Development of a Reduced Complexity Plant Canopy Physics Surrogate Model for use in Chemical Transport Models: A Case Study with GEOS-Chem v12.3.0

Sam J. Silva[1,2], Colette L. Heald[1], Alex B. Guenther[3]

5   [1]Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA
[2]Now at: Pacific Northwest National Laboratory, Richland, WA, USA
[3]Department of Earth System Science, University of California Irvine, Irvine, CA, USA

Corresponding Authors: Sam J. Silva (sam.silva@pnnl.gov) and Colette L. Heald
10   (heald@mit.edu)

Abstract. Biosphere-atmosphere interactions strongly influence the chemical composition of the atmosphere. Simulating these interactions at a detailed process-based level has traditionally been computationally intensive and resource prohibitive, commonly due to complexities in calculating radiation and light at the

15   leaf level within plant canopies. Here we describe a surrogate canopy physics model based on the MEGAN3 detailed canopy model parameterized using a statistical learning technique. This surrogate canopy model is designed specifically to rapidly calculate leaf-level temperature and photosynthetically active radiative (PAR) for use in large-scale chemical transport models (CTMs). Our surrogate model can

20   reproduce the dominant spatiotemporal variability of the more detailed MEGAN3 canopy model to within 10% across the globe. Implementation of this surrogate model into the GEOS-Chem CTM leads to small local changes in ozone dry deposition velocities of less than 5%, and larger local changes in isoprene emissions of up to ~40%, though annual global isoprene emissions remain largely consistent (within

25   5%). These changes to surface-atmosphere exchange lead to small changes in

surface ozone concentrations of ±1 ppbv, modestly reducing the northern hemispheric ozone bias, in which is common to many CTMs, here from 8 to 7 ppbv. The use of this computationally efficient surrogate canopy model drives emissions of isoprene and concentrations of surface ozone closer to observationally constrained values. Additionally, this surrogate model allows for the further development and implementation of leaf-level emission factors in the calculation of biogenic emissions in the GEOS-Chem CTM. Though not the focus of this work, this ultimately enables a complete implementation of the MEGAN3 emissions framework within GEOS-Chem, which produces 570 Tg yr^{-1} of isoprene in 2012.

1. Introduction

The biosphere plays an important role in modulating the abundance and variability of trace gases and aerosol in the atmosphere. Direct emissions of gas phase species are drivers of the majority of the natural reactivity in the atmosphere, and are important precursor sources to pollutants and climate relevant species like ozone and particulate matter (Guenther et al., 2012; IPCC, 2013; Safieddine et al., 2017). On the other hand, vegetation serves as a direct sink for these same species through a process known as dry deposition (Lelieveld and Dentener, 2000; Silva and Heald, 2018). The physical structure of the vegetation can also influence the production and loss of atmospheric constituents through changes to atmospheric turbulent transport and reductions in the actinic flux below the canopy (e.g. Makar et al., 2017). Additionally, chemical reactions occurring within the plant canopy act as a source and sink for reactive species in the above-canopy atmosphere (Goldstein et

al., 2004; Makar et al., 1999). Ultimately, the balance between the role vegetation plays as a chemical source and sink is a controlling factor for the abundance and variability of trace gases and aerosol across many regions of the globe (e.g. Geddes et al., 2016; Silva et al., 2016; Unger, 2014). It is thus important to properly account for these processes when simulating the composition and chemistry of the atmosphere.

Explicitly simulating biosphere-atmosphere interactions necessitates a detailed representation of physical, chemical, and biological processes that occur at the scale of an individual plant. This is typically achieved by integrating a set of energy and radiative balance equations vertically throughout a canopy (e.g. Ashworth et al., 2015, 2016; Goudriaan and Laar, 1994). This sort of physical model of the canopy calculates the environmental parameters that drive the biological and chemical processes which ultimately impact the atmospheric fluxes of trace gases and aerosol (Guenther et al., 2012; Lamb et al., 1996). These canopy models tend to be computationally quite expensive, and are based on measurements taken at very fine resolution (e.g. meter or less), while most atmospheric chemical transport models operate on the 10-200 km scale. Reconciling these differences in scale and addressing the steep computational requirements inherent in both canopy models and atmospheric chemical transport models are critical challenges in simulating chemically relevant interactions between the biosphere and the atmosphere.

Given the computational costs, atmospheric chemical transport models approximate canopy physics and the resulting effects on biosphere-atmosphere interactions

70 through various parameterizations. (e.g. Guenther et al., 2006; Wesely, 1989; Zhang

 et al., 2003). Most of these parameterizations are based on observed relationships,

 and intended to reduce the computational load around the calculation of the

 temperature of leaves and the amount of light (specifically photosynthetically active

 radiation, PAR) reaching leaves throughout the canopy. These model

75 parameterizations commonly assume that the temperature of leaves is equal to the

 air temperature just above the plant canopy (e.g. Guenther et al., 2006; Millet et al.,

 2010) or are based on parameterizations that ignore leaf temperature entirely (e.g.

 Wesely, 1989). The parameterizations for leaf level PAR vary widely; from assuming

 that the PAR reaching leaves in the canopy is equal to the flux of PAR incident on the

80 top of the canopy, to having some sort of reduced complexity multiplicative factor

 that represents the bulk canopy effects (e.g. shading of leaves, Guenther et al., 2006;

 Wang et al., 1998). To our knowledge, the overall impact of these parameterized

 assumptions on the fidelity of modern chemical transport models has not been

 comprehensively characterized. However for biogenic isoprene emissions, these

85 canopy approximations can lead to regional differences of greater than 20% relative

 to a fully detailed canopy model (Guenther et al., 2006).

 Direct representation of these processes is a necessary step to improve model

 reliability and validity, particularly in a rapidly changing Earth System (Committee

 on the Future of Atmospheric Chemistry Research et al., 2016). Currently, many

90 processes related to canopy energy and radiative balance are not represented in

 models of atmospheric chemistry due to computational constraints. In this work, we

 present a reduced complexity canopy model to calculate leaf temperature and PAR

for use in large-scale chemical transport models. This reduced complexity model

removes the need for approximating bulk effects of plant canopies on leaf-level PAR

95   and leaf temperature, and it allows for a more explicit process-based representation

of these effects on biosphere-atmosphere interactions at the leaf level. Our reduced

model reproduces the output of the more detailed vegetation model well, without

the large computational overhead.

2. MEGAN3 Canopy Model

100   We develop and implement a computationally efficient surrogate of the MEGAN3.0

canopy model (https://bai.ess.uci.edu/megan, last accessed 04/09/2019), an

update from previous versions of MEGAN (Guenther et al., 2006, 2012). This canopy

model calculates leaf temperature and leaf-level PAR for a 5-layer plant canopy for

both sunlit and shaded leaves, where each canopy layer represents a fraction of the

105   total plant canopy. The model is originally based largely on Goudriaan and Laar

(1994) and a brief description follows; for more information see Guenther et al.

(1999, 2006, 2012).

In the MEGAN3 canopy model the fraction of sunlit leaves in the canopy decreases

exponentially as a function of the local solar elevation angle, canopy leaf area index

110   (LAI), a clustering coefficient that accounts for leaf geometries, and a canopy

transparency coefficient representing the fraction of the canopy that does not

intercept incident radiation. The leaf temperature is calculated from a system of

energy balance equations based on Goudriaan and Laar, (1994) and Leuning et al.

(1995), with a maximum absolute difference between air temperature and leaf

115 temperature of 10℃. Leaf-level PAR is computed as a function of incoming radiation incident to the canopy top, the sunlit fraction of leaves, LAI, and a suite of geometric and radiative look up table characteristics, predominantly based on Goudriaan and Laar (1994), Leuning et al. (1995), and Spitters (1986). The full MEGAN3 canopy model takes as input: time (day and hour), geographical location (latitude and

120 longitude), air temperature, incident radiation on the top of the canopy, wind speed, humidity, air pressure, LAI, and a set of canopy characteristics (canopy biomass distribution, clustering coefficients, etc.) in the form of a look up table that varies by six vegetation types. The six vegetation types are: needleleaf trees, tropical forest trees, temperate broadleaf trees, shrubs, herbaceous plants, and crops. It is

125 important to note that differing canopy model choice and parameter selection can result in substantial changes to the ultimate estimates of biosphere-atmosphere exchange (Keenan et al., 2011).

The MEGAN3 canopy model has been specifically developed for use in simulating biogenic emissions, and has been extensively applied in related studies (e.g. Chen W.

130 H. et al., 2018; Geron et al., 2016; Guenther et al., 2006). Additionally, the MEGAN framework has been widely adopted across a variety of regional and global models (e.g. GEOS-Chem, WRF-CHEM, and CESM). Thus the MEGAN3 canopy model is a good candidate for surrogate model development because it enables a direct implementation of improved process-based canopy physics into a variety of 3D

135 models without the need for substantial model architecture development.

3. Surrogate Model Development

To begin the surrogate model development, we first use a variable selection

approach to evaluate and rank which of the suite of model input variables are most

important for the simulation of both leaf level PAR and temperature. To do this, we

140   use a machine learning regression method for model simplification and

parameterization, specifically LASSO (Least Absolute Shrinkage and Selection

Operator, Hastie et al., 2001). As applied here, LASSO is a regression method that

calculates linear coefficients through a modified least squares cost function, with the

addition of a penalized L1 norm (the sum of the absolute value of the coefficients).

145   While LASSO was originally developed as a complete regression method, we follow

the recommendations of Hastie et al. (2001) and use LASSO only for variable

importance ranking and dimensionality reduction of the input variable space to the

model.

We apply the linear LASSO method for rankings across a full year of 3-hourly

150   simulated canopy physics from the MEGAN3 canopy model at the global scale for the

year 2012. Input meteorology is from MERRA-2 assimilated meteorological fields at

2°x2.5° horizontal resolution (Gelaro et al., 2017), and the vegetation distribution

from the Olson 2001 land maps (Olson et al., 2001). LAI is derived from the MODIS

TERRA MOD15A2 Product (Myneni et al., 2002, 2007) re-gridded to 2°x2.5°

155   horizontal resolution and a monthly timescale. This input data is identical to that

used in the GEOS-Chem chemical transport model (CTM) described below for direct

comparison with prior work and ease of implementation into that CTM. The spatial

distribution of vegetation and LAI are summarized in Figures 1 and 2, respectively.

