MODELLING GLOBAL ISOPRENE EMISSIONS WITH A SIMPLIFIED PROCESS BASED APPROACH DRIVEN BY THE P-MODEL

Bikem Pastine

A thesis submitted in partial fulfilment of the requirements for the degree of Master of Science in Computational Methods in Ecology and Evolution.

Imperial College London

August 2023

Declaration and Acknowledgements

The data used in this project are available freely online, and I am responsible for assembling, processing, and cleaning the data. Details of the sources can be found in Appendix 2 and are referenced throughout the project.

Thank you to Shirley Cai for guiding me in assembling the data to run the P-Model. Thank you to Dr. Sandova for providing the soil moisture data required to calculate the β parameter and for his time explaining the SPLASH model so patiently. Thank you to Dr. Orme for his insights when developing this project, they are much appreciated.

The model presented in this project is a simplified version of the isoprene model developed by my supervisor Dr. Morfopoulos. The modelling details were developed together with Dr. Morfopoulos. I am responsible for the integration of the model into the P-model and the design of the subsequent computational campaign. Dr. Morfopoulos advised me in developing the analysis and patiently and kindly guided me to the world of vegetation modeling. This project is inspired by her work, and I too am inspired by her. Thank you.

I would like to thank my Zelimhan Akhmiev for his wisdom and kindness. I hope I can return your generosity, gentleness, and support when you inevitably study for your third master's degree.

Finally, thank you to my parents. Thank you for reading the drafts of this work, and your endless emotional support. Thank you for believing in me and pushing me to be the smartest and most thoughtful version of myself. Most of all, thank you for giving me this opportunity to study, my world has tripled in size thanks to you.

Abstract

Isoprene is a biogenic volatile organic compound emitted by plants that significantly impacts the Earth's system. This study introduces a simplified adaptation of the Energetic Status Model (Morfopolous et al., 2014) for isoprene emission, assimilated into the P-Model for vegetation productivity (Stocker et al., 2020). While the simplified global-scale isoprene emission model omits plant functional type-based basal emission parameterization, it accurately mirrors aggregated spatial and temporal fluctuations in isoprene emissions. It demonstrates superior performance to MEGAN—a widely applied empirical isoprene model—when compared with an independent dataset. The model captures biome-specific emission patterns in most cases and indicates that integrating land cover-based parameterization might further enhance the model's accuracy in specific areas. This research suggests a re-evaluation of the prevalent reliance on plant functional type-based parameterizations. It also points to directions of future research based on variables of importance identified using LASSO regularized linear regression.

Table of Contents

Introduction

Isoprene is a volatile compound produced in leaves and emitted at a rate approximately equal to methane (McGenity et al., 2018). It has the molecular form 2-methyl-1,3-butadiene and its carbon-carbon double bonds make it highly reactive (Morfopoulos, 2014). Because of its reactivity, isoprene is influential in both the atmosphere (Wells et al., 2020) and in plant physiology (Lantz et al., 2019). Accurate and robust models of isoprene emissions are therefore important for improving earth system models (Pacifico et al., 2009).

Ninety percent of isoprene emissions are from terrestrial plants (Pacifico et al., 2009). The carbon cost of isoprene is high and so isoprene is primarily produced by fast growing species (Harrison et al., 2012): about 2% of carbon removed from the atmosphere during land ecosystems' photosynthesis is re-emitted as isoprene (ibid.). While isoprene synthesis is a trait found in all climates and types of species (Lantz et al., 2019), the highest emissions are from tropical and temperate forests, and the lowest from agricultural land and the boreal zone (Pacifico et al., 2009).

Isoprene emissions benefit plants through increased heat tolerance, protection against oxidative stress, and improved membrane stability (Pollastri et al., 2014). Some literature suggests that these benefits are the direct effects of isoprene on the physiology and chemistry of the thylakoid membrane (ibid.). Alternatively, isoprene may act as a signaling molecule which alters gene expression and thus offers secondary benefits (Zuo et al., 2019). While the exact mechanisms remain elusive, isoprene offers an advantage to emitting plants, especially in environments with high temperatures (Lantz et al., 2019).

Modelling isoprene is not only useful due to its effects on plants, but also due to its influences on the atmosphere and the wider climate system (Morfopoulos, 2014). The highly reactive nature of isoprene makes it a key factor in modulating atmospheric oxidation capacity (Wells et al., 2020). Isoprene competes for hydroxide and nitric oxide in the atmosphere and thus increases the half-life of greenhouse gases like tropospheric ozone and methane (Pasifico et al., 2009). Isoprene also increases the atmospheric concentration of biogenic secondary aerosols, which impact cloud cover, albedo, and precipitation patterns (Lantz et al., 2019). It has been argued that the majority of the model uncertainty in future air quality projections comes from the uncertainties in isoprene estimates (Gomez et al., 2023). Thus, most modelling efforts have been focused on its atmospheric implications.

This project aims to improve our understanding of biogenic isoprene emission using a vegetation modelling perspective. Literature taking this approach is still emerging and these efforts have the potential to yield two key benefits beyond improving emission models:

- Constraining isoprene emissions may improve vegetation productivity estimates by encorporating the losses and gains in productivity from isoprene emission. Empirical isoprene models predict that biogenic isoprene emissions will double by 2100 (Guenther et al., 2006), so isoprene may become even more important when projecting vegetation productivity in a changing climate.

- Accurate isoprene emission estimates may improve solar-induced chlorophyll fluorescence (SIF) based Gross Primary Productivity (GPP) models which require non-photochemical quenching (NPQ) to be estimated. As xanthophyll—the key pigment in facilitating NPQ—and isoprene are both produced by the methylerythritol 4-phosphate (MEP) pathway, isoprene is ideally positioned to act as a proxy for NPQ (Penuelas et al., 2013; Morfopolous et al. 2021). Such SIF based models would be better future proofed than traditional GPP models as they would more accurately capture large scale vegetation responses to climate extremes (ibid.).

Approaching isoprene emission modeling from a vegetation productivity perspective promises improvements to the emission estimates, vegetation productivity estimates, and may open a new avenue for GPP modelling techniques.

Isoprene Models in the Literature

Approaches to isoprene emission modeling in the plant sciences generally fall into two categories: the empirical approach developing on the G95 model proposed in Guenther et al. (1995) and the mechanistic approach which is based on the N99 photosynthesis driven model (Niinemets et al., 1999).

The G95 and the MEGAN Model

In the G95 algorithm, isoprene is modelled in an empirical form. The model relies on predicted vegetation productivity metrics such as Leaf Area Index (LAI), net primary productivity (NPP), and land cover obtained using satellite remote sensing data. Each vegetation cover is assigned a basal emission rate at standard conditions. These rates are determined in branch enclosure studies for a limited number of plant functional types (PFT) (Guenther et al., 1995). The basal rates are modified through serial multipliers in the model according to ambient temperature and light conditions. G95 is the first model to account for the effect of light on isoprene emissions and so leads the way for all the empirical and mechanistic models that follow (Sharkey and Monson, 2017).

The Model of Emissions of Gases and Aerosols from Nature, known as MEGAN (Guenther et al., 2006), dominates the literature on isoprene and has been used in several earth systems models (Weber et al., 2023). Though the modelling system has several forms based on specific parameterizations, it is similar to G95 in its general form. In MEGAN, vegetation cover is simulated by vegetation productivity models and PFT based emission factors are modified by serial multipliers relating to leaf age, soil moisture level, solar angle, relative humidity, wind speed, light availability, temperature, past temperature and light, and more (Guenther et al., 2006).

MEGAN's isoprene predictions vary up to a factor of four depending on the specific parameterizations (Langford et al., 2017). MEGAN also struggles to predict isoprene increases during droughts (Filella et al., 2018) and is highly sensitive to the estimated baseline emission factors (Situ et al., 2014). The model is vulnerable to errors because of its overparameterization and its empirical approach (Harrison et al., 2013).

The N99 and the Energetic Status Model

Niinemets et al. (1999) presents a physiological model of isoprene emission in which isoprene production is determined by the electron transport rate from Photosystem II and the energy requirement for isoprene synthesis. N99 postulates that there is a competition for reducing power between photosynthetic sugar production, MEP, and other pathways. This theory allows the model to draw redox equivalencies and

chemically derive the energy requirement of isoprene as compared to sugar synthesis. N99 obtains input parameters from process-based leaf photosynthesis models and ambient temperature measurements (Niinemets et al., 1999).

The model takes the form:

$$I = \varepsilon \times \alpha \times J \times f(T) \quad (1)$$

Where I is the isoprene emission rate, ε is the fraction of electrons directed towards isoprene production, α is the equation for the reduction requirement for isoprene production, J is the light limited photosynthetic electron transport rate, and $f(T)$ is a function of isoprene synthase activity algorithm developed by Guenther et al. (1995).

The N99 model is adapted as the basis of other semi-mechanistic models including Arneth et al. (2007) which adds empirical parameters relating to leaf age and CO2 inhibition. In the literature, N99 based models also modulate ε by PFT based basal emission rates and when run at a global scale, emissions are related to A_J (the light limited assimilation rate) instead of J (Pacifico et al., 2011). Variations of N99 are included in a small set of earth systems models (Cao et al., 2021).

This project is based on the M14 Energetic Status Model (Morfopoulos et al., 2014) which improves upon key mechanistic inconsistencies in the N99 model, specifically within the terms α and ε. Morfopoulos (2014) argues that N99 links isoprene production too closely to carbon assimilation and diminishes the importance of the MEP pathway. It further argues that the relation of emissions to A_J at a global scale lead to an inconsistency in the model whereby the photosynthetic system is assumed to be substrate limited, even when the system is light limited (ibid.).

The M14 model parameterizes the fraction of electrons used in isoprene production as a function of the balance between the total supply of photosynthetic reducing power (approximated as J) and the electron demand for carbon fixation (J_v) known as the energetic status of the plant. When the demand of rubisco for reducing power exceeds the supply of electrons, then most electrons are diverted to sugar synthesis. However, when photosynthesis is limited by rubisco regeneration and there is an excess in the electron supply, more electrons are used to produce isoprene (Morfopoulos et al., 2014).

The model is given by:

$$I = \varepsilon \times J \times f(T) \quad (2)$$

$$\varepsilon = c_1 + c_2[J - J_v] \quad (3)$$

Where J and $f(T)$ are the same as in the N99 model, J_v is the rubisco limited electron transport, and c_1 and c_2 are empirical scaling parameters (Morfopoulos et al., 2014).

Although similar to the N99 model, the M14 model represents the key patterns of isoprene emissions without any additional parametrization. For instance, the model reproduces variations in isoprene emission when there is an excess of light, the temperature peak of isoprene emission, and the decreases in isoprene emission with increasing CO_2. These are features of isoprene emission behavior that extensions of the N99 model address through empirical multipliers (Pacifico et al., 2011). The M14 model also captures the increase in isoprene emission during mild drought, a feature few other models are able to replicate.

Despite the incomplete understanding of the mechanisms governing isoprene production, different modelling approaches produce global isoprene emissions estimates that are remarkably similar (Arneth et al., 2008). This may be because both empirical and mechanistic models use the PFD based parameterization of basal emission rates (ibid.). The parameterization schemes for PFT require researchers to subjectively lump species into limited PFT categories, meaning that the model results are inevitably highly adaptable at the stage of this categorization (ibid.). PFT based parameterization is the largest known source of model error in isoprene models, introducing an error of at least ± 10% into global estimates (Pacifico et al., 2011).

