most abundant non-methane hydrocarbon found in the atmosphere [1]. It is predominantly emitted by terrestrial vegetation, with a global emission rate of 400-750 Tg C year^{-1} [2]. It is highly reactive due to the presence of double bonds, and readily oxidises in the atmosphere, contributing to ozone production and to secondary organic aerosol (SOA). These fine aerosol particles can have various effects, including adverse health effects, visibility reduction, climate change [3].

Isoprene, however, is unable to be measured by satellite. This means that it is difficult to obtain large spatial and temporal coverage, as measurements must be obtained in situ. Instead, formaldehyde is used as a proxy for isoprene. Formaldehyde (HCHO) is commonly formed from the oxidation of volatile organic compounds (VOCs), of which isoprene is a major component. Isoprene is also the dominant VOC driving the variability of formaldehyde [4]. Formaldehyde can also be emitted through activities such as biomass burning and fossil fuel combustion, however chemical production is the dominant source. It has a short lifetime of several hours during midday, as it is destroyed through reactions with OH and photolysis. This means that measured formaldehyde must have been produced by recent VOC oxidation [5]. Therefore, formaldehyde can be a useful proxy for isoprene.
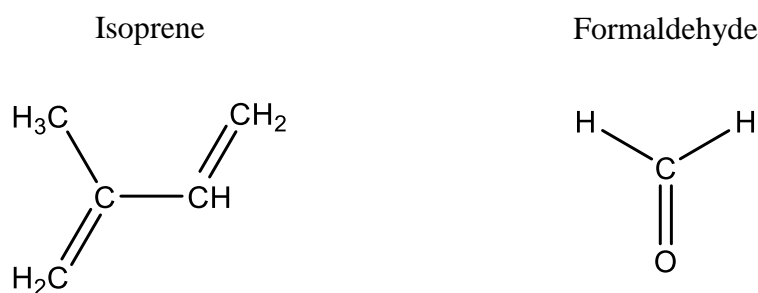

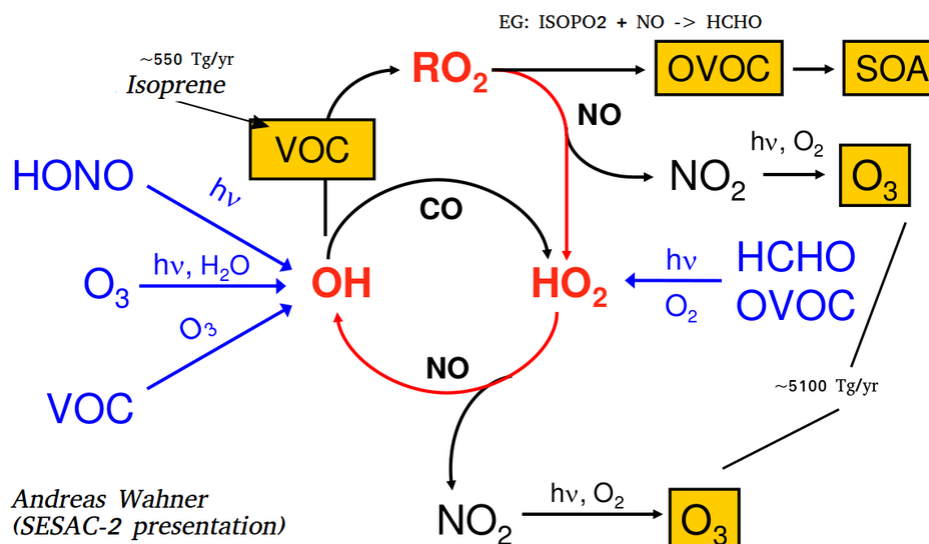*Figure 1*: Structure of isoprene (left) and formaldehyde (right).

Figure 2: Radical cycling enabled by NO, which produces formaldehyde as a byproduct. Figure obtained from Andreas Wahner and edited by Jesse Greenslade.

Formaldehyde is measured by satellite in columns. I will be using data obtained from the Ozone Monitoring Instrument (OMI) aboard the NASA Aura satellite. OMI is a nadir-viewing near-UV/Visible CCD spectrometer. As the measurements are obtained in a column, there is a potential for smearing, where the measured formaldehyde has undergone horizontal transport and no longer aligns with the location of the emission of the isoprene from which it was produced. However, due to the short lifetime of isoprene, Palmer et. al. (2003) show that smearing is not an issue, except possibly in winter when other VOC oxidations make a major contribution to the formaldehyde [6]. As formaldehyde peaks in Australia in summer, we will be focusing on the summer months.

Previous studies have been performed to create a model which can be used to predict isoprene emissions at both a global and regional level. One such model is the Model of Emissions of Gases and Aerosols from Nature (MEGAN). This model uses leaf area index (LAI), solar radiation, temperature, soil moisture, and plant functional type (PFT) to give estimates of biogenic VOCs, including isoprene and formaldehyde [7].