In general, forested land classes have the highest LAI values, and are spread

160 throughout the tropics and the northern latitudes. Crops, grasses, and shrubs are located predominantly in transitionary landscapes and near regions of larger population (e.g. India, Central North America, etc.), and tend to have lower LAI values.

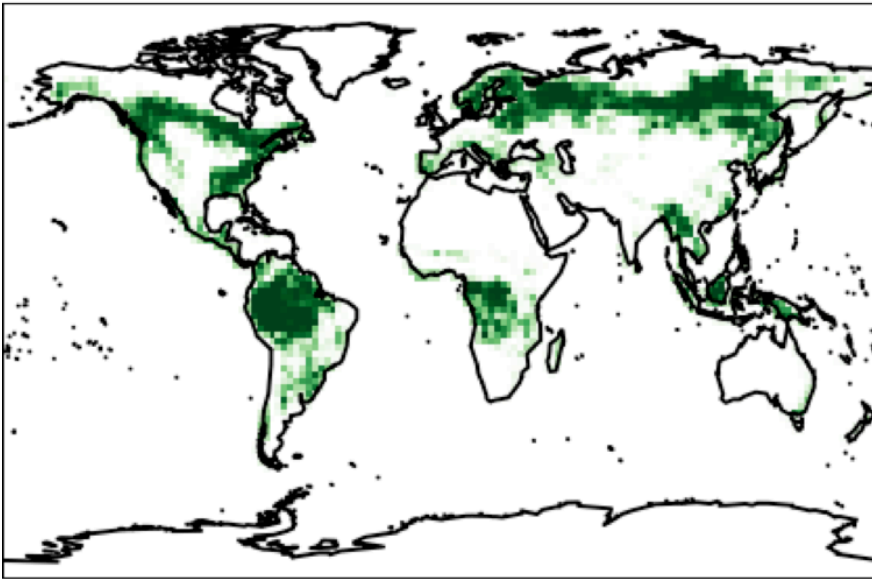
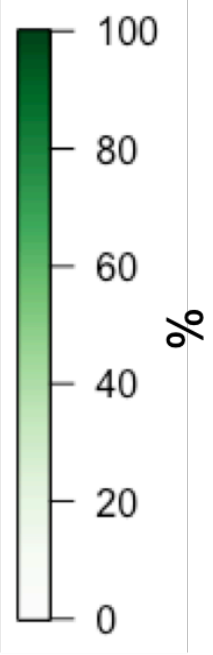
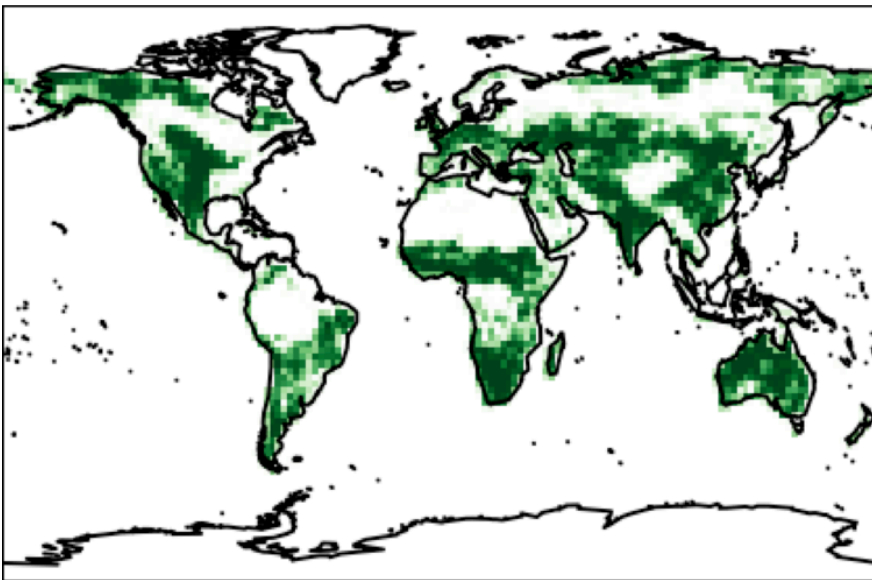



Figure 1. The percent of each 2°x2.5° gridbox occupied by each vegetation class used in this work. Panel a) is forested vegetation, and panel b) is crops, grasses, and shrubland.

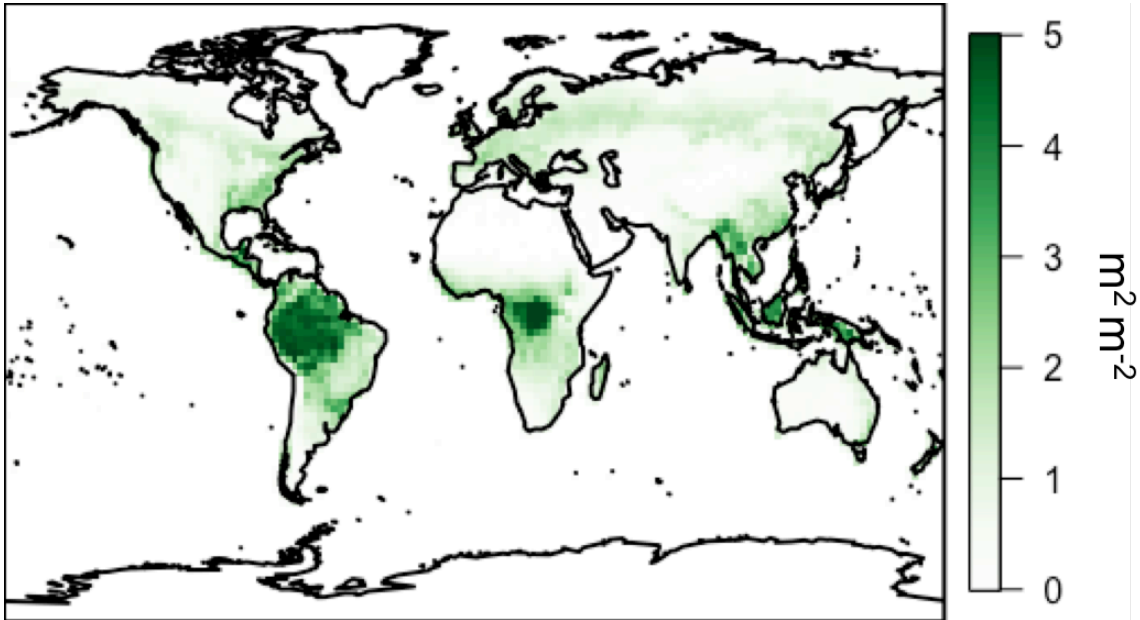



Figure 2. Annual Average LAI from MODIS for the year 2012.

The LASSO importance rankings are remarkably consistent for both sunlit and shaded leaves and for all vertical levels of the canopy. For each quantity, the two highest ranked variables are consistent at each layer throughout the canopy, and have substantially larger importance to the final result than any additional variable. For brevity we discuss only those first two variables here. The two most important variables for the calculation of leaf temperature are air temperature and wind speed. Air temperature dominates in importance for the calculation of leaf temperature, with a larger coefficient emerging at a higher L1 norm penalty weighting. Other variables that are physically important in nature (e.g. solar

radiation) do not appear important in the LASSO rankings due in part to how they

180   covary with air temperature, and how the rankings are derived separately for sunlit

and shaded leaves. For the calculation of leaf PAR we find that the two most

important variables are PAR out of the lowermost atmospheric gridbox (incident on

the canopy), and the local vegetation LAI. We use these selected variables to develop

a simplified parameterization for leaf temperature and PAR.

185   We model leaf temperature for a given canopy level, i ($T_{i,leaf}$, K) as linear with 2-

meter air temperature (T_{air}, K):

(1) $T_{i,leaf} = A_i + B_i * T_{air}$

Where A_i and B_i are fitted parameters per canopy level (i). This ignores the addition

of the second most important variable, wind speed. However, the addition of wind

190   speed to the regression only improves the performance of the model by less than

1% in total bias and R^2; thus for simplicity, we neglect this variable.

For the calculation of leaf-level PAR at a given canopy level ($PAR_{i,leaf}$, $\mu mol\ m^{-2}\ s^{-1}$),

we use an exponential Beers-Law analogue, including the influence of Leaf Area

Index (LAI) and the PAR incident to the top of the canopy (PAR_{toc}, $W\ m^{-2}$):

195   (2) $PAR_{i,leaf} = PAR_{toc} * exp(C_i + D_i * LAI)$.

Where C_i and D_i are fitted parameters per canopy level (i).This exponential

functional form is chosen due to the observed and simulated relationships between

LAI and canopy light interception (Engel et al., 1987; Goudriaan and Monteith, 1990) following a similar functional form.

200   We fit equations 1 and 2 for all layers of the canopy and for sunlit and shaded leaves, resulting in 20 total free parameters necessary to model the entire plant canopy across the globe. In this regression method, we ignore the role of differing vegetation classes and apply the regression agnostic to vegetation type. This is done to keep the total necessary number of free parameters low (20 versus 120), and

205   because this more parsimonious model performs quite well (see Section 3.1) without the need for additional vegetation type specific coefficients. The resulting surrogate model coefficients are summarized in Table 1.

	Canopy Level	A	B	C	D
Sunlit Leaves	1	-13.891	1.064	1.083	0.002
	2	-12.322	1.057	1.096	-0.128
	3	-1.032	1.031	1.104	-0.298
	4	-5.172	1.050	1.098	-0.445
	5	-5.589	1.051	1.090	-0.535
Shaded Leaves	1	-12.846	1.060	0.871	0.015
	2	-11.343	1.053	0.890	-0.141
	3	-1.068	1.031	0.916	-0.368
	4	-5.551	1.051	0.941	-0.592
	5	-5.955	1.053	0.956	-0.743

210   **Table 1. Regression Coefficients for the Canopy Surrogate Model. Canopy level 1 represents the top of the canopy.**

The final quantity necessary for the canopy model is the fraction of sunlit and shaded leaves. Here, that fraction in each layer of the plant canopy is calculated

directly following the MEGAN3 code, (see Guenther et al., 2006, 2012), without any

model simplification. The sunlit fraction is calculated as follows:

215   (3) $K_b = 0.5 * \frac{C_1}{\sin \beta} \frac{\alpha_1}{\sin \beta}$

(4) $Sunlit\ Fraction = \exp\left(K_b * \frac{LAI}{1-\alpha_2} * f\right)$

Where β is the solar angle above the horizon, K_b is the extinction coefficient for

black leaves, C_1 is the canopy clustering coefficient, C_2 is the canopy transparency,

and f is the fraction of biomass in the canopy light travels through to reach a given

220   leaf (the vector [0.05, 0.23, 0.5, 0.77, 0.95]). Consistent with the MEGAN3 parent

canopy model, we assume a Gaussian distribution of biomass in the canopy,

centered in the middle canopy layer, and a canopy transparency of 0.2, and a leaf-

clustering coefficient of 0.9.