To this author's knowledge, this thesis presents the first global-scale isoprene-emissions model using a simplified version of the M14 model and is the first to model emissions globally without relying on PFT based emission parameterization. I introduce a simplified, general model of isoprene emission, termed the F-Model, and investigate the efficacy of the foundational elements of our current mechanistic understanding of isoprene emission in plants at a monthly timestep, and at a 0.5° by 0.5° spatial resolution to identify areas of strong and poor model fit. Areas where the model performs well can be capitalized on while areas of weak model fit reveal some insight, guiding further research to improve our understanding of the processes governing isoprene production.

Methods

The F-Model: A simplified process-based model for isoprene emission

The F-model is a simplified version of the M14 model (Morfopolous et al., 2014). Instead of a variable value for ε, this model assumes a constant isoprene factor, F, that is calculated empirically and is given by:

$$I = F \times J \times f(T) \tag{4}$$

Here, I represents isoprene emission, J denotes light-limited electron flux, and $f(T)$ is the isoprene synthase temperature dependency algorithm described by Niinemets et al. (1999). Appendix 1 presents the form of $f(T)$. The standard temperature input for this function is set to 30°C (ibid.).

The underlying assumption of the F-Model is that some constant fraction, F, of the total electrons generated in Photosystem II are used in the MEP pathway to produce the precursor to isoprene, DMADP. Subsequently, isoprene synthase converts DMADP to isoprene following its enzyme kinetics function, $f(T)$.

This model is not a complete or exact representation of plant physiology, and hence it cannot provide reliable predictions of future isoprene emissions. The purpose of the model is to investigate the degree to which broadscale global isoprene emission patterns are captured by this simplified model, both spatially and temporally. The key anticipated shortcoming of this model is that it is unlikely that F is constant through space and time as it is expected to vary according to energetic status, vegetation cover, soil moisture stress, the electron demand of other non-photosynthetic pathways, and more.

To run the F-Model at a global scale, J needs to be calculated using a global vegetation model. I use the P-model for this purpose.

The P-Model

The P-model for vegetation productivity combines the Farquhar–von Caemmerer–Berry model of photosynthesis (Farquhar et al.,1980; von Caemmerer & Farquhar, 1981) and light use efficiency (LUE) modelling with an optimality principle that governs the carbon assimilation versus transpiration trade off in plants (Stocker et al., 2020). This model is primarily mechanistic, is forced with satellite data, and is not reliant on parameterization based on vegetation type classification. It predicts GPP as accurately as parameter heavy, empirically driven models (ibid.). The pyrealm package in python is used to run this model, the package is available for download on GitHub (github.com/ImperialCollegeLondon/pyrealm).

The P-model is governed by the coordination hypothesis. Under mean daytime conditions over weekly time scales, this hypothesis states that plants operate close to the intersection of light-limited and rubisco-limited assimilation rates, (A_J) and (A_v), respectively (Prentice et al., 2013; Wang et al., 2017). In absence of water stress, A_v equals A_J at the monthly timescale of this study. This is noteworthy in relation to the M14 model in which J and J_v are defined from the Farquhar model as follows:

$$J = 4A_J \frac{C_i + 2\Gamma^{\cdot}}{C_i - \Gamma^{\cdot}} \quad (5)$$

$$J_v = 4A_v \frac{C_i + 2\Gamma^{\cdot}}{C_i - \Gamma^{\cdot}} \quad (6)$$

Hence, under the coordination hypothesis, J_v and J are equal. As a result, it is not possible to simulate the energetic status by this method using the P-model when water stress is not modelled. The P-model remains useful to power the proposed F-Model as the energetic status term is not included.

I calculate J by Equation (5) using the parameters generated by the P-Model. The model is run without accounting for the C4 photosynthesis pathway. Soil moisture stress is also not accounted for despite the literature on the impact of drought stress on biogenic isoprene emissions (Jiang et al., 2018). This is because the P-Model applies the soil moisture stress parameter (β) as a retroactive penalty on GPP (Stocker et al., 2020) and does not supply a built-in method to back calculate other parameters. Nevertheless, β generated from the SPLASH model (Davis et al., 2017) and supplied by Dr. Sandoval are used in the analysis of the F-Model results.

Meteorological Data

I use reanalysis data on meteorological conditions to run the modified P-model from 2005 to 2014. The information on the data sources is in Appendix 2. The data are averaged to a monthly temporal resolution and to a 0.5° by 0.5° spatial resolution using the CDO toolbox in bash.

Isoprene Observations

I compare my model predictions to isoprene emission estimates from two data products henceforth referred to as OMI and CrIS.

OMI is a product based on formaldehyde (HCHO) measurements by the Ozone Monitoring Instrument (OMI) sensor on the Aura satellite (Bauwens et al., 2016). HCHO is an oxidation product of isoprene but is also produced by other pathways such as fires and anthropogenic sources (ibid.). OMI HCHO

observations are used to back calculate isoprene emissions by source inversion modelling using the IMAGESv2 chemistry transport model (Stavrakou et al., 2013), the GFED4s fire inventory (Randerson et al., 2018), and the MEGAN-MOHYCAN inventory of biogenic isoprene emission (Guenther et al., 2006, 2012; Muller et al., 2008). The OMI data is available monthly from 2005 to 2014 at a 0.5° by 0.5° spatial resolution.

CrIS is based on spectroscopic measurements of isoprene in the atmosphere by the Cross Track infrared Sounder (CrIS) on board the Suomi NPP satellite (Fu et al., 2019). Spectral signals of isoprene are converted to isoprene columns using the full-physics MUSES algorithm (Fu et al., 2018). The CrIS data is promising, but, due to its recent release, few independent studies quantify its reliability at a global scale. The CrIS data is available monthly from 2012 to 2020 at a 0.5° by 0.5° spatial resolution (Wells and Millet, 2022).

Computation

The F-Model/P-Model is run using python and bash on the Imperial HPC service. The results are analyzed using a combination of R and python. Figures are produced in Matplotlib in python.

Results

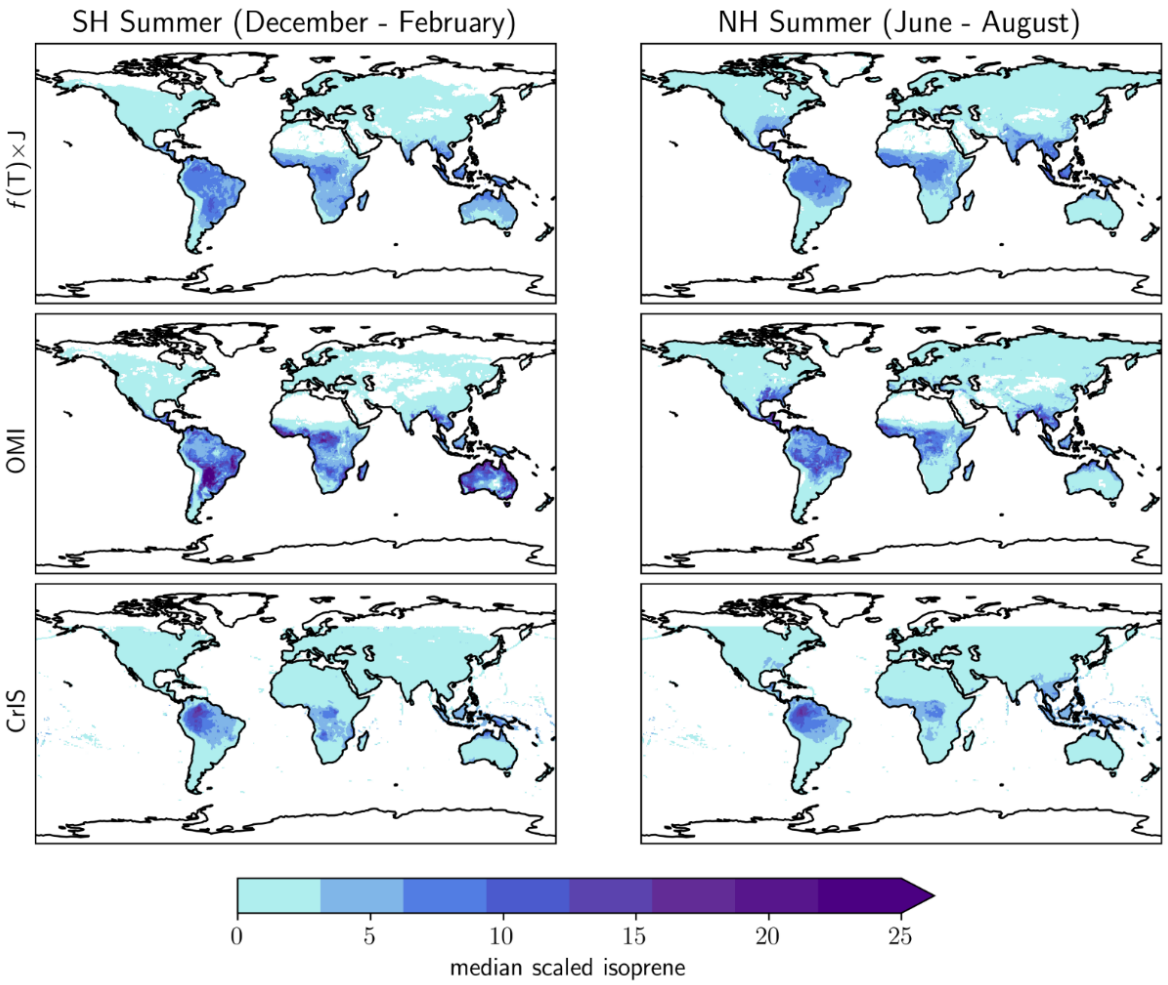
To allow for direct comparison, the data is subset to the overlap between OMI and CrIS; the results from 2012 to 2014 (inclusive) are presented.

The data is cleaned before analysis. Extreme emission events are expected during periods of high stress (Morfopoulos et al., 2021), so traditional outlier detection is inappropriate. Instead, tight data clusters at extremely high or low values—identified by visual inspection—are interpreted as sensor errors and hence removed. The cleaned datasets consist of roughly 1.8 million data points in the F-Model, 1.9 million in the CrIS, and 1.6 million in the OMI dataset.

Since the constant *F* in the F-Model is an empirical parameter, its estimation requires isoprene observations, the very data I use to evaluate the model. So, it is not possible to compare the model's absolute level with OMI or CrIS. Hence, I focus on the correlations between them. The data is presented in a standardized form while correlations are run on the original units. The units of the F-Model are μmol isoprene/m^2/month, OMI is in kg isoprene/grid cell/month, and CrIS reports molecules/cm^2/month. The OMI data units are converted to 'kg isoprene/m^2/month' to remove the area distortion bias resulting from the angular rather than distance-based sizing of the grid cells. I then scale the data by the median of each dataset. Median scaling is preferred over maximum or mean scaling due to the right skewed nature of isoprene emissions.

Spatial Accuracy

Figure 1 shows the average spatial patterns of isoprene emissions for the F-Model and the OMI/CrIS measurements.