The highest isoprene emitters are believed to be eucalyptus trees [8]. Due to the high number of eucalypts in Australia, the isoprene emissions there are estimated to be quite high. These predicted high emissions, however, do not match in situ observations. This is a problem for air quality chemistry models, and needs to be further investigated to better inform the models. Field measurements have also shown that concentrations of isoprene and monoterpenes are of the same magnitude in South-Eastern Australia, with neither species dominating. Both of these suggest that more research should be performed focusing on Australia, as it displays some interesting characteristics [9].
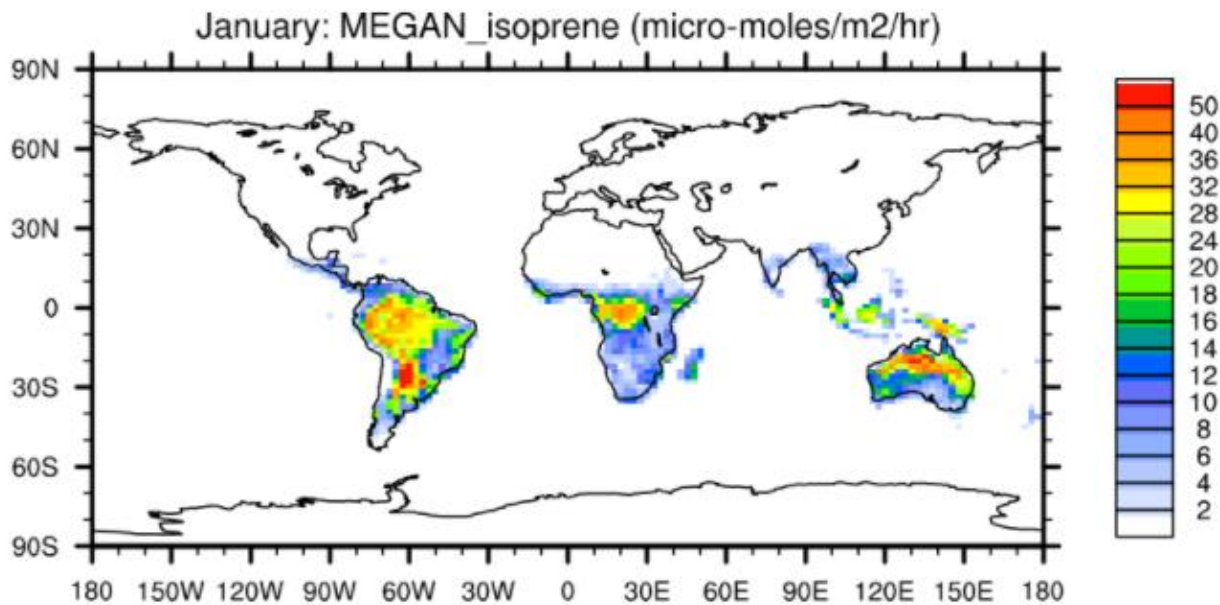

Figure 3: Global isoprene emissions for January 2000 as simulated with MEGAN2.1 algorithms. Figure reproduced from Guenther et. al. [7].

Previous studies by Emmerson et. al. (2016) have shown that MEGAN does not provide a good representation of the emissions in Australia, with isoprene emissions being overestimated, and monoterpenes, another biogenic VOC, being underestimated. There was no single scaling factor for the emission factors which could align the estimates with the observations for all seasons and conditions, suggesting that multiple emission factors need to be adjusted, or that the relationship between isoprene emission and the driving variables needs adjustment. There are several theories as to why the model gives such an inaccurate estimate, such as the fact that the emission factors of the model were based on young eucalyptus trees, which have been found to emit more isoprene than the older trees which appear in higher numbers in Australia. Other theories are the inability of the model to respond to the extremely high temperatures which Australia experiences, or its poor response to times of low soil moisture, such

as droughts [9]. As there is doubt as to which relationships cause the bad representation, this project aims to look at the various relationships between these driving variables and formaldehyde as a proxy for isoprene, without actually using the model.

This project will look for both spatial and temporal relationships between formaldehyde and the driving variables to better understand the drivers of formaldehyde variability and, by proxy, isoprene variability. This can then better inform the MEGAN model and improve it for Australia. Spatial analysis will be performed on formaldehyde, PFTs, and LAI, whilst temporal analysis will be performed on formaldehyde, soil moisture, solar radiation, and temperature.

As said previously, formaldehyde columns will be obtained from OMI. Other possible data sources, as seen in Table 1 below, include European Space Agency Climate Change Initiative for soil moisture, NASA's MODIS instrument for LAI, NCAR's Community Land Model for PFT, and NASA GEOS Reanalysis for temperature and solar radiation.

Table 1: Possible data sources for formaldehyde and the driving variables.