From this relatively simple three-function parameterization (Leaf Temperature,

225   Leaf PAR, and Sunlit Leaf fraction), we are able to implement more physically

realistic parameterizations for biosphere-atmosphere interactions in atmospheric

chemical transport models.

3.1. Surrogate Model Performance

Here, we evaluate the surrogate model performance for all vegetation globally.

230   **3.1.1 Temperature**

The surrogate model simulated annual canopy average leaf temperature

distribution and performance for 2012 is summarized in Figures 3 and 4. In Figure

3, the average for each canopy layer is calculated as a sum of the sunlit and shaded leaves, weighted by the sunlit fraction of that layer. In turn, the canopy average is calculated as the weighted sum of the layer averages, weighted by the fraction of the canopy biomass in each layer. The annual average temperature is shown in Figure 3a, where it largely follows a latitudinal gradient. The warmest temperatures are ~310 K in the tropics, and the coldest average leaf temperatures are ~280 K in the northern high latitude boreal regions. The surrogate model is linear with the 2-meter "near-surface" air temperature, and therefore follows that spatial distribution directly.

The surrogate model for leaf temperature performs well, with the annual average spatial R^2 and mean bias relative to the full model shown in Figures 3b and 3c, respectively. Across all regions, the R^2 is very high, indicating that a linear relationship between 2-meter air temperature and canopy average temperature is a good approximation for capturing the variability of the full MEGAN3 canopy model. The temperature R^2 drops below 0.90 only in coastal regions/gridboxes that contain very little vegetation, representing less than 5% of all vegetated areas. The temperature surrogate bias is also generally quite low, as shown in Figure 3c. The majority of regions have an absolute mean bias of less than 1 K, and more than 90% of the annual average mean biases are less than 2 K. The surrogate model generally computes temperatures that are biased cool over highly vegetated tropical and sub-tropical regions, and slightly overestimates temperature over northern boreal forests (by ~0.1 K). The most substantial overestimations occur in or near the hot and arid regions of the globe, and always in regions where there is little vegetative

cover at all. On a relative scale these biases are quite small; all are less than 1% of the total magnitude. The model development process described in Section 3 removed the vegetation type discrimination in the parent MEGAN3 canopy model. This leads to some spatial coherence in the global bias patterns, where regions

260   dominated by coniferous forests, grasses, and shrubs (e.g. boreal northern hemisphere and the western United States) tend to be slightly biased warm, and the other vergetation types are biased slightly cool.

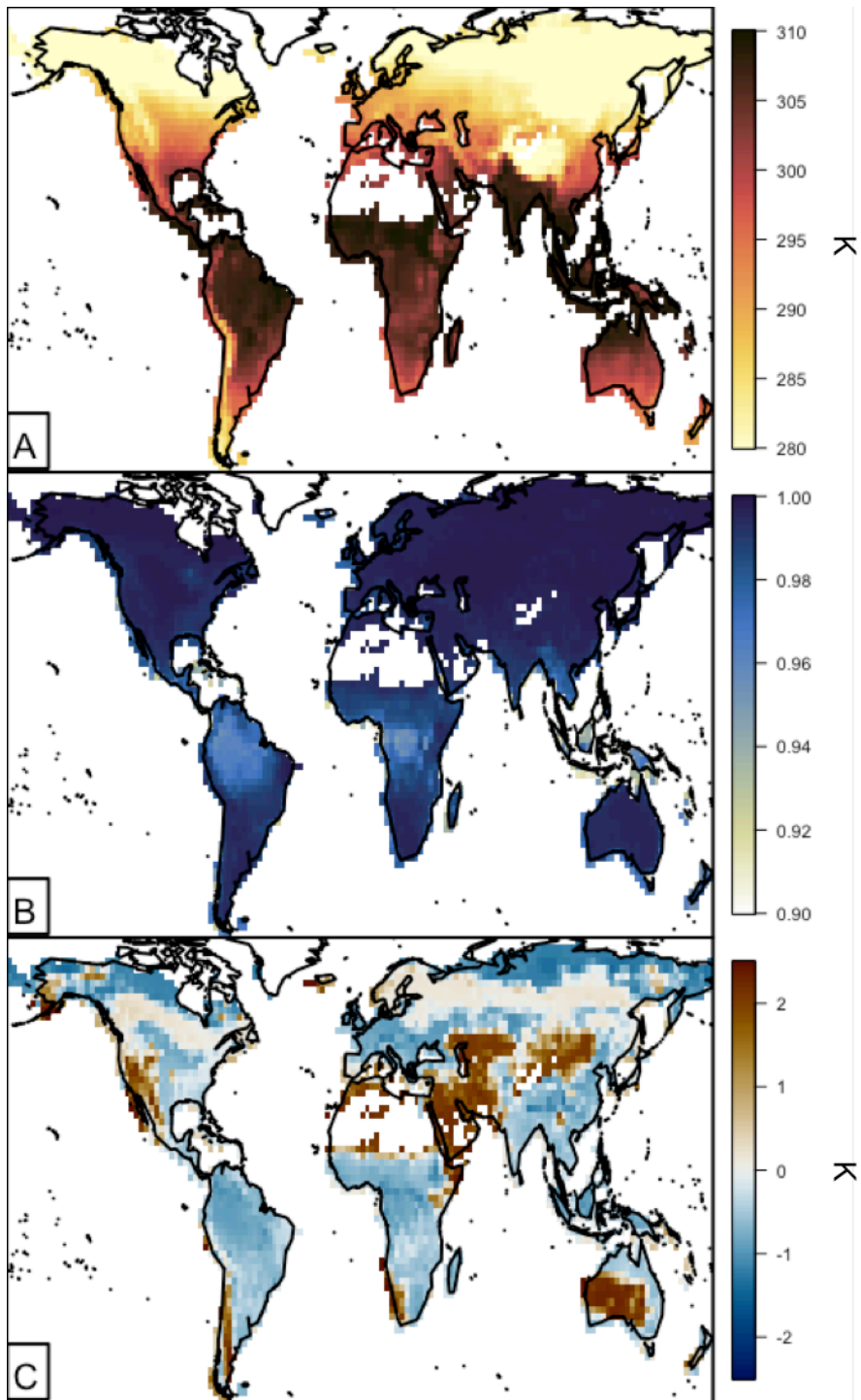



Figure 3. Surrogate model performance for the annual canopy average leaf temperature in 2012. Panels are as follows: A) Annual average surrogate model leaf level temperature (Kelvin), B) R² between the surrogate and the full model leaf level temperature, C) Annual average leaf level temperature bias (surrogate-full model, K)

265

The vertical profile of annual average leaf temperature is shown in Figure 4a. The broad shape of the canopy profile is consistent across vegation types and the globle.

270   The upper canopy layers are cooler than the lower canopy layers, as an insulating effect from air temperature occurs within the canopy. The higher order variability (e.g. small differences between adjacent layers at the top and bottom of the canopy) stems from the more detailed representation of canopy energy balance in the full MEGAN3 model, which includes the influence of terms like PAR, relative humidity,

275   LAI, and wind speed. However this higher order variability is quite consistent, allowing for it to be reproduced in the simplified surrogate model.

Similar to the spatial performance, the overall surrogate model performs well throughout the canopy. The surrogate model temperature R^2 is shown in Figure 4b. The values are all near 1.0, with the lowest value of 0.97 in the middle of the canopy,

280   where the transition from cooler to warmer leaves is slightly more difficult for the surrogate model to capture. As demonstrated in Figure 4c, the bias throughout the canopy is low as well. On a global annual average, the surrogate model is biased cool, but only slightly (on both a relative and absolute scale). The highest magnitude bias is at the top canopy layer, at -0.04 K.

285

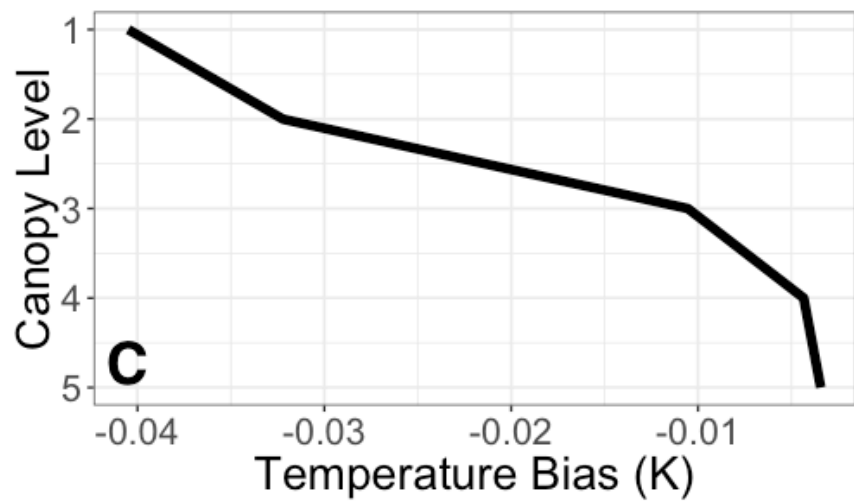
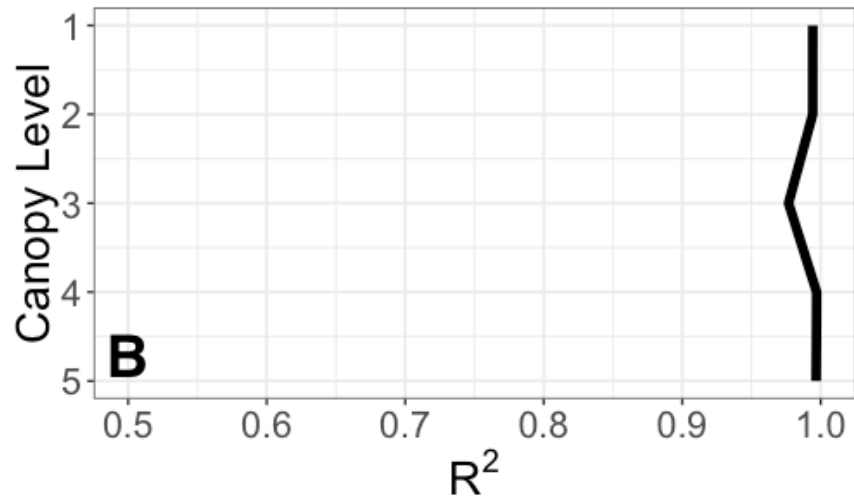
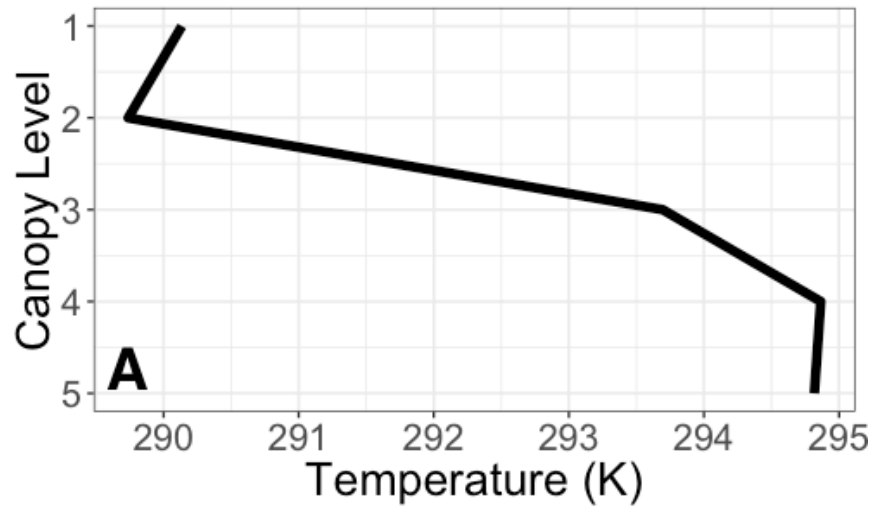