227 **Figure 1***: Comparison of the median scaled average isoprene emissions during Southern Hemisphere*
228 *(SH) summer and Northern Hemisphere (NH) summer from the F-Model, OMI, and CrIS. Refer to*
229 *Appendix 3 for monthly average emission maps from the F-Model (2005-2014).*

230 The F-Model and the isoprene products share similar broad scale emission patterns: isoprene emissions in
231 the SH summer are consistently high in Amazonia, the Congo River Basin, Northern Australia, and
232 Southeast Asia; and in the NH summer, Amazonia, the Congo Basin, and Southeast Asia are isoprene
233 hotspots, and there are high emissions in parts of South Asia and the Coastal Plain of North
234 America. Additionally, the seasonal shift from SH high emissions to NH high emissions in their
235 respective summers is reproduced by the F-Model. However, there are also noticeable differences
236 between the model and the isoprene products, namely differences in the transition gradients between low
237 and high emitting areas and the intensity of emission hotspots.

The F-Model exhibits a more gradual gradient from high emitting to low emitting regions, while the isoprene products show abrupt switches in emission paradigm. This abruptness could be due to land use differences that are not accounted for by the F-Model. For instance, during the NH summer in the southern USA, the model shows smooth emissions along the southeastern mixed forests while OMI shows a strip of low emissions coinciding with the Mississippi Alluvial Plain which is primarily used in agriculture (Environmental Protection Agency, 2023). As broadleaf crops are not isoprene producers (Sharkey et al., 2007), the drop in OMI estimates in this region is expected but cannot be explained by the F-Model. This is visualized at a higher resolution in Appendix 4B.

The intensity of the emission hotspots is another observable difference. For instance, the northern Amazon hotspot in the SH summer is more extreme in the CrIS data than the F-Model output. This is especially true for the OMI dataset which has more intense isoprene emission extremes than both CrIs and the F-Model. This observation is supported by the Fisher-Pearson skewness coefficients of the datasets: the skewness of the F-Model output is 1.3 compared to 2.9 and 2.6 for OMI and CrIS, respectively. While all the data are right skewed, the F-Model does not reproduce emissions as extreme as observed in OMI and CrIS. This could be because the F-Model does not account for stress conditions when excess light energy may be diverted to isoprene synthesis.

There is a notable discrepancy between the CrIS and OMI emission patterns and the underlying causes warrant further investigation beyond this project's scope. In some instances where a pattern is present in one dataset and absent in the other, the F-Model falls between the two. For example, in the SH summer in South America OMI has a major isoprene hotspot in the Gran Chaco Forest while CrIS does not present it. Additionally, OMI has a pattern of lower emissions along the Amazon River Basin while the CrIS does not. In these instances, the F-Model reproduces both patterns (see at a greater resolution in Appendix 4A).

For a quantitative comparison of the F-Model to the isoprene products, I perform Pearson’s correlations on a number of subsets of the data. The data satisfies the assumptions of Pearsons correlation. The overall correlation coefficients for $f(T) \times J$ and OMI is 0.705 and 0.724 with CrIS ($p<0.001$). OMI and CrIS have a correlation coefficient of 0.503 ($p<0.001$)

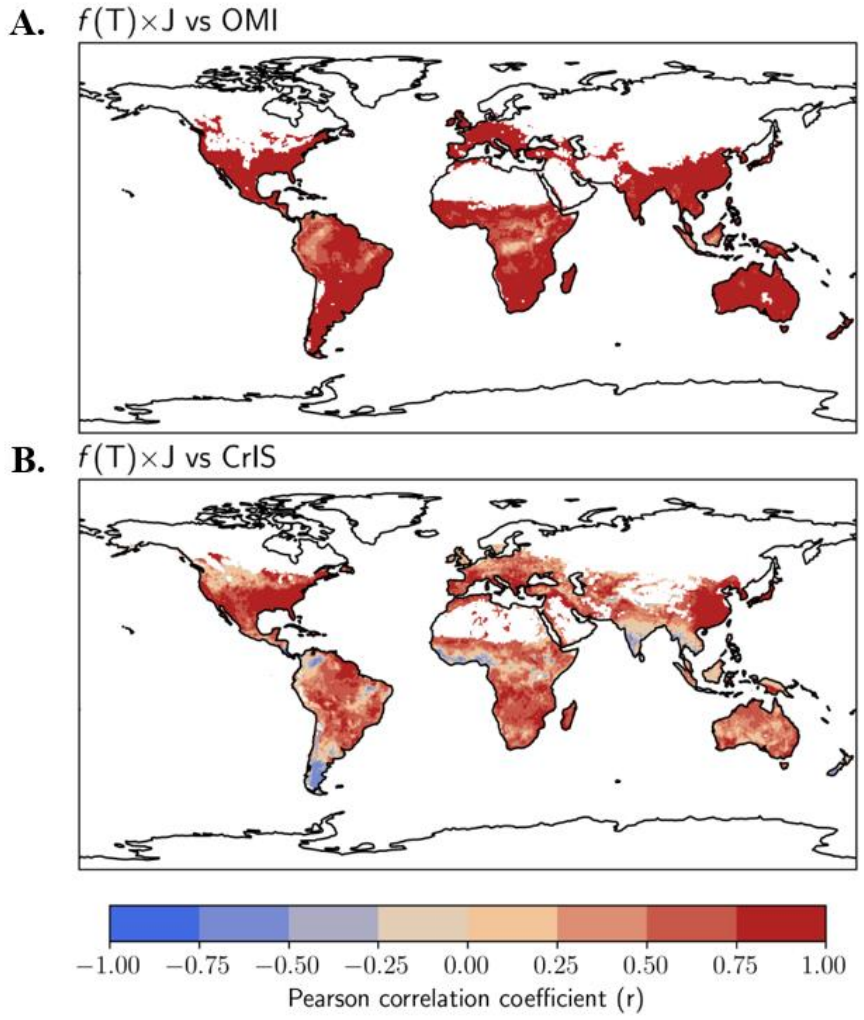


Figure 2*: Pearson's correlation coefficient by grid between* ***A.*** *the F-Model and OMI, and* ***B.*** *the F-Model and CrIS. The p-value map is in Appendix 5.*

Figure 2 presents the spatial variation in the quantitatively measured F-Model fit when compared to OMI and CrIS. The F-Model is highly correlated with OMI in most places, while it diverges from CrIS in many geographical locations.

Areas of low model performance offer insight into potential improvements. The F-Model is negatively correlated with CrIS at the southern tip of India and South America, as well as along the Gulf of Guinea. These regions coincide with mosaiced cropland and evergreen broadleaf forests (Acharya and Punya, 2013; Eva et al., 2004; Ardo, 2015). Additionally, both maps indicate a poor fit along the edge of the Amazon, Congo Basin, and Borneo rainforests. These are regions that may be encroached by

agroforestry, logging, pasture, and cropland. These results are a further indication that improvements could be made to the F-Model by incorporating land use into the model.

Temporal Accuracy

Figure 3 shows average seasonal patterns in emissions as differences from the mean.

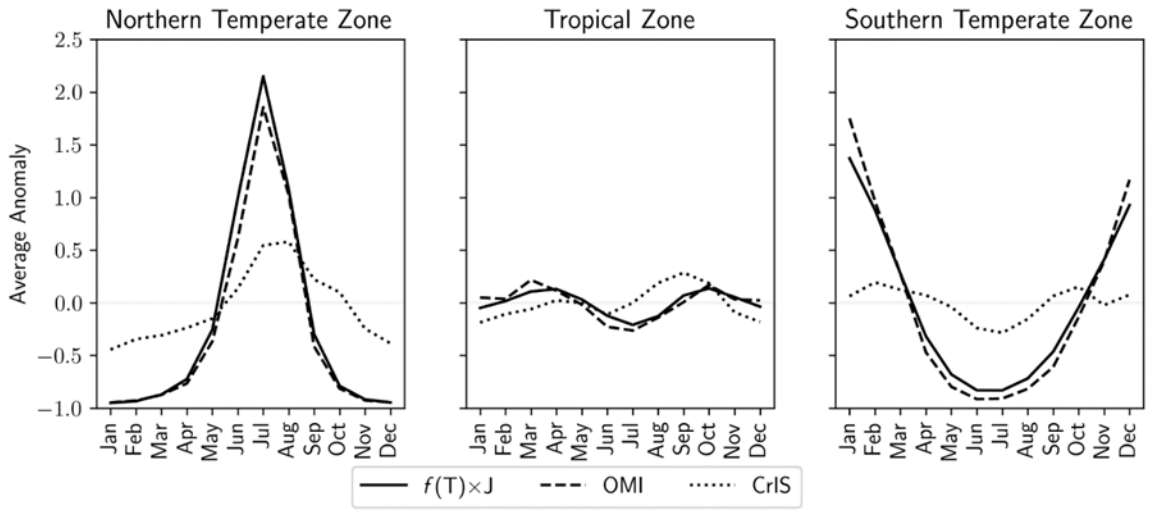
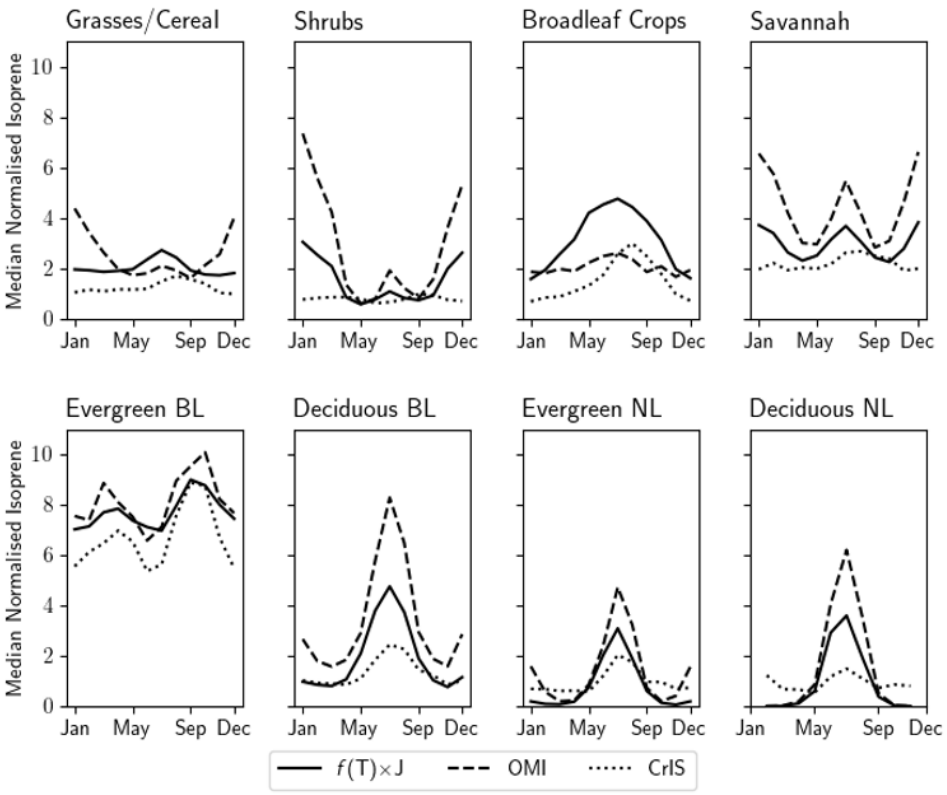


Figure 3*: Regional seasonal monthly isoprene emission anomalies, relative to the mean, from the F-Model, OMI, and CrIS (2012-2014).*

The F-Model reproduces the seasonal fluctuations in the OMI data in all zones. However, the CrIS data diverges from the seasonal behavior of the F-Model and OMI. The CrIS data is less variable than OMI and the F-Model, and in the NH and in the tropical zone, has a phase shift in reaching peak emissions. The reason for this difference remains elusive.