Variable	Possible Data Source
Formaldehyde	OMI
Soil Moisture	European Space Agency Climate Change Initiative
PFT	NCAR's Community Land Model
LAI	NASA's MODIS
Temperature	NASA GEOS Reanalysis
Solar Radiation	NASA GEOS Reanalysis

Firstly, for spatial analysis, K-means clustering will be used. K-means clustering is a useful tool for identifying groups in complex datasets without the need to specify the number of groups or what they will be sorted by. Initial cluster centres are chosen from the data. The remaining data points are added to the cluster with the closest cluster centre, then the cluster centre is updated to be the new mean of the cluster. Once all data has been assigned to a cluster, the clusters are refined through redistribution of data where there is now a closer cluster centre in another cluster. Again, the cluster centres are updated with the new mean of the cluster and data is continually refined in this way until no data has changed clusters [10].

Formaldehyde columns will be obtained by satellite and will be used as monthly means, using only the summer months. As formaldehyde can be emitted from biomass burning, any fires present during the time the data was obtained will need to be masked to ensure that the formaldehyde measured was from VOC oxidation. I will use MODIS to obtain data about fires.

This clustering will be performed individually on the formaldehyde column data and on each of the relevant driving variables. Both the mean and standard deviation will be used to ensure that I am clustering regions that co-vary. By mapping these clusters spatially, similarities in location of clusters between the variables can be seen, which can then be further investigated.

Secondly, temporal analysis will be performed. I will be looking at the interannual variability of monthly mean values by using the average of the clusters found previously. By performing linear regression with formaldehyde and each of the relevant emission factors against time, it can be seen how they change temporally. These temporal changes can then be compared between the different variables to look for similarities, and to see whether the emission factors are represented by the data. There is also the possibility of the use of multiple regression if indicated.

References

[1] Zindler, C., Marandino, C., Bange, H., Schütte, F. and Saltzman, E. (2014). Nutrient availability determines dimethyl sulfide and isoprene distribution in the eastern Atlantic Ocean. *Geophysical Research Letters*, 41(9), pp.3181-3188. DOI: doi:10.1002/2014GL059547

[2] Kameyama, S., Yoshida, S., Tanimoto, H., Inomata, S., Suzuki, K. and Yoshikawa-Inoue, H. (2014). High-resolution observations of dissolved isoprene in surface seawater in the Southern Ocean during austral summer 2010–2011. *Journal of Oceanography*, 70(3), pp.225-239. DOI: 10.1007/s10872-014-0226-8

[3] Carlton, A. G., Wiedinmyer, C., and Kroll, J. H. (2009). A review of Secondary Organic Aerosol (SOA) formation from isoprene. *Atmospheric Chemistry and Physics*, 9, pp. 4987-5005. DOI: 10.5194/acp-9-4987-2009

[4] Palmer, P. I., et al. (2006), Quantifying the seasonal and interannual variability of North American isoprene emissions using satellite observations of the formaldehyde column. *Journal of Geophysical Research*, 111(D12), p.315, DOI: 10.1029/2005JD006689.

[5] Wolfe, G., Kaiser, J., Hanisco, T., Keutsch, F., de Gouw, J., Gilman, J., Graus, M., Hatch, C., Holloway, J., Horowitz, L., Lee, B., Lerner, B., Lopez-Hilifiker, F., Mao, J., Marvin, M., Peischl, J., Pollack, I., Roberts, J., Ryerson, T., Thornton, J., Veres, P. and Warneke, C. (2016). Formaldehyde production from isoprene oxidation across NOx regimes. *Atmospheric Chemistry and Physics*, 16(4), pp.2597-2610.

[6] Palmer, P. (2003). Mapping isoprene emissions over North America using formaldehyde column observations from space. *Journal of Geophysical Research*, 108(D6).

[7] Guenther, A., Jiang, X., Heald, C., Sakulyanontvittaya, T., Duhl, T., Emmons, L. and Wang, X. (2012). The Model of Emissions of Gases and Aerosols from Nature version 2.1 (MEGAN2.1): an extended and updated framework for modeling biogenic emissions. *Geoscientific Model Development*, 5(6), pp.1471-1492.

[8] Evans, R. C., Tingey, D. T., Gumpertz, M. L., and Burns, W. F. (1982). Estimates of isoprene and monoterpene emissions rates in plants. *Botanical Gazette*, 143, pp.304–310.

[9] Emmerson, K., Galbally, I., Guenther, A., Paton-Walsh, C., Guerette, E., Cope, M., Keywood, M., Lawson, S., Molloy, S., Dunne, E., Thatcher, M., Karl, T. and Maleknia, S. (2016). Current estimates of biogenic emissions from eucalypts uncertain for southeast Australia. *Atmospheric Chemistry and Physics*, 16(11), pp.6997-7011.

[10] Brodley, C. (2001). Proceedings of the eighteenth International Conference on Machine Learning (ICML-2001), June 28-July 1, 2001. San Francisco, CA: Morgan Kaufmann Publishers, pp.577-584.