Figure 4. Surrogate model performance for the annual average vertical canopy temperature profile in 2012. Panel A shows the vertical average surrogate model leaf-level temperature (Kelvin). Panel B shows the surrogate model R^2 against the full model. Panel C shows the leaf level temperature bias (K) of the surrogate model compared to the full model.

3.1.2 Photosynthetically Active Radiation

The annual canopy average leaf-level PAR for 2012 is shown spatially in Figure 5a. In Figure 5, canopy temperature averages are calculated using the same method as for canopy temperature. Annual average PAR varies from ~200 µmol m^{-2} s^{-1} to greater than 600 µmol m^{-2} s^{-1}. This spatial variability is a function of both PAR incident on the top of the canopy (largely related to cloud cover and solar angle) and the canopy LAI. Leaf level PAR in the surrogate model varies linearly with incident PAR, and decreases exponentially with LAI. Additionally, the reduction under high LAI is exacerbated due to a higher fraction of shaded leaves in high LAI canopies, which have substantially lower average PAR. The highest leaf-level PAR values are generally located in arid regions, where LAI and the number of cloudy days are quite low. The lowest values are located in the equatorial tropical rainforests and the northern boreal forests. The low leaf level PAR values in the rainforest are coincident with the highest LAI values globally, leading to very strong shading effects below the first canopy layer. The northern boreal forests have low leaf level PAR in part due to relatively high LAI, but also due to reduced incoming PAR in the winter months when the solar angle is low.

The annual average performance of the surrogate leaf-level PAR relative to the full model is shown in Figures 5b and 5c. The temporal R^2 over a full year in Figure 5b, is generally quite high indicating that the surrogate formulation captures the majority of the temporal variability inherent in the full model. The R^2 values range from 0.92

to 1.0. The highest R^2 values are in regions with low LAI, where the effects of shading and other canopy physical processes are greatly reduced. The worst model R^2 performance is over two characteristic regions: higher elevations with low vegetation densities (i.e. global deserts in Central Asia and South America), and

315 those with the highest LAI. Both regions represent extreme scenarios for canopy radiative physics. The higher elevation regions with low vegetation densities have very little leaf shading at all throughout the year, and thus the canopy physics represented with a simple exponential decay is no longer as relevant. In the high LAI regions, the elevated importance of shading and resulting complexity in the PAR

320 calculation is more challenging for the simplified representation of the surrogate model. However, this poor performance still has a quite high R^2, with the lowest value of 0.92. The annual average model biases are generally within ±40 μmol m^{-2} s^{-1}, with a few more extreme values reaching ±200 μmol m^{-2} s^{-1}. The surrogate model is broadly biased high over regions with lower LAI and slightly low over regions

325 with high LAI. In a relative sense, these changes are nearly all within 10-15%, with a maximum normalized mean bias of 0.4.

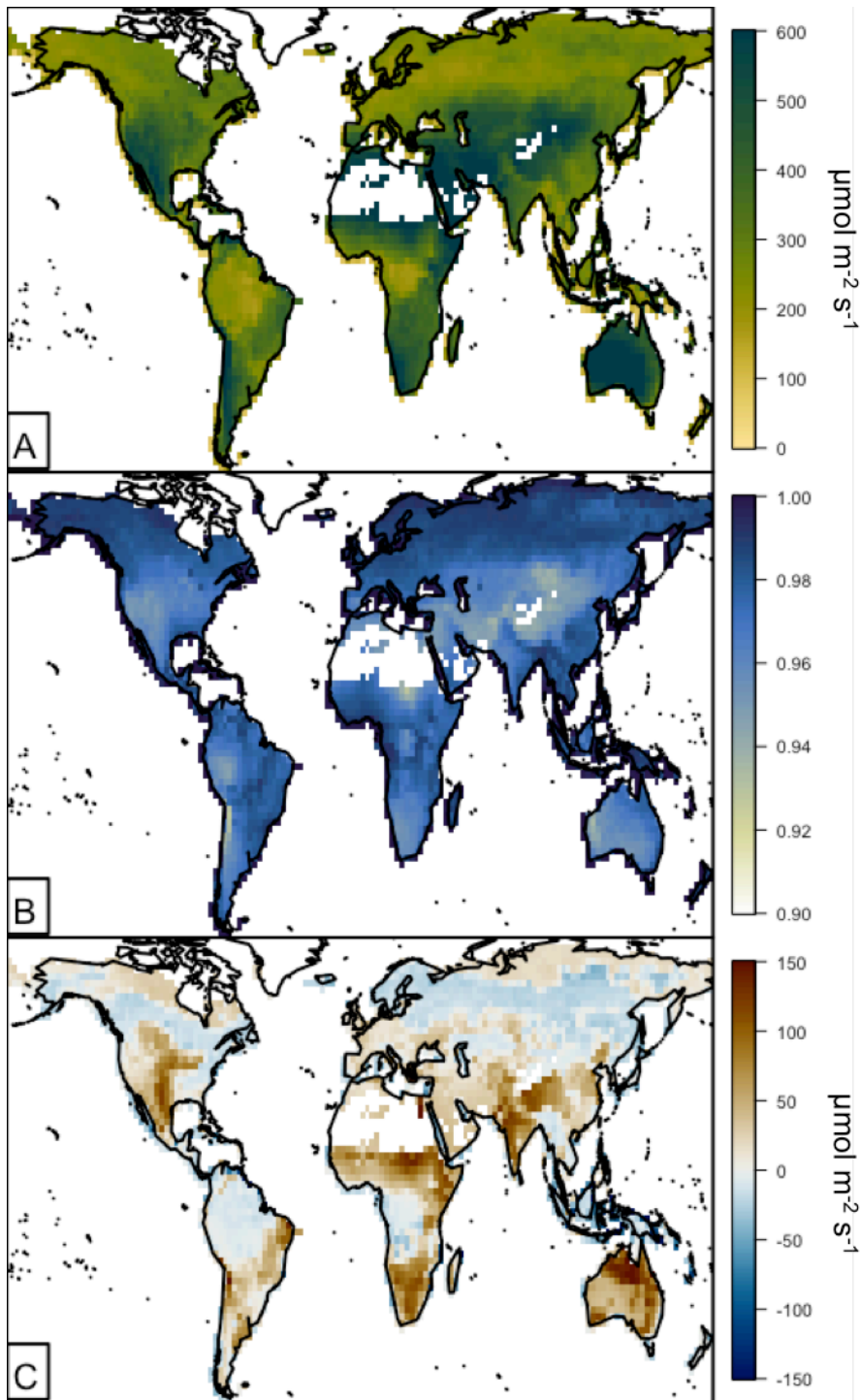



Figure 5. Surrogate model performance for the annual canopy average leaf PAR in 2012. Panels are as follows: A) Annual average surrogate model leaf level PAR (µmol m⁻² s⁻¹), B) R² between the leaf level PAR simulated using the surrogate and the full model, C) Annual average leaf level PAR bias (surrogate-full model, µmol m⁻² s⁻¹).

330

The average vertical distribution of leaf-level PAR throughout the canopy and the associated surrogate model performance are shown in Figure 6. To explore the additional dependence on LAI, the quantities shown are separated into three LAI

335   ranges. These are as follows: a low range with LAI less than 0.5, a midrange with LAI between 0.5 and 5, and a high LAI range containing canopies with total LAI greater than 5. The low LAI range represents ~40% of all vegation throughout the year, the middle range represents nearly 60%, and the high LAI range contains only a small fraction of all vegetation (~1%).

340   The average distribution of PAR across canopy levels is shown in Figure 6a. As LAI increases, there is a substantial reduction in leaf level PAR deeper into the vegetated canopy. This is particularly obvious with the high LAI range, consistent with substantial shading and light interception above the bottom of densely vegetated canopies. On the other hand, the variability throughout the low LAI canopies is quite

345   small. This LAI dependence explains in part the relatively low canopy average leaf-level PAR throughout the tropical forests in Figure 5a. The variability in the PAR at the top canopy layer (Canopy Level 1 in Figure 6a) stems from two major sources. The first is simply the spatial distribution of these LAI ranges in relationship to the annual average incident PAR to the canopy top. Very high LAI values occur primarily

350   over the tropics, where sunlight is consistently high throughout the year, and the seasonal effects of changing solar angles is small. The opposite is true for many of the regions with smaller LAI values, which are distributed more evenly across the globe. A second order effect in the MEGAN3 canopy model is that of in-layer attenuation of light and shading throughout the canopy, where leaves in a given

355   layer may intercept light and shade leaves lower within that same layer. This has the

effect of reducing the layer average leaf-level PAR as a function of leaf geometries

and LAI, and explains why the highest canopy layer average leaf-level PAR is not the

same as the average PAR incident on the top of the canopy.