Figure 4 investigates seasonal performance across biomes. The biome filter used in this study is the MODIS Terra Type 3 classification product (Friedl and Sulla-Menashe, 2015).

A.



B.

	Grasses / Cereal	Shrubs	Broadleaf Crops	Savannah	Evergreen BL	Deciduous BL	Evergreen NL	Deciduous NL
$f(T)\times J$ vs OMI	0.692	0.776	0.608	0.748	0.529	0.877	0.693	0.919
$f(T)\times J$ vs CrIS	0.630	0.477	0.484	0.707	0.613	0.763	0.448	0.404

Figure 4*: F-Model performance by biome.* ***A.*** *Average median normalized isoprene.* ***B****. Pearson's r correlation coefficient by biome. All results are statistically significant at all standard levels (p < 0.001).*

In many cases the F-Model reproduces relative emission levels and seasonal patterns across biomes, even though there is no biome-specific parameterization in the model. This is particularly evident in the evergreen BL biome, where the F-Model's replication of the seasonal fluctuations in both CrIS and OMI stands out. There are also cases where the F-Model reproduces trends and relative emission levels seen in OMI but is unable to capture the variation in CrIS. Additionally, CrIS predicts lower relative emissions and less extreme fluctuations than OMI in every biome. This is especially evident in the shrublands and in the deciduous NL biome. While the F-Model indicates lower relative emissions than the OMI data in all cases except in croplands, it does not project emissions as low and stable as CrIS. The F-Model produces a middle ground between the two isoprene products, yet it aligns more closely with OMI than with CrIS as further evidenced in the table in Figure 4B.

Two biomes with poor F-Model performance are shrublands and broadleaf crops. In the shrubland biome, there is low agreement between OMI and CrIS about the magnitude and pattern of emissions and the F-Model aligns more strongly with OMI. The model replicates the patterns in OMI here but is unable to produce the high emissions in the SH summer months. This could be a result of photosynthetic stress being left out of the F-Model in the form of $[J - J_v]$ and water stress parameterization. In the cropland biome on the other hand, emissions are overestimated by the model as it does not parameterize land use: it has no way to account for the fact that broadleaf crops are not isoprene emitters.

The F-Model vs MEGAN

Figure 5 compares the performance of the F-Model to the performance of the empirical MEGAN model. MEGAN 2.1 data is from Guenther et al. (2012) published online by the Irvine Biosphere Atmosphere Interactions Group (Guenther, 2019).

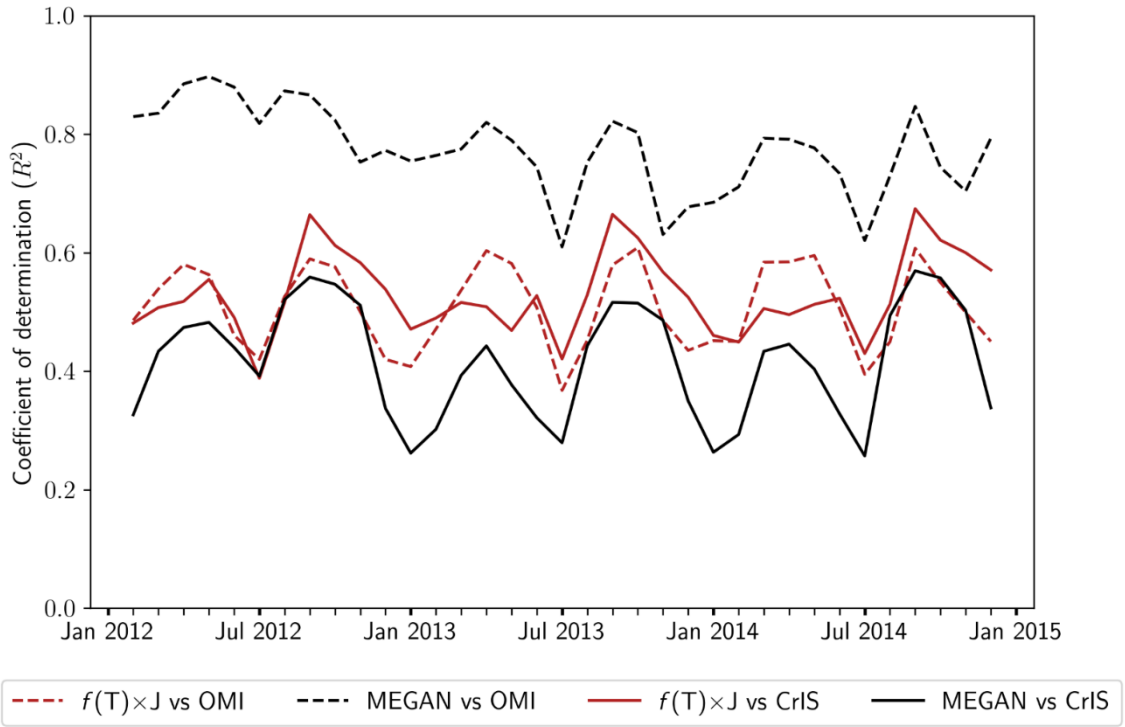


Figure 5: *Monthly R^2 value between F-Model and the OMI and CrIS data compared to their correlation with MEGAN. All results are statistically significant at all standard levels ($p < 0.001$).*

MEGAN's fit depends on which isoprene emissions product it is compared to. When compared to OMI, MEGAN has an R^2 of 0.76 ($p < 0.001$). This high correlation between the MEGAN and OMI is expected as OMI is partially derived using MEGAN. When compared to the fully independent dataset CrIS,

MEGAN is not able to sustain its high explanatory power. MEGAN only explains 41% ($p < 0.001$) of the variation in CrIS and given the highly parameter rich nature of the model, this performance is underwhelming.

On the other hand, the F-Model performs similarly well when compared to both isoprene products. The F-Model explains 50% and 52% ($p < 0.001$) of the variation in OMI and CrIS respectively, as measured by the overall R^2. The result is particularly interesting given the low linear agreement between OMI and CrIS ($R^2 = 0.25$, $p < 0.001$).

Analysis of F

F contains information about the amount of isoprene synthase in a plant, as well as the rate of electron supply to each gram of isoprene synthase. While the F-Model assumes *F* to be constant, the physiological understanding of isoprene production suggests that this value should vary through space and time.

This section investigates the variation of the implied *F* across time, calculated as OMI/[$f(T) \times J$]. *F* is calculated in relation to the OMI dataset as it is the more established isoprene emissions product. The subset of the results for the entire timeseries of OMI are analyzed in this section: from 2005 to 2014 (inclusive).

Patterns in F

The results of the previous section suggest that the model's missing parameterization of land use may be important in some places. Differences in *F* are therefore investigated against this dimension.

Figure 6 shows the variation in implied *F* over space and time.

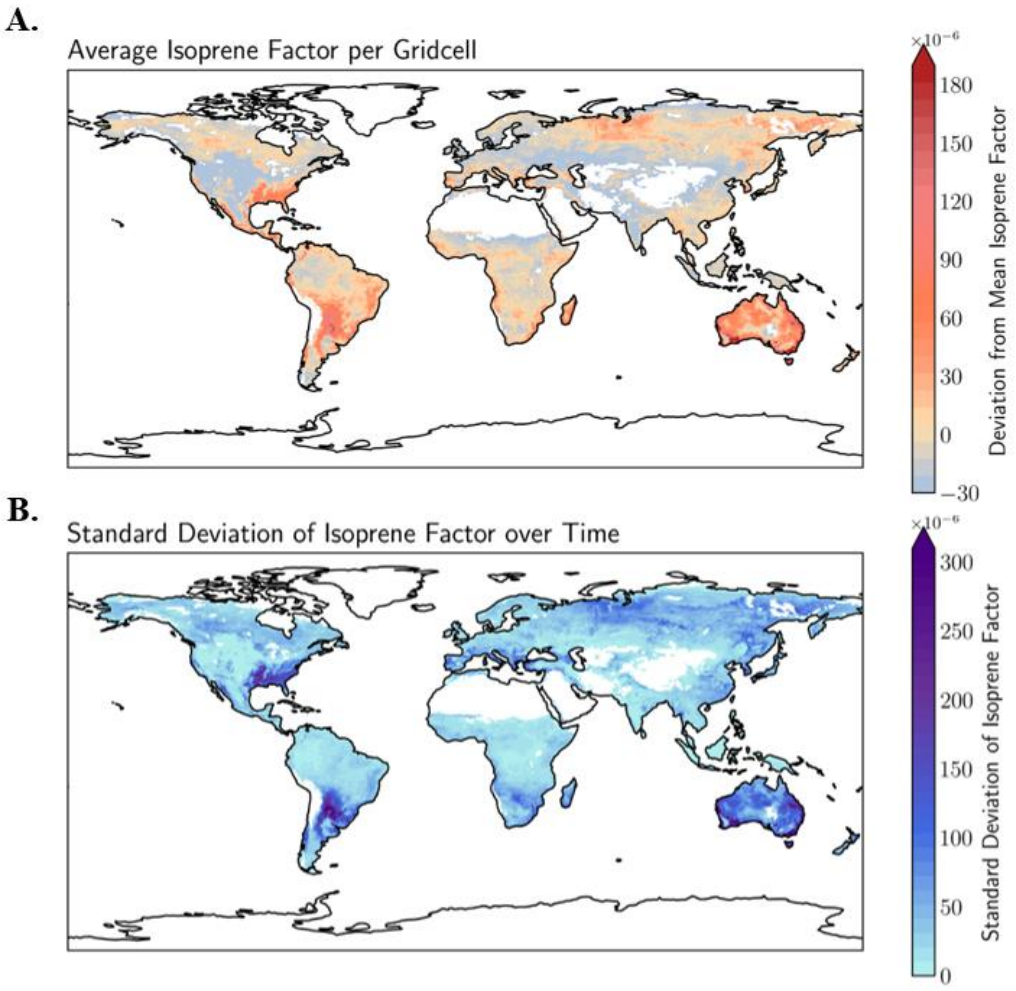


Figure 6: *The variation over space and time in the isoprene factor, F, calculated relative to OMI from 2005 to 2014.* ***A.*** *mean F per grid cell (2005 to 2014) compared to the global average.* ***B.*** *standard deviation of F from its grid cell average.*

In Figure 6A, *F* is close to the global average in most places, indicating that the assumption of a constant *F* through space may be broadly reasonable. However, there are hotspots of higher-than-average *F* values along the Mediterranean, the Russian and North American boreal forest tree line, Australia, the American Coastal Plains, and the Gran Chaco Forest in South America. Bar Australia, the other four areas are classified by MODIS as savannah biomes. In the savannah, the F-Model reproduces seasonal isoprene fluctuations well but is unable to capture the magnitude of emissions (Figure 4A). The variation in implied *F* through time is also particularly high in the Coastal Plains, Australia, and the Gran Chaco Forest (Figure 6B). Perhaps, the effect of photosynthetic stress is particularly potent in the savannah biome and in these regions and hence the higher implied *F* value.

Figure 6A also shows areas of lower-than-average *F* values. These areas are primarily along the Eurasian Steppe, in central and western North America, and in western India. These regions are characterized by grassland biome and exhibit low temporal variability in *F* (Figure 6B). The grassland biome has low isoprene emissions globally and is modelled moderately well by the F-Model (Figure 4). This could be indicative of species in the grasslands producing less isoprene synthase.

The distributions and medians of *F* by biome can be seen in Figure 7 below.

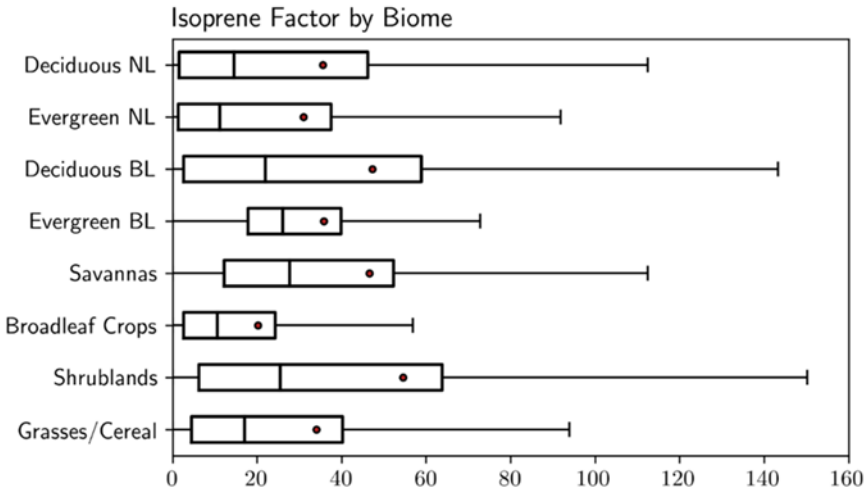


Figure 7: *F from 2005 to 2014 derived by the F-Model and the OMI product bound by biome. The red dots represent the mean. Statistical outliers are not visualized. Medians between all groups are statistically different from each other by the Kruskal-Wallis non-parametric test ($H = 278324$, $p < 0.001$) and Dunn's test with Bonferroni correction for multiple comparisons ($p < 0.001$ in all cases).*

Grasslands indeed seem to have a lower average and less variation in *F* than savannahs. It is possible that this difference in *F* between biomes is a result of differences in environmental conditions in these places or it may reflect true biological differences in the species found in these biomes. While the median *F* for each category is significantly different, there is a high variation within each group, and it is difficult to conclude that the behavior of *F* is primarily defined by biome.

LASSO Regression

To identify which factors are most promising for future work, I use a linear model with ten potential explanatory variables including biome, photosynthetic variables derived from the P-Model, climate variables, and soil moisture stress. I apply LASSO regularization to this regression as LASSO finds a balance between model accuracy and parsimony by applying a penalty term (λ) that drives the regression

coefficient estimates towards zero. This approach aligns with the objective of identifying possible relationships between photosynthetic variables and *F* to guide future mechanistic model development. LASSO regularization is also less sensitive than maximum likelihood to multicollinearity, a critical concern when the explanatory photosynthetic variables are derived from the same climate data by the P-Model. I performed the LASSO penalized regression using the glmnet package in R (Friedman et al., 2023).

Due to the highly right skewed distribution of *F*, I log transform and z-normalize it. While the resulting distribution is leptokurtic, it better approximates normality. The explanatory variables are also z-normalized, and the coefficients are reported in the z-normalized scale.

There is little biological evidence to suggest that the relationships between *F* and the explanatory variables should be linear, nor is it obvious what the shapes of the relationships would be. It is therefore critical that these results only be interpreted as indications of a relationship. The results of the analysis cannot be definitive at this stage.

K-fold cross validation is used to identify appropriate values of λ, with k =10. The regression coefficients are reported for three λ values. The first is at the λ with the minimum cross-validated error ($\log(\lambda) \approx -9.18$) as recommended, the next two are arbitrary ($\log(\lambda) = -5.50$ and -4.00) giving an idea of parameter compression for higher λ and marginally increased cross-validation error. The cross-validation curve of λ can be found in Appendix 6.

	log(λ)		
	-9.18	**-5.50**	**-4.00**
Biome:			
Intercept (Grasses/Cereals)	-0.11	-0.09	-0.06
Decidious NL	0.47	0.43	0.24
Evergreen NL	-0.36	-0.26	-
Decidious BL	0.23	0.19	0.13
Evergreen BL	0.05	-	-
Savannas	0.24	0.09	-
Broadleaf Crops	0.03	-	-
Shrubland	0.10	-	-
Soil Moisture Stress (β)	0.14	0.14	0.12
LUE	-2.04	0.06	0.09
J	-2.04	-	-
Jmax	-	-	-
GPP	1.71	0.16	-
Vcmax	0.00	-0.55	-0.37
Ambient Temperature	0.58	0.52	0.43
Shortwave Radiation	0.29	0.31	0.33
FPAR	0.48	0.50	0.52
R2	0.683	0.678	0.653

Table 1: *Lasso regression results with F as the dependent variable at three λ values. P-values are not reported as they are incompatible with the lasso penalty. The R^2 values reported are virtually identical to the adjusted R^2 values due to the sample size (over 3 million observations). See nomenclature table for unfamiliar acronyms.*

Reducing the number of predictor variables from 17 to 10 resulted in minimal loss in predictive power as measured by R^2 (Table 1) indicating that most of the explanatory power was preserved by the reduced set of variables. Furthermore, the benefit of LASSO regularization over least squares is underscored by the behavior of the coefficients of *J* and GPP in the model. Where least squares would predict high coefficients for these variables (as LASSO does when λ is low), LASSO rapidly shrinks them towards zero as λ increases. This behavior indicates these variables may be overfitting the data or have a correlated variable remaining in the model.

The most robust predicters of *F* are primarily beyond the leaf scale: temperature, shortwave radiation, and FPAR (see Appendix 7 for the coefficient trace plot). These variables are stable in their explanatory power at a range of regularization strengths. They are therefore prime candidates for further investigation, especially given that neither N99 nor the M14 explicitly parametrize these variables in their calculations of ε. None of these of variables are particularly highly correlated with each other (the highest correlation is between shortwave radiation and temperature: $r = 0.59$, $p < 0.001$), and neither temperature nor

shortwave radiation are highly correlated with any other variable tested. However, FPAR is moderately correlated with J_{max}, J, and GPP ($p>0.001$) (correlation matrix in Appendix 8). It is possible that the true relationship is with J rather than FPAR as may be expected from N99 and the M14 models, but this model indicates that there may be a more direct relationship between F and these environmental variables that is yet to be accounted for.

V_{cmax} is also a robust explanatory candidate in this analysis. V_{cmax} has not been explicitly linked to emissions in the literature to this author's knowledge. However, it is mechanistically sensible that V_{cmax} could be a suitable candidate for further research given the importance placed on J_v in the M14: J_v is the rubisco limited electron transport rate while V_{cmax} is the maximum rate of carboxylation by rubisco. In Morfopoulos (2014), V_{cmax} is presented as a modulating factor for J_v, such that the value of J_v is directly proportional to V_{cmax} when all else is constant. The results from the LASSO regression seem to corroborate the M14 model in this way.

The results present soil moisture stress as a consistently informative parameter. While it is not necessarily the most important parameter in the linear model, it is not correlated with any other predictor making its inclusion paramount.

Another notable takeaway from Table 1 is that, as the level of regularization increases, many of the biome dummy variables show reduced importance or are deemed unnecessary by the model. Based on this result, it is worth investigating if PFT based parameterization in isoprene emission models is causing overfitting to observational data and to what extent these parameters are mechanistically necessary to capture the variation in emissions.

Discussion and Conclusion

The F-Model is a parsimonious and simplified representation of isoprene emissions based on the M14 model (Morfopolous et al., 2014). The model reproduces broadscale patterns in emissions through space and time as compared to the OMI and CrIS data products. This is a positive indication for the reliability of the model building blocks of physiological isoprene emissions models like the N99 and M14 model. It is a further positive indication of the benefit of the mechanistic modelling approach that the F-Model seems to outperform MEGAN when compared to an independent dataset CrIS at the global scale.

The parameterization of basal emission rates by PFT in the literature has been cited as problematic. The F-Model does not contain any input data on PFT specific emission rates, and yet reproduces relative emissions and seasonal fluctuation observed in the OMI isoprene product in most cases. While the spatial patterns are well reproduced broadly, the F-Model struggles to replicate emissions in croplands and fragmented landscapes. This might be due to the lack of representation of land covers in the model. Further analysis suggests that while some biome-based parameterization may be a worthwhile addition, other variables appear to be better candidates for investment.

While the ubiquity of PFT based basal emission rates is questioned, differing isoprene emission behaviors between species is not disputed. It is possible for basal factors to be a useful parameterization at an individual and canopy scale but not when modelling emissions globally because of the abstractions made when upscaling the model. For instance, a high emitting species of fern (Azolla *filiculoides)* commonly grows alongside non-emitting rice plantations and increases total emission in the area (Arneth et al., 2008). Satellite based biome classifications are likely to categorize the land as agricultural and researchers may assign the basal emission rate of rice to the whole region, missing out on the true variability. A further abstraction is likely to occur when assigning satellite-based distributions of biomes to PFT groups. In this thesis, a difference in the emission behavior between grasslands and savannahs is observed, but the PFT categories in MEGAN are not narrow enough to account for these differences (Weber et al., 2023). Though basal emission rates undoubtedly change from species to species, the inevitable abstractions made to upscale lab-based observations may make PFT based parameterization less robust when modelling isoprene globally.

Analysis in the isoprene emission factor F, suggests that the influence of ambient temperature, shortwave radiation, and FPAR may still be unaccounted for and may influence the fraction of reducing power going toward isoprene emissions. Neither of the ε values in the M14 nor in the N99 directly account for these variables and so their inclusion may yield some benefits to models.

The F-model has been applied in a limited scope in this project and many unanswered questions are left to future work. An evaluation of the model during high emission events is of particular interest. Modelling emissions during periods of high vegetative stress is critical for improving the ability of vegetation models to constrain plant productivity. Furthermore, it is during such extreme periods where isoprene emission estimates could be substantially improved by a mechanistic approach over a statistical one. Future works would benefit if soil moisture stress were accounted for in the F-model. This requires the P-Model to provide tools to back calculate photosynthetic parameters based on soil moisture stress penalization. Of further interest is to investigate the performance of the F-Model when driven by different vegetation models to isolate the sources of errors observed in this work.

Overall, the performance of the simplified model indicates that the foundation of the Energetic Status Model is aligned with coarse scale temporal and spatial variation in isoprene emissions, and that accounting for vegetation stress and anthropogenic landscape modification may be of great benefit.