Figures 6b and 6c summarize the statistical performance of the surrogate model

360   vertically through the canopy in terms of the R^2 and the mean bias, respectively.

Overall, the surrogate model reproduces the PAR variability as compared to the full

parent model well. For both the low and middle LAI ranges (LAI less than 5), all R^2

values are greater than ~0.9. The only substantially lower R^2 are from the lower

canopy in high-LAI regions, where PAR is generally quite small (see Figure 6a). The

365   surrogate model struggles somewhat to capture this lower canopy variability due in

large part to the increased complexity of resolving canopy shading and radiative

physics in high LAI canopies. However, the ultimate influence on the total canopy-

scale bias is generally low.

The vertical distribution of that bias is shown in Figure 6c. Broadly, the absolute

370   PAR bias is low (less than 5% on a relative scale) and decreases throughout the

canopy. All biases are positive except for the top canopy layer for high LAI range

canopies, and this poor fit is likely related to the limited representation of high LAI

regions in the full dataset (only ~1% of all vegetated area), and isn't present if a

lower cutoff for high LAI ranges is used (e.g. LAI > 3). The decreasing magnitude

375   throughout the canopy is largely related to the decreasing overall leaf-level PAR (see

Figure 6a). It important to note that the bias terms are all sensitive to the choice of

LAI bin ranges, and the variability in bias at each level can be quite large (e.g. above

50 μmol m^{-2} s^{-1} in the top canopy layer). For both the high and middle LAI ranges,

the absolute magnitude of the PAR bias decreases throughout the canopy, and the

380   bias remains relatively constant for low LAI range vegetation. On a relative scale,

these biases are all quite small with a normalized mean bias usually less than 5%.

The exception to this is the lowest layer of the high LAI range canopies. In this low

canopy layer the magnitude of the bias is quite low, as is the total magnitude of leaf-

level PAR, the resulting difference between small numbers leads to a relatively large

385   normalized mean bias of ~0.3.

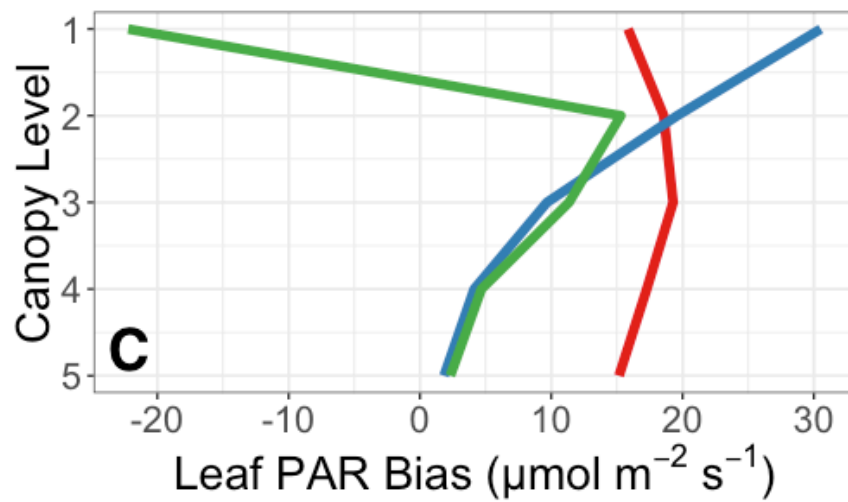
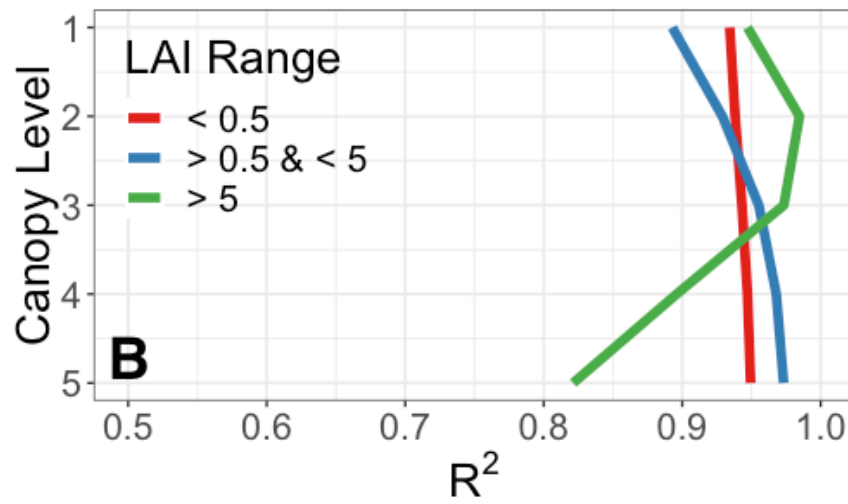
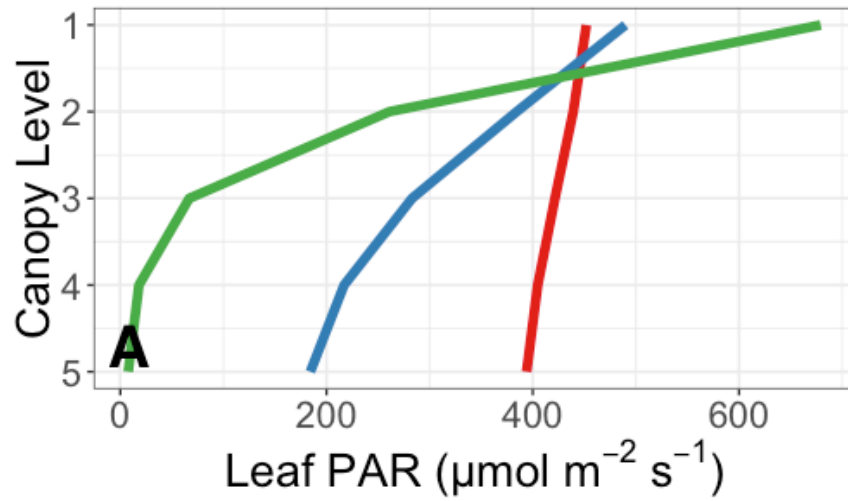



Figure 6. Surrogate model performance for the annual average vertical canopy PAR profile in 2012, as a function of LAI. Panel A shows the vertical average surrogate model leaf-level PAR (µmol m^{-2} s^{-1}) for low LAI (red), mid-range LAI (blue) and high LAI canopies. Panel B shows the surrogate model R^2 against the full model. Panel C shows the leaf level PAR bias (µmol m^{-2} s^{-1}) of the surrogate model compared to the full model. Level 1 is the top of the canopy.

An essential function of canopy models used in CTMs is to calculate the amount of

light that falls on already light-saturated leaf surfaces. This regulates the effect of a

change in PAR incident on the canopy on various biological and physical processes

(e.g. biogenic emissions). We estimate the fraction of leaves that are light-saturated

using the γ_{PAR} formulation from the MEGAN algorithm (Guenther et al 2006, 2012).

This variable aims to capture the amount of light saturation on a given leaf, and

ranges from 0 to 1, with higher values corresponding to more saturated leaves. To

explore light saturation, we examine cases where the γ_{PAR} value is greater than 0.9. A

scatterplot of the annual average fraction of leaves that are light-saturated ($\gamma_{PAR} \geq$

0.9) per model gridbox for both the full model and the surrogate model throughout

the canopy is shown in Figure 7. The surrogate model reproduces the full model

fraction of light-saturated leaves well, generally to within ~5%, with a median bias

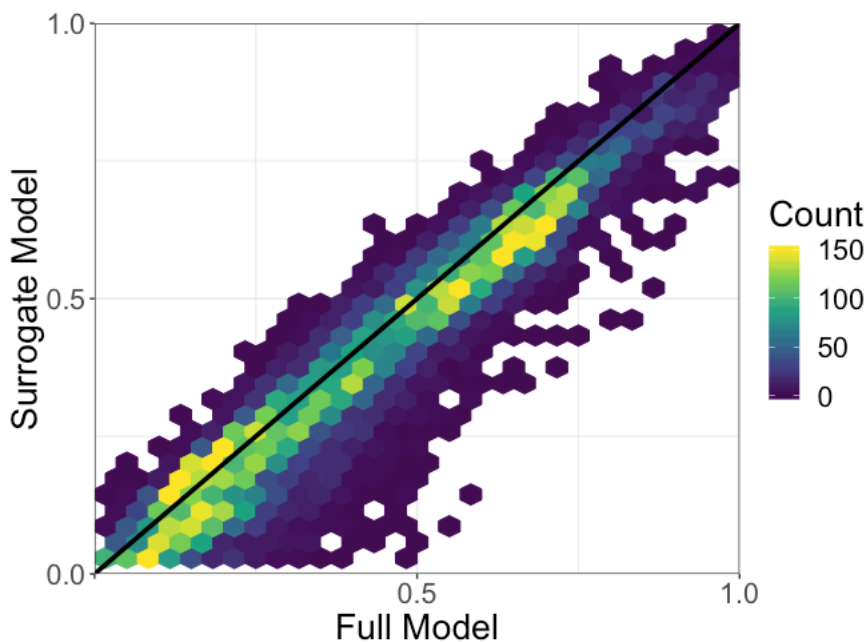
of -2%.

405

Figure 7. The annual grid box average fraction of light saturated leaves as simulated by the full and surrogate models throughout the canopy for the year 2012. The colorbar represents the number of observations in a given hex. The 1:1 line is shown in black.

Ultimately, this assessment demonstrates that the surrogate model reproduces the

410   parent MEGAN3 canopy model well for both leaf temperature and leaf-level PAR.

The exponential relationship between leaf-level PAR and canopy incident PAR and

the linear relationship between leaf temperature and near-surface air temperature

captures the majority of the information inherent in the parent model. Some higher

order variability in absolute magnitude of the variables is missing from this

415   surrogate model, however the biases are generally all within ~10%.