Nomenclature

Acronyms

BL	Broadleaf
CrIS	Cross track infrared sounder (and isoprene estimates produced by this sensor)
DMADP	Dimethylallyl pyrophosphate (isoprene precursor)
FPAR	Fraction of absorbed photosynthetically active radiation
G95	Guenther et al. (1995) empirical model of isoprene emissions
GPP	Gross primary productivity
LAI	Leaf area index
LASSO	Least absolute shrinkage and selection operator
LUE	Light use efficiency
M14	The Energetic Status Model for isoprene emissions (Morfopolous et al. 2014)
MEGAN	The model of emissions of gases and aerosols from nature (Guenther et al. 2006)
MEP	Methylerythritol 4-phosphate pathway (for isoprene production)
N99	Niinemets et al. (1999) mechanistic model for isoprene emissions
NH	Northern Hemisphere
NL	Needleleaf
NPP	Net primary productivity
NPQ	Non-photochemical quenching
OMI	Ozone monitoring instrument (and isoprene estimates produced by this sensor)
PFT	Plant functional type
SH	Southern Hemisphere
SIF	Solar-induced chlorophyll fluorescence
Suomi NPP	Suomi national polar-orbiting partnership satellite

Greeks

α	Redox requirement for isoprene production (N99)
β	Stocker's soil moisture stress penalty factor from the SPLASH model
ε	Fraction of electrons towards isoprene production (N99 and M14)
λ	Penalty term in LASSO regression

Romans

A_J	Light limited assimilation rate
A_V	Rubisco limited assimilation rate
c_1	Empirical constant 1 (M14)
c_2	Empirical constant 2 (M14)
F	Empirical constant (F-Model)
I	Isoprene emission rate
$f(T)$	Isoprene synthase activity algorithm (see Appendix 1)
J	Light limited electron flux
J_{max}	Maximum rate of light limited electron flux
J_v	Rubisco regeneration limited electron flux
V_{cmax}	Maximum rate of carboxylation by rubisco

Data and Code Availability Statement

The data generated as part of this project can be found at: *https://doi.org/10.5281/zenodo.8265993*

The sources of the publicly available datasets are given in Appendix 2.

The code used in this project can be found at: *https://github.com/bikempastine/Isoprene_PModel*

Code for the P-Model can be found at: *https://github.com/ImperialCollegeLondon/pyrealm*

References

Acharya, P. and Punia, M. (2013) 'Comparison of modis derived land use and land cover with Ministry of Agriculture reported statistics for India', Journal of Applied Remote Sensing, 7(1), p. 073524. doi:10.1117/1.jrs.7.073524.

Ardö, J. (2015) 'Comparison between remote sensing and a dynamic vegetation model for estimating terrestrial primary production of Africa', Carbon Balance and Management, 10(1). doi:10.1186/s13021-015-0018-5.

Arneth, A. et al. (2007) 'Process-based estimates of terrestrial ecosystem isoprene emissions: Incorporating the effects of a direct CO2-isoprene interaction', Atmospheric Chemistry and Physics, 7(1), pp. 31–53. doi:10.5194/acp-7-31-2007.

Arneth, A. et al. (2008) 'Why are estimates of global terrestrial isoprene emissions so similar (and why is this not so for monoterpenes)?', Atmospheric Chemistry and Physics, 8(16), pp. 4605–4620. doi:10.5194/acp-8-4605-2008.

Bauwens, M. et al. (2016) Nine Years of global hydrocarbon emissions based on source inversion of omi formaldehyde observations [Preprint]. doi:10.5194/acp-2016-221.

Cao, Y. et al. (2021) 'Ensemble projection of global isoprene emissions by the end of 21st century using CMIP6 models', Atmospheric Environment, 267, p. 118766. doi:10.1016/j.atmosenv.2021.118766.

Cucchi M., Weedon G. P., Amici A., Bellouin N., Lange S., Müller Schmied H., Hersbach H., Cagnazzo, C. and Buontempo C. (2021): Near surface meteorological variables from 1979 to 2019 derived from bias-corrected reanalysis, version 2.0. Copernicus Climate Change Service (C3S) Climate Data Store (CDS), DOI: 10.24381/cds.20d54e34

Davis, T.W. et al. (2017) 'Simple process-led algorithms for simulating habitats (Splash v.1.0): Robust indices of radiation, evapotranspiration and plant-available moisture', Geoscientific Model Development, 10(2), pp. 689–708. doi:10.5194/gmd-10-689-2017.

Environmental Protection Agency, U.S. (2023) Ecoregions of North America, EPA. Available at: https://www.epa.gov/eco-research/ecoregions-north-america.

Eva, H.D. et al. (2004) 'A land cover map of South America', Global Change Biology, 10(5), pp. 731–744. doi:10.1111/j.1529-8817.2003.00774.x.

Farquhar, G.D., von Caemmerer, S. and Berry, J.A. (1980) 'A biochemical model of photosynthetic CO2 assimilation in leaves of c3 species', Planta, 149(1), pp. 78–90. doi:10.1007/bf00386231.

Filella, I. et al. (2018) 'A Modis photochemical reflectance index (PRI) as an estimator of isoprene emissions in a temperate deciduous forest', Remote Sensing, 10(4), p. 557. doi:10.3390/rs10040557.

Friedl, M., Sulla-Menashe, D. (2015). 'MCD12C1 MODIS/Terra+Aqua Land Cover Type Yearly L3 Global 0.05Deg CMG V006' [Data set]. NASA EOSDIS Land Processes Distributed Active Archive Center. https://doi.org/10.5067/MODIS/MCD12C1.006

Friedman J, Tibshirani R, Hastie T (2010). "Regularization Paths for Generalized Linear Models via Coordinate Descent." Journal of Statistical Software, **33**(1), 1–22. doi:10.18637/jss.v033.i01.

Fu, D. et al. (2018) ‘Retrievals of tropospheric ozone profiles from the synergism of airs and omi: Methodology and validation’, Atmospheric Measurement Techniques, 11(10), pp. 5587–5605. doi:10.5194/amt-11-5587-2018.

Fu, D. et al. (2019) ‘Direct retrieval of isoprene from satellite-based infrared measurements’, Nature Communications, 10(1). doi:10.1038/s41467-019-11835-0.

Gomez, J. et al. (2023) ‘The projected future degradation in air quality is caused by more abundant natural aerosols in a warmer world’, Communications Earth & Environment, 4(1). doi:10.1038/s43247-023-00688-7.

Guenther, A. (2019) ‘MEGAN21’ [Data set], University of California at Irvine BAI. Available at: https://bai.ess.uci.edu/megan/data-and-code/megan21.

Guenther, A. et al. (1995) ‘A global model of natural volatile organic compound emissions’, Journal of Geophysical Research, 100(D5), p. 8873. doi:10.1029/94jd02950.

Guenther, A. et al. (2006) ‘Estimates of global terrestrial isoprene emissions using Megan (model of emissions of gases and aerosols from nature)’, Atmospheric Chemistry and Physics, 6(11), pp. 3181–3210. doi:10.5194/acp-6-3181-2006.

Guenther, A.B. et al. (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. doi:10.5194/gmd-5-1471-2012.

Harris, I. et al. (2020) ‘Version 4 of the Cru Ts monthly high-resolution gridded Multivariate Climate Dataset’, Scientific Data, 7(1). doi:10.1038/s41597-020-0453-3.

Harrison, S.P. et al. (2012) ‘Volatile isoprenoid emissions from plastid to planet’, New Phytologist, 197(1), pp. 49–57. doi:10.1111/nph.12021.

Jiang, X. et al. (2018) ‘Isoprene emission response to drought and the impact on Global Atmospheric Chemistry’, Atmospheric Environment, 183, pp. 69–83. doi:10.1016/j.atmosenv.2018.01.026.

Langford, B. et al. (2017) ‘Isoprene emission potentials from European oak forests derived from canopy flux measurements: An assessment of uncertainties and inter-algorithm variability’, Biogeosciences, 14(23), pp. 5571–5594. doi:10.5194/bg-14-5571-2017.

Lantz, A.T. et al. (2019) ‘Isoprene: New insights into the control of emission and mediation of stress tolerance by gene expression’, Plant, Cell & Environment, 42(10), pp. 2808–2826. doi:10.1111/pce.13629.

McGenity, T.J. et al. (2018) ‘Microbial cycling of isoprene, the most abundantly produced biological volatile organic compound on Earth’, *The ISME Journal*, 12(4), pp. 931–941. doi:10.1038/s41396-018-0072-6.

Morfopoulos, C. (2014) A unifying model for isoprene emission by plants. Thesis. Imperial College London.

Morfopoulos, C. et al. (2014) ‘A model of plant isoprene emission based on available reducing power captures responses to atmospheric CO2’, New Phytologist, 203(1), pp. 125–139. doi:10.1111/nph.12770.Stoc

Morfopoulos, C. et al. (2021) ‘Vegetation responses to climate extremes recorded by remotely sensed atmospheric formaldehyde’, Global Change Biology, 28(5), pp. 1809–1822. doi:10.1111/gcb.15880.

Müller, J.-F. et al. (2008) ‘Global isoprene emissions estimated using Megan, ECMWF analyses and a detailed canopy environment model’, Atmospheric Chemistry and Physics, 8(5), pp. 1329–1341. doi:10.5194/acp-8-1329-2008.

Myneni, R., et al. (2015). MYD15A2H MODIS/Aqua Leaf Area Index/FPAR 8-Day L4 Global 500m SIN Grid V006 [Data set]. NASA EOSDIS Land Processes Distributed Active Archive Center. https://doi.org/10.5067/MODIS/MYD15A2H.006

Niinemets et al. (1999) ‘A model of isoprene emission based on energetic requirements for isoprene synthesis and leaf photosynthetic properties for liquidambar and quercus’, Plant, Cell & Environment, 22(11), pp. 1319–1335. doi:10.1046/j.1365-3040.1999.00505.x.

Pacifico, F. et al. (2009) ‘Isoprene emissions and climate’, Atmospheric Environment, 43(39), pp. 6121–6135. doi:10.1016/j.atmosenv.2009.09.002.

Pacifico, F. et al. (2011) ‘Evaluation of a photosynthesis-based biogenic isoprene emission scheme in Jules and simulation of isoprene emissions under present-day climate conditions’, Atmospheric Chemistry and Physics, 11(9), pp. 4371–4389. doi:10.5194/acp-11-4371-2011.

Peñuelas, J. et al. (2013) ‘Photochemical reflectance index as an indirect estimator of foliar isoprenoid emissions at the ecosystem level’, Nature Communications, 4(1). doi:10.1038/ncomms3604.

Pollastri, S., Tsonev, T. and Loreto, F. (2014) ‘Isoprene improves photochemical efficiency and enhances heat dissipation in plants at physiological temperatures’, Journal of Experimental Botany, 65(6), pp. 1565–1570. doi:10.1093/jxb/eru033.

Prentice, I.C. et al. (2013) ‘Balancing the costs of carbon gain and water transport: Testing a new theoretical framework for plant functional ecology’, Ecology Letters, 17(1), pp. 82–91. doi:10.1111/ele.12211.

Randerson, J.T. et al. (2018) Global Fire Emissions Database, Version 4, (GFEDv4). ORNL DAAC, Oak Ridge, Tennessee, USA doi:10.3334/ORNLDAAC/1293

Sharkey, T.D, et al. (2007) ‘Isoprene emission from plants: Why and how’, Annals of Botany, 101(1), pp. 5–18. doi:10.1093/aob/mcm240.

Sharkey, T.D. and Monson, R.K. (2017) ‘Isoprene research - 60 years later, the biology is still enigmatic’, Plant, Cell & Environment, 40(9), pp. 1671–1678. doi:10.1111/pce.12930.

Situ, S. et al. (2014) ‘Uncertainties of isoprene emissions in the Megan model estimated for a coniferous and broad-leaved mixed forest in southern China’, Atmospheric Environment, 98, pp. 105–110. doi:10.1016/j.atmosenv.2014.08.023.

Stavrakou, T. et al. (2013) ‘Key Chemical Nox sink uncertainties and how they influence top-down emissions of Nitrogen Oxides’, Atmospheric Chemistry and Physics, 13(17), pp. 9057–9082. doi:10.5194/acp-13-9057-2013.

Stocker, B.D. et al. (2020) ‘P-model V1.0: An optimality-based light use efficiency model for simulating ecosystem gross primary production’, Geoscientific Model Development, 13(3), pp. 1545–1581. doi:10.5194/gmd-13-1545-2020.

Tans, P. and Keeling, R. (2005) Atmospheric carbon dioxide variations at Mauna Loa Observatory, Hawaii, Global Monitoring Laboratory. Available at: https://gml.noaa.gov/ccgg/trends/data.html (Accessed: 21 May 2023).

von Caemmerer, S. and Farquhar, G.D. (1981) 'Some relationships between the biochemistry of photosynthesis and the gas exchange of leaves', Planta, 153(4), pp. 376–387. doi:10.1007/bf00384257.

Wang, H. et al. (2017) 'Towards a universal model for carbon dioxide uptake by plants', Nature Plants, 3(9), pp. 734–741. doi:10.1038/s41477-017-0006-8.

Weber, J. et al. (2023) 'Updated isoprene and terpene emission factors for the interactive BVOC (ibvoc) emission scheme in the United Kingdom Earth System Model (UKESM1.0)', Geoscientific Model Development, 16(10), pp. 3083–3101. doi:10.5194/gmd-16-3083-2023.

Wells, K.C. and Millet, D.B. (2022) ROCR isoprene retrievals from the Cris satellite sensor, ROCR Isoprene Retrievals from the CrIS Satellite Sensor doi:10.13020/5n0j-wx73

Wells, K.C. et al. (2020) 'Satellite isoprene retrievals constrain emissions and atmospheric oxidation', Nature, 585(7824), pp. 225–233. doi:10.1038/s41586-020-2664-3.

Zuo, Z. et al. (2019) 'Isoprene acts as a signaling molecule in gene networks important for stress responses and Plant Growth', Plant Physiology, 180(1), pp. 124–152. doi:10.1104/pp.18.01391.

Appendices

Appendix 1: Isoprene synthase temperature dependency algorithm f(T) as in Niinemets et al. (1999)

$$f(T) = \frac{S_S(T)}{S_S(30°C)}$$

$$S_S(T) = \frac{e^{\frac{c - \Delta H_s^a}{RT}}}{1 + e^{\frac{\Delta S^S T - \Delta H_s^d}{RT}}}$$

Here, c is a unitless scaling constant with the value 35.478, ΔH^{as} is the activation energy with the value of 83129 J/mol, R is the gas constant with the value 8.314 J/mol/K, T is temperature in Kelvin, ΔS^S is the entropy term with the value 8875 J/K/mol, and ΔH^{ds} is the deactivation energy with the value of 284,600 J/mol.

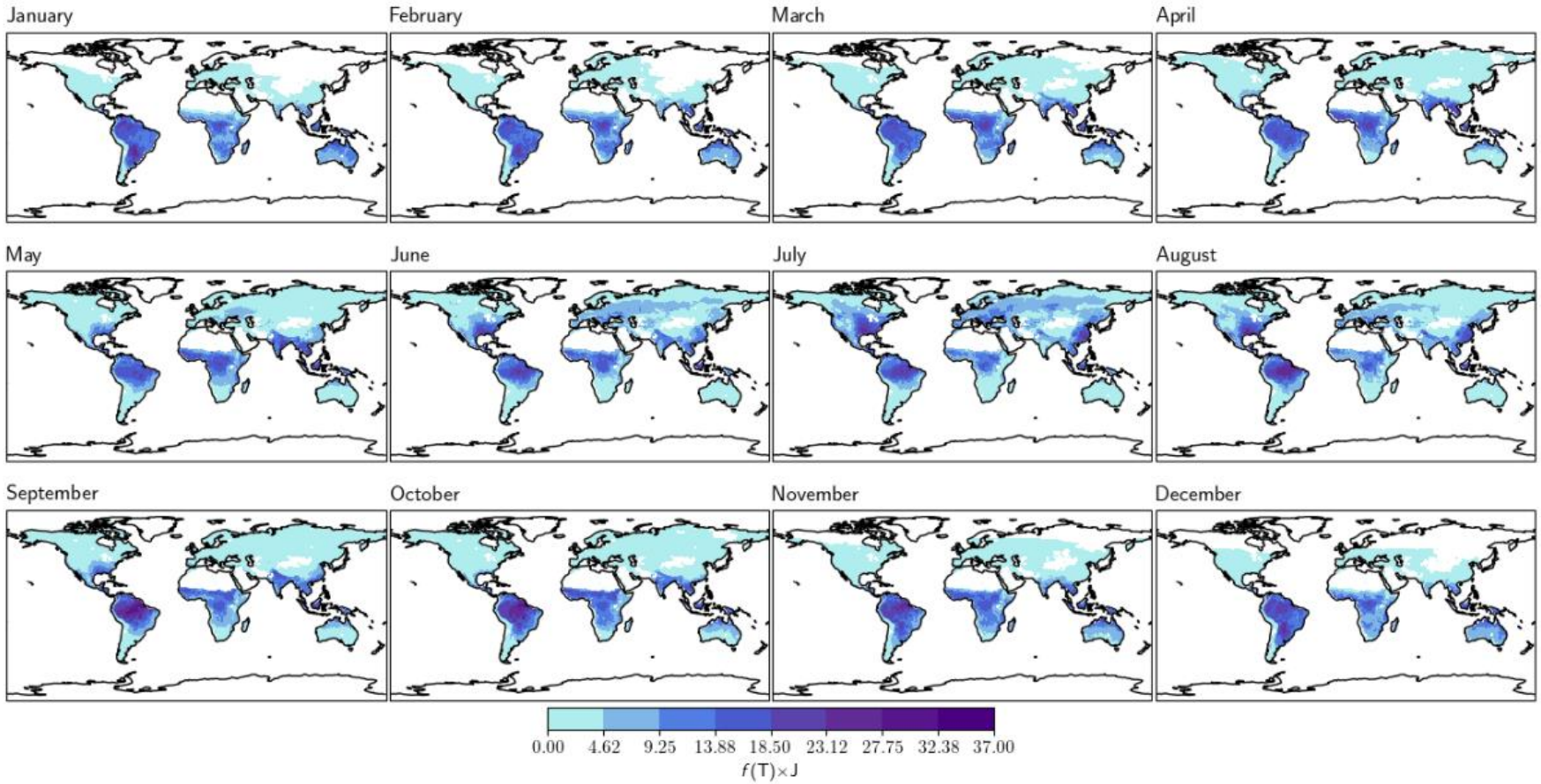
Appendix 2: Table of data sources

	Source	Spatial Resolution	Temporal Resolution	Download point	Citation
P-Model					
Air Temperature	WFDE5: bias-adjusted ERA5 reanalysis data	0.5° × 0.5°	Hourly	Copernicus	Cucci et al., (2021)
Shortwave Radiation	WFDE5: bias-adjusted ERA5 reanalysis data	0.5° × 0.5°	Hourly	Copernicus	Cucci et al., (2021)
Vapor Pressure Deficit	CRU TS4.04: high-resolution gridded data of month-by-month variation in climate	0.5° × 0.5°	Daily	CEDA Archive	Harris et al., (2020)
FPAR	MODIS/Aqua Leaf Area Index/FPAR 8-Day L4 Global 500 m SIN Grid	500 meters	8 Day	Provided by Dr. Morfopolous	Myneni et al., (2015)
Elevation	WFDE5: bias-adjusted ERA5 reanalysis data	0.5° × 0.5°	-	Copernicus	Cucci et al., (2021)
CO2	Mauna Loa Observatory, Hawaii, Global Monitoring Laboratory	Single point *	Monthly	Global Monitoring Laboratory	Tans and Keeling, (2005)
Analysis					
Soil Moisture Stress	Provided by Dr. Sandova	0.5° × 0.5°	Monthly	Contact email: d.sandoval17@ic.ac.uk	Davis et al., (2020)
Biome	MODIS Terra Type 3 classification product	0.5° × 0.5°	-	United States Geological Survey	Friedl and Sulla-Menashe, (2015)
CrIS	Global isoprene measurements from CrIS	0.5° × 0.5°	Monthly	Data Repository for U of M	Wells and Millet, (2020)
OMI	Satellite-derived isoprene emission estimates based on OMI HCHO (2005-2014)	0.5° × 0.5°	Monthly	Belgian Institute for Space Aeronomy	Bauwens et al., (2015)

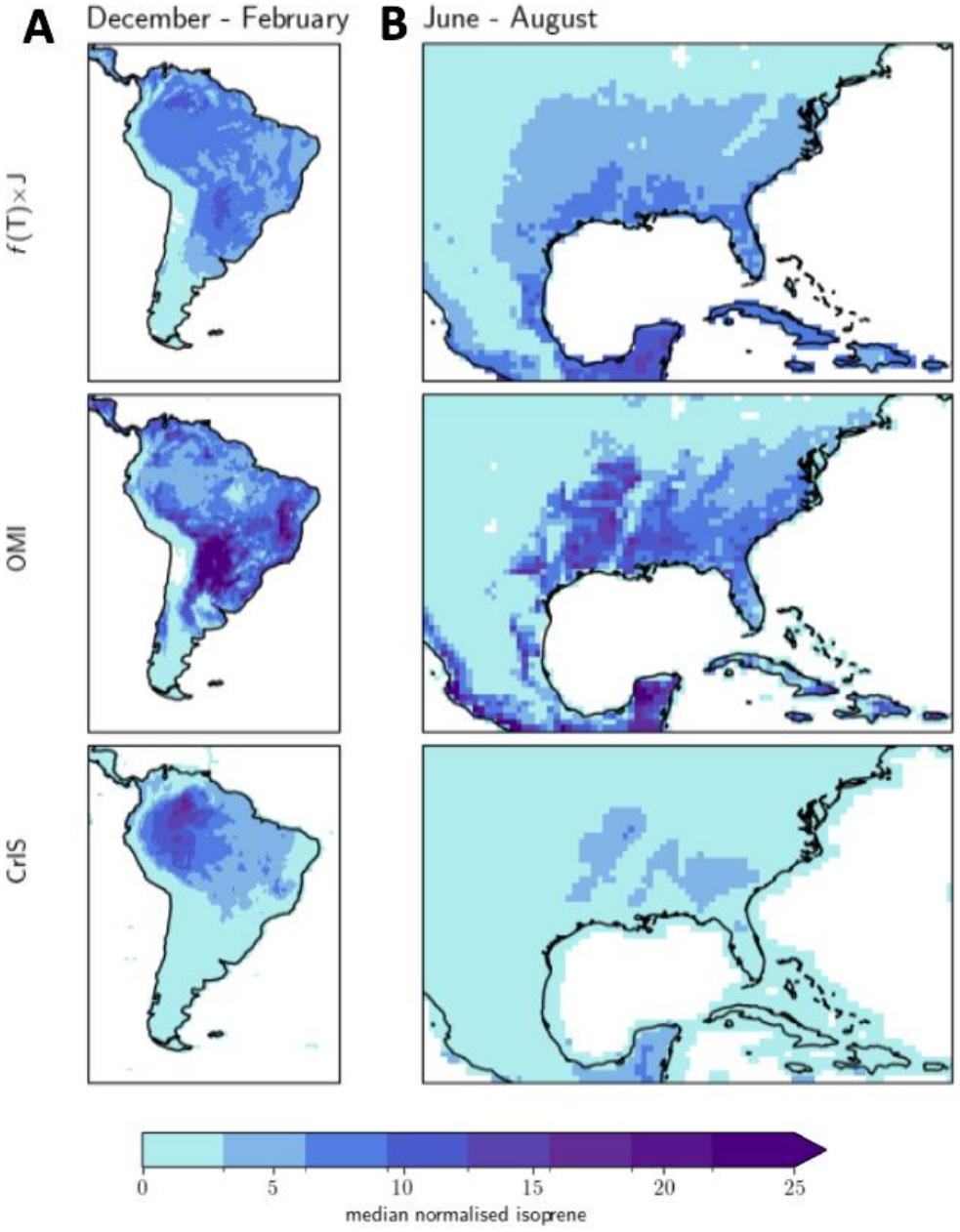
* CO2 levels are assumed spatially constant globally.