4. Chemical Transport Model Description

We evaluate the impact of the canopy model parameterization on atmospheric

composition using the GEOS-Chem v12.3.0 chemical transport model (www.geos-

chem.org). GEOS-Chem is a computational model for simulating atmospheric

420   chemistry, including a detailed HO_x-NO_x-BrO_x tropospheric chemical mechanism

(Bey et al., 2001; Mao et al., 2013; Travis et al., 2016). We drive GEOS-Chem with

MERRA-2 Meteorology at 2°x2.5° spatial resolution, with 47 vertical layers (Gelaro

et al., 2017). The timesteps for convection and chemistry are 10 and 20 minutes,

respectively. Identically to the canopy model input data, we use LAI values from the

425   MODIS-Terra MOD15A2 product (Myneni et al., 2002, 2007), and plant functional

types from the Olson 2001 dataset (Olson et al., 2001). Fire emissions are from the

Global Fire Emissions Database v4 (GFED4, Giglio et al., 2013), and global

anthropogenic emissions are from the Community Emissions Data System inventory

(CEDS, Hoesly et al., 2018). Regional emissions over the United States, Africa, and

430   Asia are from the NEI 2011 (Travis et al., 2016), DICE-Africa (Marais and

Wiedinmyer, 2016), and MIX (Li et al., 2017) emissions inventories, respectively.

Soil NO_x emissions are calculated following Hudman et al. (2012). Simulations are

shown for the years 2012 and 2013, with the first year discarded for spin up when

considering gas-phase chemical impacts.

435   **4.1 MEGAN Emissions**

The biogenic emissions scheme in GEOS-Chem v12.3.0, MEGAN2.1, is based on

Guenther et al. (2006, 2012) and Millet et al. (2010). The emissions of a given

compound are calculated from base canopy-level emission factors multiplied by

"activity factors" representing standard processes that govern biogenic emissions

440   (temperature, PAR, light-dependence, etc.), and "stress factors" modeling the effect

of vegetative stress (heat, drought, etc.) on biogenic emissions. Each of these activity

and stress factors vary with the environmental state. The base emission factor itself

varies with vegetation type, and these activity factors respond to leaf temperature,

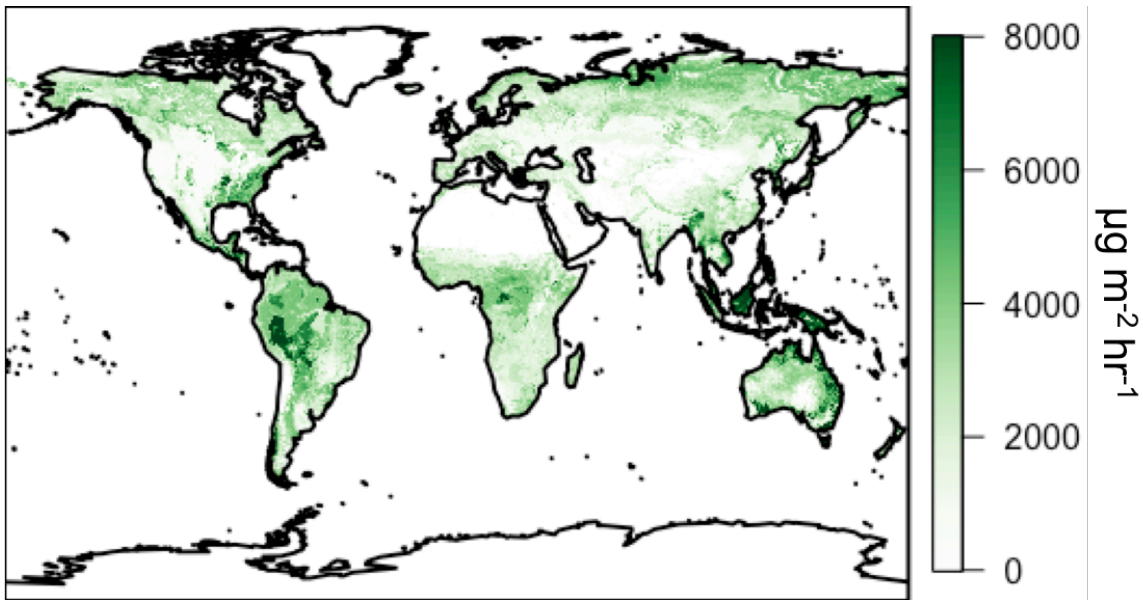
leaf-level PAR, leaf age, leaf area index, soil moisture, and atmospheric CO_2

445   concentrations. The base emission factors used in this work are consistent with

those used in GEOS-Chem v12.3.0; an example for isoprene is shown in Figure 8. The

emission factors are highest in forested regions, and lowest over areas with little

vegetation (e.g. deserts). These emission factors are regridded from the original

resolution of 0.25°x0.3125° to match the GEOS-Chem resolution of 2°x2.5°.

450   **Figure 8. Base Isoprene emission factors used in this work.**

As GEOS-Chem v12.3.0 has no representation of plant canopy physics, the LAI,

temperature, and PAR activity factors are all re-parameterized following Guenther

et al. (2006) in the standard model. In this parameterization, leaf temperature is

455   assumed to equal air temperature in the calculation of the temperature activity

factor. The LAI and PAR activity factors are calculated in an approach known as the

Parameterized Canopy Environment Emission Activity (PCEEA) approach that does

not include any description of the vertical distribution of vegetation, and only

includes response to the LAI, PAR incident to the top of the canopy, and the solar

460   zenith angle.

We modify the MEGAN implementation in GEOS-Chem to allow for the

representation of canopy physics decribed in Section 3. In order to properly scale all

emission factors to the plant canopy using a canopy model, a normalization factor

must be applied at a set of standard environmental and ecological conditions

465   (Guenther et al., 2006, 2012). This normalization factor varies depending on the

choice of those standard conditions and the canopy model used. In MEGAN2.1 these

standard conditions are: LAI of 5, current air temperature of 303K, current incident

PAR at the canopy top of 1500 μmol m^{-2} s^{-1}, and a 10%/80%/10% split of growing,

mature, and senescent leaves (Guenther et al., 2012; Kaiser et al., 2018). We

470   calculate other necessary standard conditions, specifically the 24-hour average air

temperature and PAR, from the meteorological fields conditional on locations that

meet the previously described standard conditions. In situations where all of the

previous instantaenous standard conditions (e.g. Current Temperature = 303K,

Current PAR = 1500 μmol m^{-2} s^{-1}, and LAI = 5) are jointly met to within ±10%, we

475   calculate the 24-hour average prior meteorological conditions from the global

reanalysis fields, and use the mean of those calculations as the standard 24-hour

average conditions. The resulting standard conditions for 24-hour average

temperature and 24-hour average PAR are 298.5 K and 740 μmol m^{-2} s^{-1},

respectively. These standard conditions result in a normalization factor of 0.21

480   using the surrogate canopy model surrogate developed in this work. The value of

0.21 is lower than those used in implementations of previous MEGAN model versions in other models such as CLM (0.3) and WRF-Chem (0.57) (Guenther et al., 2012). It is important to note that the normalization factor approximately scales with the square of the current temperature conditions and linearly with the current

485   PAR conditions. For the 24-hour average conditions, the scaling is reduced to approximately linear for temperature and as a square root for PAR. Given this, small deviations from these standard conditions (e.g. those that could arise from different 24-hour averaging methodology) can lead to substantial changes in the normalization factor. Additionally, consistent with previous work these standard

490   conditional calculations are likely variable across model meteorological configurations and should be recalculated on a model-specific basis (Guenther et al., 2012). Since this normalization factor is applied consistently to all emissions globally at all times, it linearly modulates all biogenic emissions. As such, the total emissions calculated by the MEGAN2.1 emissions framework are highly sensitive to

495   the parameter choices made in this normalization processes.

In GEOS-Chem v12.3.0 we update the activity factors associated with PAR, LAI, and temperature as well as the normalization to take advantage of our new canopy surrogate model. This enables a full implementation of the most recent MEGAN3 emission activity algorithm in the GEOS-Chem model. In the PCEEA implementation

500   of MEGAN in the base version of GEOS-Chem, activity factors are calculated separately for PAR (γ_P), LAI (γ_{LAI}), and temperature (γ_T), and then multiplied together following Guenther et al. (2006):

$$(5) \; \gamma_{PCEEA} = \gamma_{LAI}\gamma_T\gamma_P$$

Following MEGAN3 (https://bai.ess.uci.edu/megan), we implement PAR and temperature activity factors that are calculated jointly per canopy level and summed together weighted by the vertical canopy biomass distribution. In this work, as in previous non-PCEEA versions of the MEGAN framework (Guenther et al., 2006, 2012), the effect of LAI is calculated through direct multiplication of the emission factor by LAI as opposed to an activity factor formulation, along with a canopy normalization factor (C_{CE}):

$$(6) \; \gamma_{Canopy} = C_{CE}LAI\gamma_{TP}$$

$$(7) \; \gamma_{TP} = \sum_{l=1}^{5} w_l \, \gamma_P\gamma_T$$

These activity factors for PAR, LAI, and temperature are the same as those in Guenther et al. (2012), as averages throughout the canopy weighted by the biomass fraction within a given canopy layer (w_l). There is an additional canopy depth emission activity response applied to the light dependent activity factors which is intended to model the variability of emissions throughout the canopy (e.g. Harley et al., 1996). This canopy depth activity factor is a multiplicative factor that varies linearly as a function of LAI and canopy depth, with a value between 0 and 1.3. For clarity, we refer to the MEGAN emissions implementation in GEOS-Chem using the γ_{PCEEA} activity factors as "MEGAN_PCEEA" and those using the γ_{Canopy} activity factors as "MEGAN_Canopy". While γ_T in the MEGAN_PCEEA approach follows a similar functional form to that in MEGAN_Canopy, the lack of vertical canopy structure in the MEGAN_PCEEA

configuration leads to a very different treatment of the joint effects of Temperature,

525   PAR, and LAI on emissions. Specifically, the MEGAN$_{PCEEA}$ approach aims to

approximate the joint effects of shading and temperature change within a canopy,

whereas MEGAN$_{Canopy}$ aims to directly simulate those processes. We use the canopy

physics surrogate model described in Section 3 to calculate the leaf temperature and

PAR in the MEGAN$_{Canopy}$ implementation.

530   Though stress factors in the MEGAN framework allow for the additional capability

to evaluate the impact of vegetative stress processes on emissions (e.g. Geron et al.,

2016), we do not enable those processes in this study. The other activity factors

(leaf age, soil moisture, and CO_2 inhibition) are the same in both the MEGAN$_{Canopy}$

and MEGAN$_{PCEEA}$.

535   **4.2 Dry Deposition**

Dry Deposition in GEOS-Chem v12.3.0 is calculated through a resistor-in-series

approach based on the Wesely (1989) parameterization, originally described and

implemented in Wang et al. (1998). In this approach, the dry depositional flux of gas

phase species is calculated as the surface concentration of that gas multiplied by a

540   transfer velocity known as the "dry deposition velocity". A recent assessment of the

dry deposition velocity parameterization in GEOS-Chem found that biases in

simulated dry deposition velocities are in general quite low, though there is

evidence that missing key processes may be responsible for missing variability in

the simulation (Silva and Heald, 2018).

545   Prior to this work, canopy effects were not directly considered in GEOS-Chem dry

deposition, and only approximated in bulk using a polynomial decomposition

scheme (Wang et al., 1998) that calculated a single constant jointly representing

both a multiplicative factor ($1 + b/PAR_{leaf}$, $b = 50$ μmol m^{-2} s^{-1}) to the stomatal

resistance from Baldocchi et al. (1987) based on leaf-level PAR and a normalization

550   of the stomatal resistance by LAI. Here, we replace the polynomial decomposition

scheme and use the leaf-level PAR calculations from the canopy surrogate to directly

calculate the multiplicative factor, and then explicity normalize by LAI. The LAI

normalization in the original polynomial decomposition calculates values that are a

factor of ~1.7 higher than those calculated through direct normalization when using

555   the surrogate model. To maintain the same magnitude of the simulated dry

deposition velocities as in the standard model, which are generally unbiased (Silva

and Heald, 2018), we scale the stomatal resistance by a factor of 0.6.

5. Surrogate Model Integration into GEOS-Chem

Implementing the updated canopy surrogate in a global model directly impacts the

560   surface-atmosphere exchanges processes of biogenic emissions and dry deposition,

which together influence the chemical composition of the atmosphere. In this

section we outline the changes to both surface processes, focusing on isoprene

emissions and ozone dry deposition, followed by the changes to surface level ozone

concentrations in the GEOS-Chem model.

565   The impact of the canopy model on isoprene emissions in 2012 is summarized in

Figure 9. The annual average isoprene emissions using the MEGAN$_{Canopy}$ emissions

implementation are shown in Figure 9a, with the highest emissions in the tropics and subtropics, as well as the Southeast United States. Though not distinct in Figure 9, the boreal forests are a substantial emitter of biogenic species during the summer

570   months. The relatively small emissions from this region during the winter months reduce the prominence of these emissions on the annual average. The global annual total of isoprene emitted in 2012 from the MEGAN$_{Canopy}$ configuration is ~350 Tg C yr^{-1}.

The annual average differences in the simulated isoprene emissions following

575   implementation of the surrogate canopy model (MEGAN$_{Canopy}$ – MEGAN$_{PCEEA}$) are shown in Figure 9b. In general, emissions decrease over forested regions and increase over non-forested (grasses, crops, and shrubland) areas. The highest absolute changes are the decrease in the Equatorial Amazon and the increase in Northern Australia. On a relative scale, the forested and non-forested differences are

580   more apparent. This relative change (MEGAN$_{Canopy}$/MEGAN$_{PCEEA}$) is shown in Figure 9c. While there are relatively modest decreases in tropical and boreal forests, the emissions increase in heavily cropped Indian Subcontinent and sub-Saharan Africa show the largest relative change. Though the spatial variability in relative difference is substantial, the annual global isoprene emissions from the canopy model are

585   within 5% of the original model version (~340 Tg C yr^{-1}). These results are consistent with those from Guenther et al. (2006) who found that the global total biases in isoprene emissions were low, but spatial variability was large, when using a parameterized approach (MEGAN$_{PCEEA}$) over a direct canopy model implementation in the MEGAN framework (as in the surrogate model application,

590   MEGAN$_{Canopy}$). On aggregate, these changes are all well within the stated uncertainty of the MEGAN isoprene emissions of approximately a factor of 2 (Guenther et al., 2012).

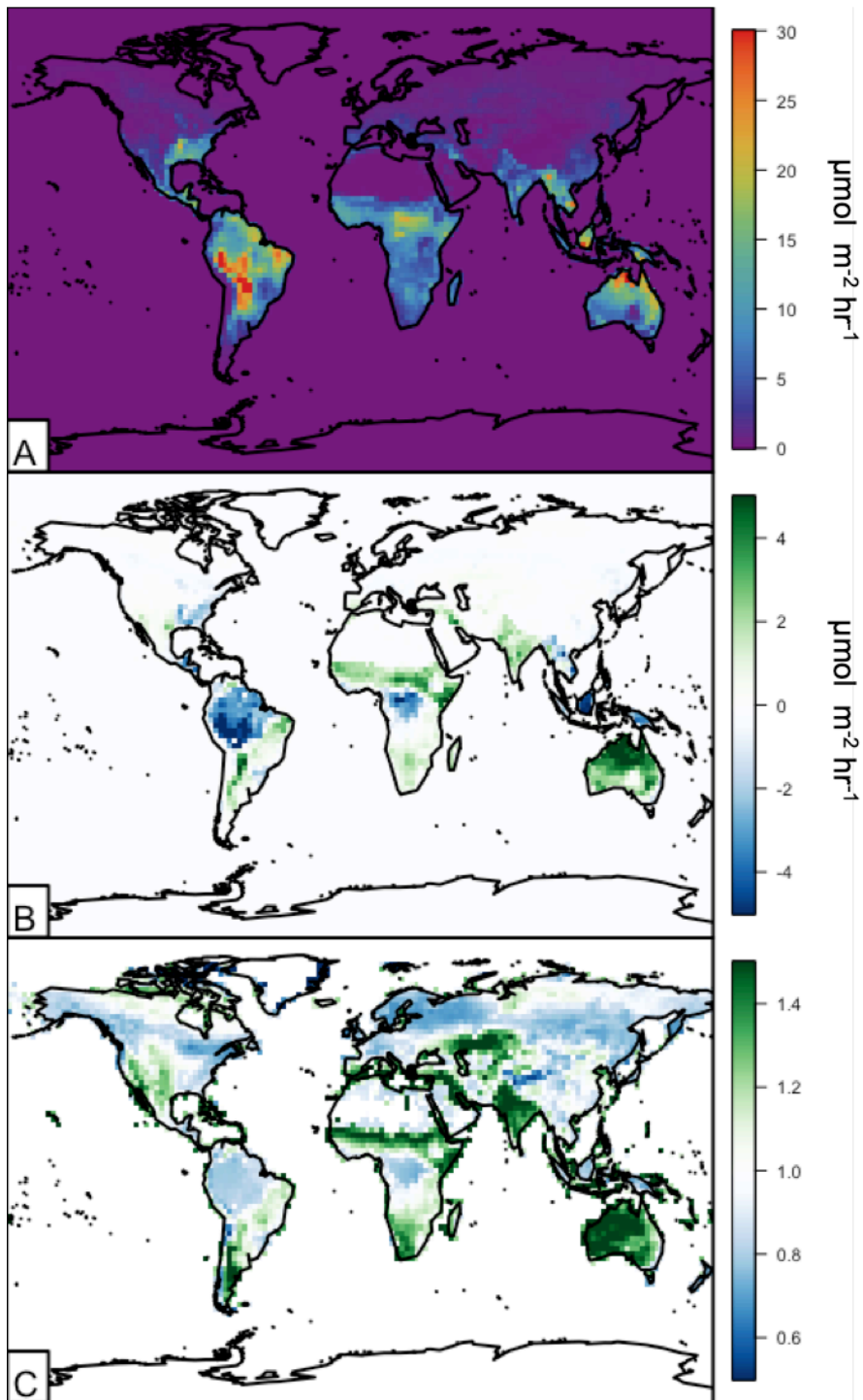



Figure 9. Annual average (2012) isoprene emissions simulated in GEOS-Chem driven by the surrogate model canopy physics (MEGAN$_{Canopy}$). Panel A shows the annual average emissions. Panel B shows the difference between the surrogate model and base version of simulated emissions. Panel C shows the relative difference between the surrogate model and base version of simulated emissions (surrogate/base model).

It is not possible to directly parse the individual process contributions to the total

600   emissions changes due to the fundamentally different coupled treatments of the

influence of temperature, PAR, and canopy structure on biogenic emissions through

the activity factors in both the MEGAN$_{Canopy}$ and the MEGAN$_{PCEEA}$ configurations.

However, a comparison of the isoprene differences between the two simulations

against LAI, leaf-level PAR, and leaf temperature (Figure 10) indicates that the

605   changes are most strongly driven by the leaf-level PAR and LAI effects. The isoprene

emissions changes are directly proportional to leaf-level PAR, inversely

proportional to LAI, and show no substantial relationship to leaf temperature. The

forested and non-forested differences in Figure 9 can be explained further from the

correlations shown in Figure 10. The forested areas with the largest decreases in

610   isoprene emissions tend to have high LAI values and lower canopy average leaf PAR,

whereas the opposite is true for the non-forest locations. The relationships in Figure

10 support the interpretation that the leaf-level PAR and LAI effects are the largest

drivers of change in biogenic isoprene emissions between the two model versions.

Overall, these results indicate that the representation of canopy radiative physics is

615   more important than thermodynamically resolving the difference between air and

leaf temperature for simulating biogenic emissions in the MEGAN framework.