Appendix 3: Map of the monthly average F-Model isoprene estimates 2005 to 2014.

Model results are in µmol isoprene/m2/month.

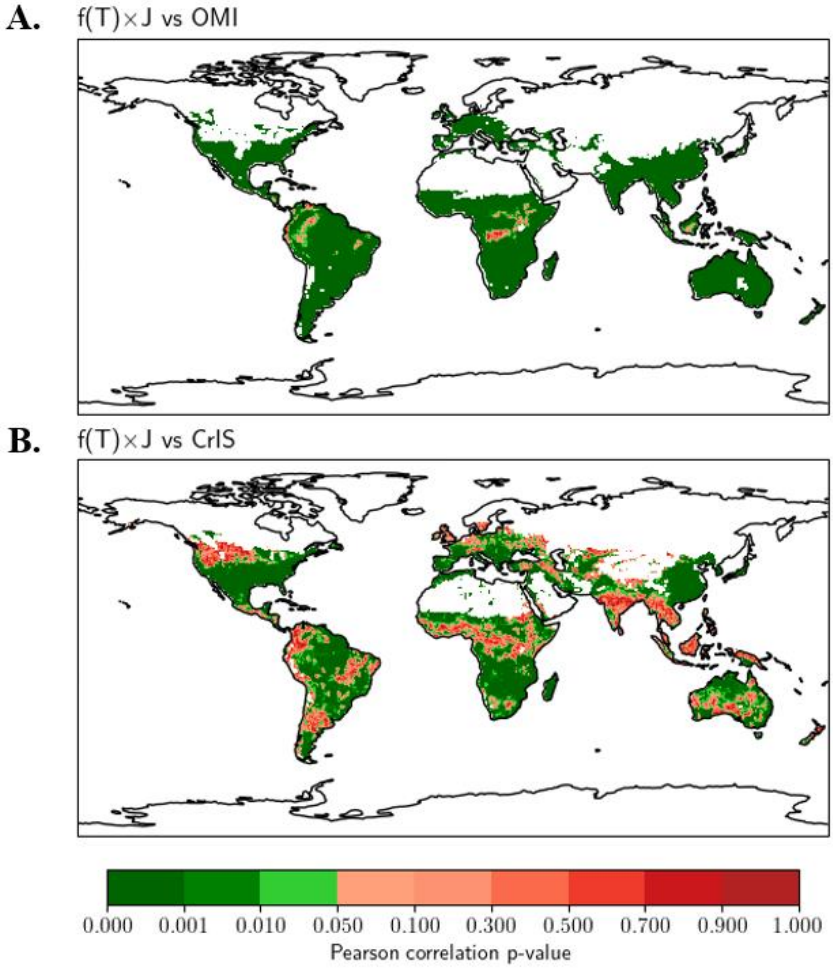


Appendix 4: Map showing the normalized average emission in two locations discussed in the text, blown up.



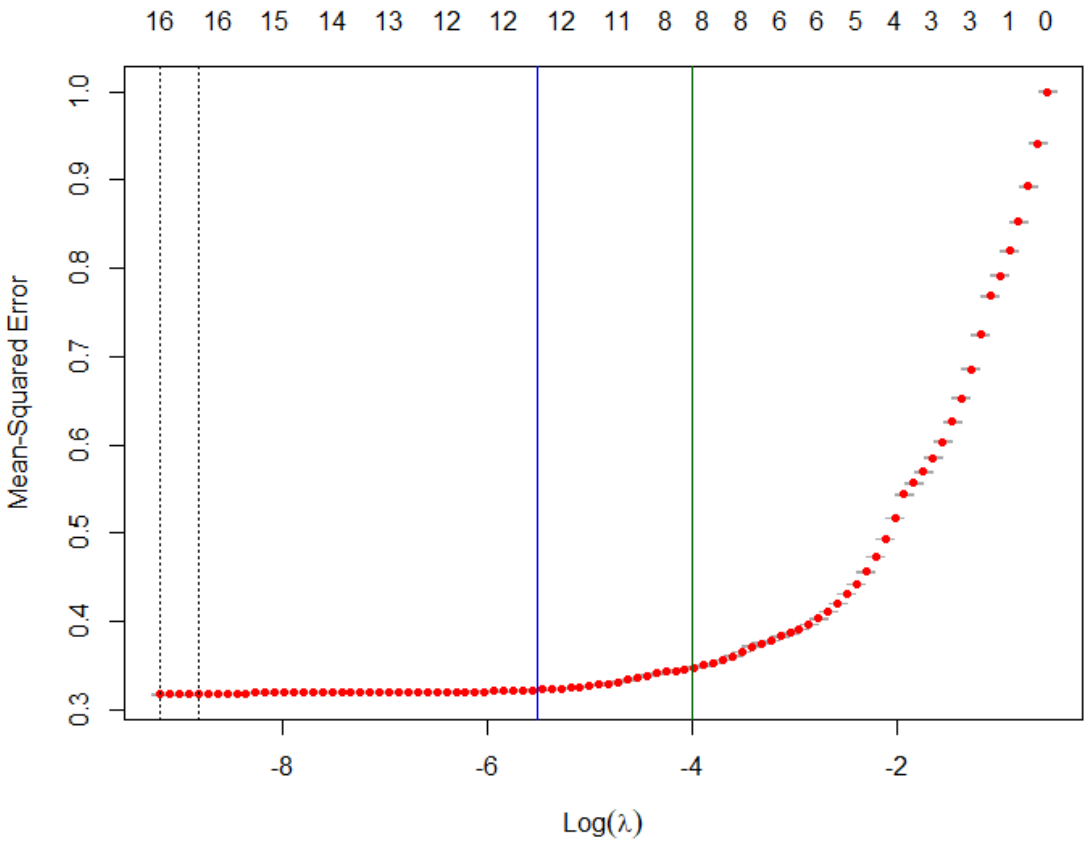
Appendix 5: Map showing Pearson's correlation p-value with OMI and CrIS Satellite Products calculated on each grid cell (2012-2014).

Map corresponding to Figure 3 in the main text.

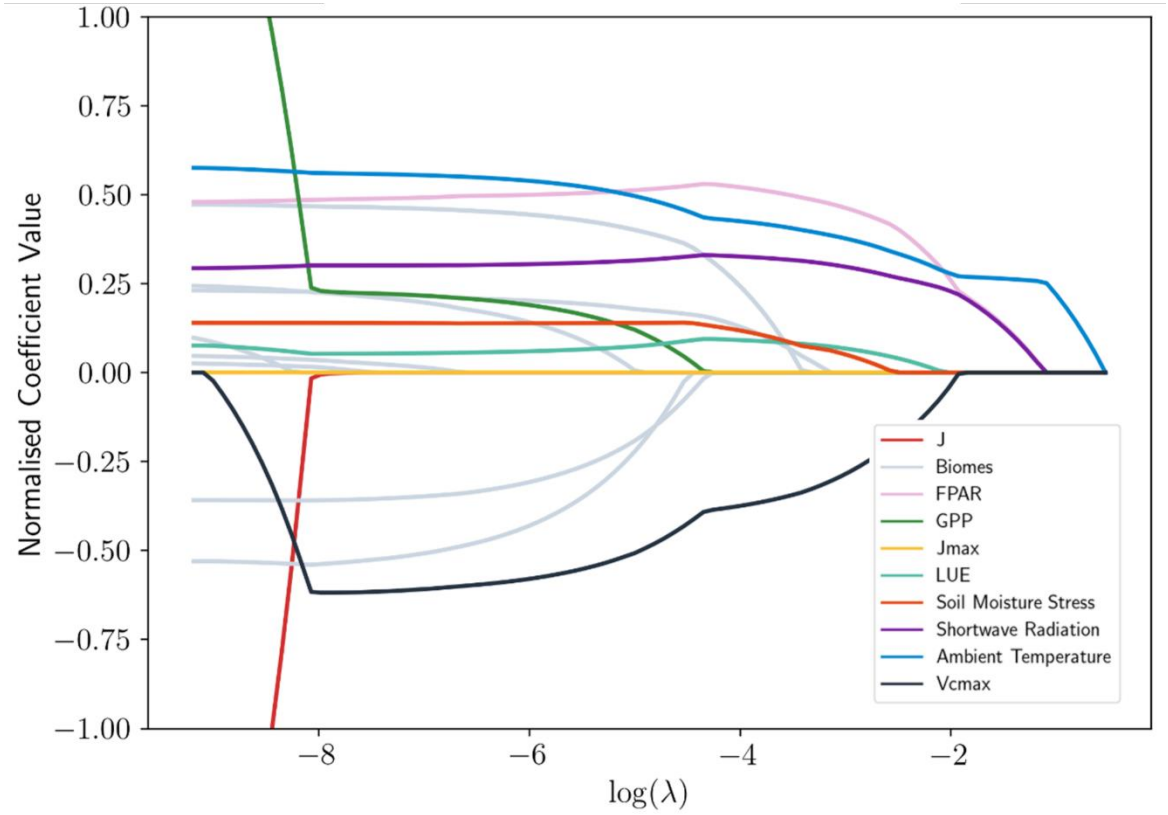


Appendix 6: Lasso Regression Cross Validation curve of λ

The dotted lines illustrate λ with the lowest mean square error across the k-folds, and one standard error above this minimum. The blue line represents the second λ reported in Figure 8 (log(λ) = -5.5) and the green line is the third and highest λ reported (log(λ) = -4).



Appendix 7: Lasso Regression Coefficient Trace Plot



Appendix 8: Pearson's correlation coefficient, r, heat map between explanatory variables used in Lasso Regression

All significant at p<0.001.