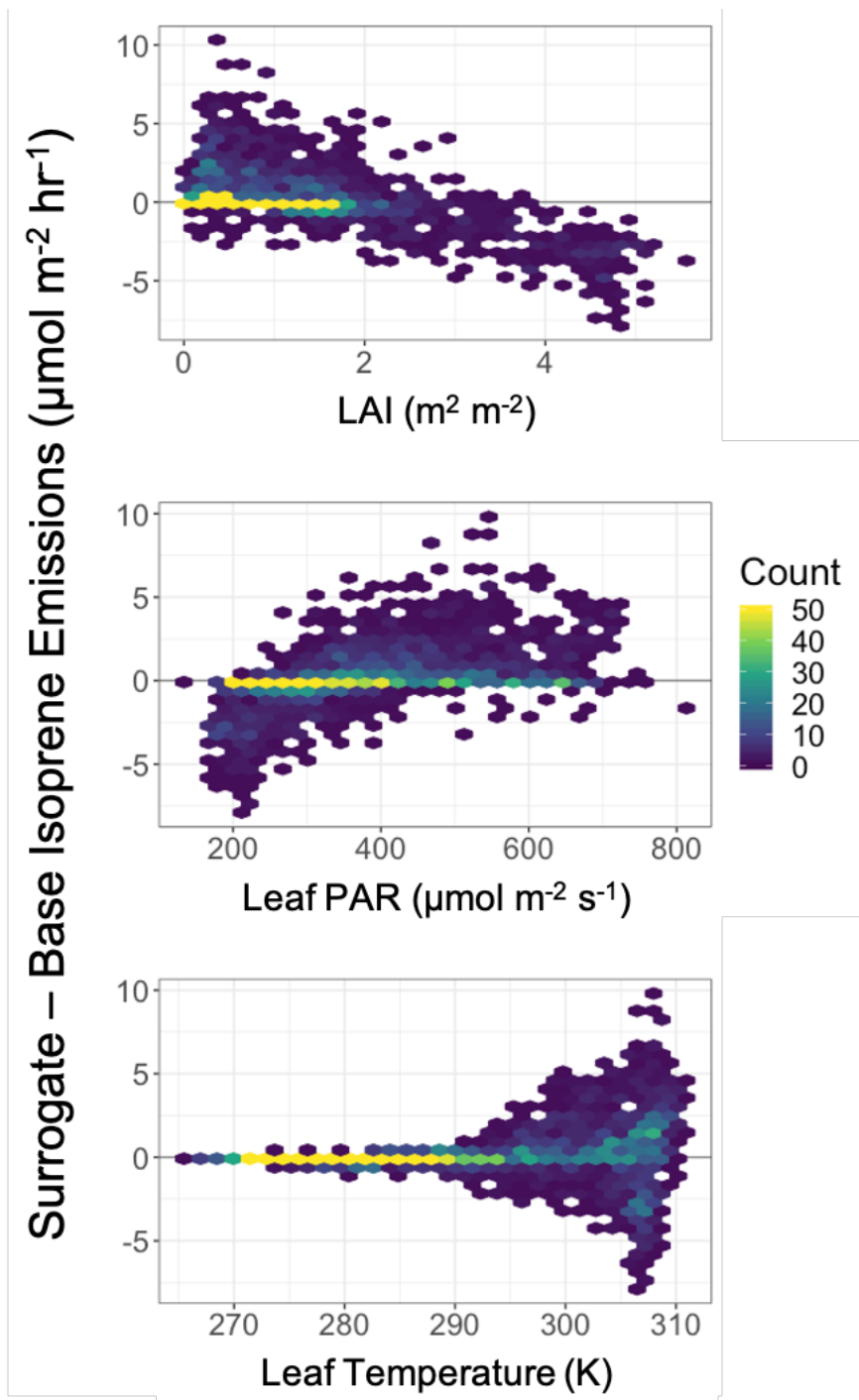



Figure 10. Difference in annual average isoprene emissions between the surrogate canopy model (MEGAN_Canopy) and the base simulation (MEGAN_PCEEA) (atoms C cm⁻² s⁻¹, see Figure 9B) as a function of LAI, Leaf Level PAR (μmol m⁻² s⁻¹), and Leaf Temperature (K). Gridboxes dominated by water were filtered removed from these figures. The colorbar represents the number of observations in a given hex.

620

There are few spatial constraints on isoprene emissions that can act as independent

validation data for the new model framework. However, recent work over the

southeast United States (Kaiser et al., 2018; Travis et al., 2016; Yu et al., 2016)

625   indicates that the base version of GEOS-Chem used here (v12.3.0), which uses

MEGAN$_{PCEEA}$, overestimates isoprene emissions by 15-40%. The MEGAN$_{Canopy}$

configuration reduces isoprene emissions in most locations in the Southeast United

States by ~10%, and locally leads to reductions as large as ~20%, bringing the

model into better agreement with these observational constraints.

630   The MEGAN emissions framework calculates the emissions of other non-isoprene

biogenic species as well, including monoterpenes. The influence of the canopy

surrogate model on monoterpene emissions is shown in Figure 11. The annual total

monoterpene emissions in 2012 from MEGAN$_{Canopy}$ are ~95 Tg C yr^{-1}. These

emissions are shown in Figure 11a and are highest over the densely vegetated

635   regions of the world, in particular the tropics. Similar to isoprene emissions,

monoterpene emissions in the northern latitude forests peak during summer

months. The implementation of the canopy surrogate model reduces global annual

total monoterpene emissions by approximately 20%. The annual average absolute

and relative changes to monoterpene emissions due to the canopy surrogate model

640   (MEGAN$_{Canopy}$ − MEGAN$_{PCEEA}$) are shown in Figures 11b and 11c, respectively.

Simulated monoterpene emissions differ from isoprene emissions in that

monoterpene emissions are more sensitive to temperature, with an additional

influence of a light independent emission factor that varies with leaf temperature

(Guenther et al., 2012). There is a fairly constant 20-30% decrease across regions

645 with lower LAI values, including the African savannahs and the Indian subcontinent.

The highest absolute changes are in transitional areas near high LAI forests, with

warmer temperatures (the tropics and subtropics). The high LAI areas of the

tropical and northern forests show smaller decreases of ~5%. These changes, while

substantial, are well within the stated uncertainty ranges in monoterpene emissions

650 of the MEGAN model (300-400%, Guenther et al., 2012).

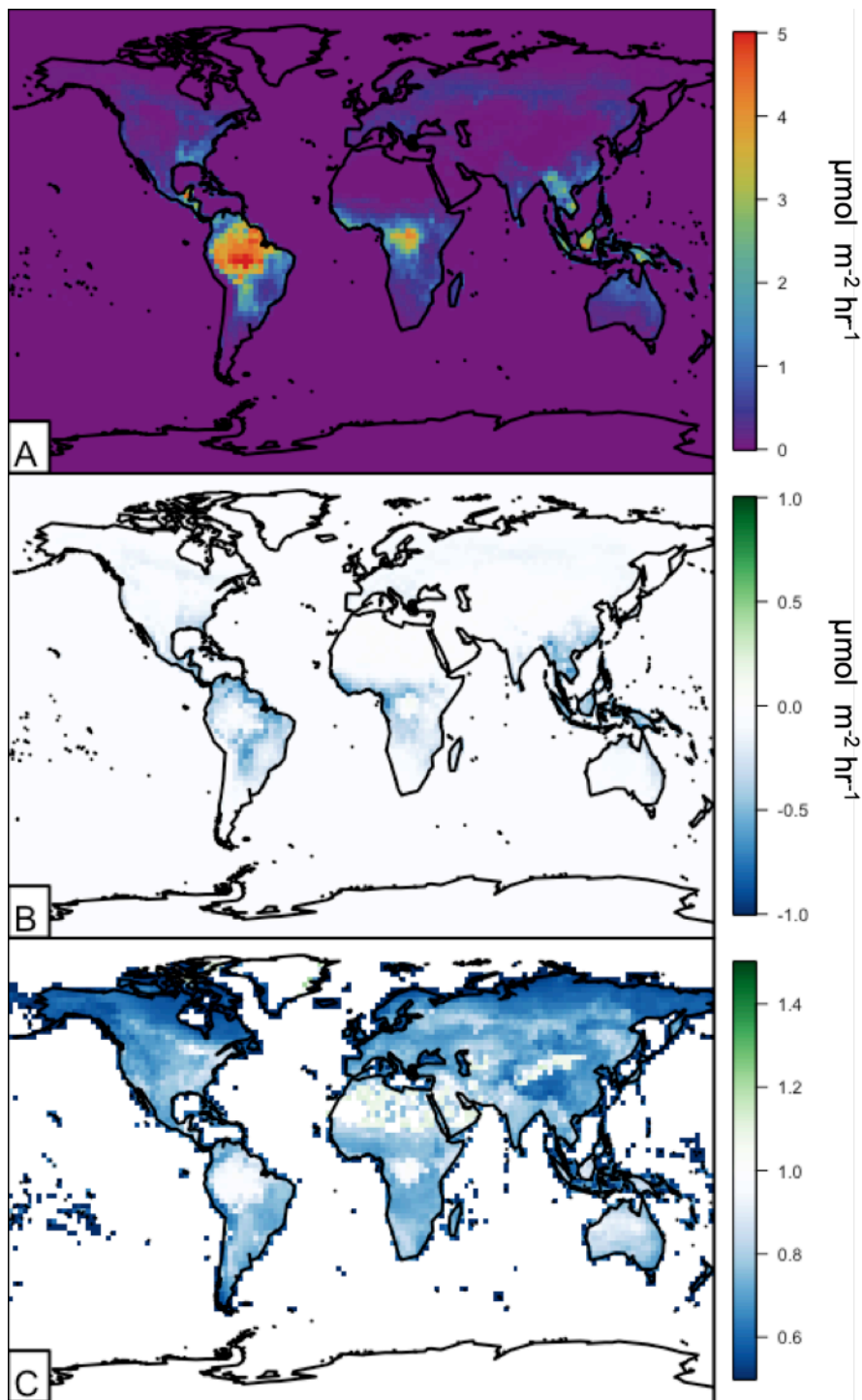



Figure 11. Annual average (2012) monoterpene emissions simulated in GEOS-Chem driven by the surrogate model canopy physics. Panel A shows the annual average emissions. Panel B shows the difference between the surrogate model and base version of simulated emissions. Panel C shows the relative difference between the surrogate model and base version of simulated emissions (surrogate/base model).

655

Changes in simulated ozone dry deposition velocities in 2012 are summarized in Figure 12. Figure 12a shows the annual average spatial distribution of ozone dry deposition velocities. The values vary from less than 0.1 cm s^{-1} over the global oceans, to above 0.5 cm s^{-1} in densely vegetated regions like the tropical rainforests.

The impact of the updated canopy model on ozone dry deposition velocities is in general quite small, with an average change of near zero (~0.004 cm s^{-1}). The annual average relative change is shown in Figure 12b, and the absolute difference in Figure 12c, both in relation to the base version of GEOS-Chem v12.3.0. These changes are nearly all within ±5%, or ±0.01 cm s^{-1}, with a maximum change of 15%, (0.04 cm s^{-1}). Relative changes track most strongly with broadleaf and coniferous forested areas. This is consistent with those regions being most sensitive to stomatal deposition (Silva and Heald, 2018), as the canopy scheme implemented here changes dry deposition only through the calculation of the stomatal resistance term.

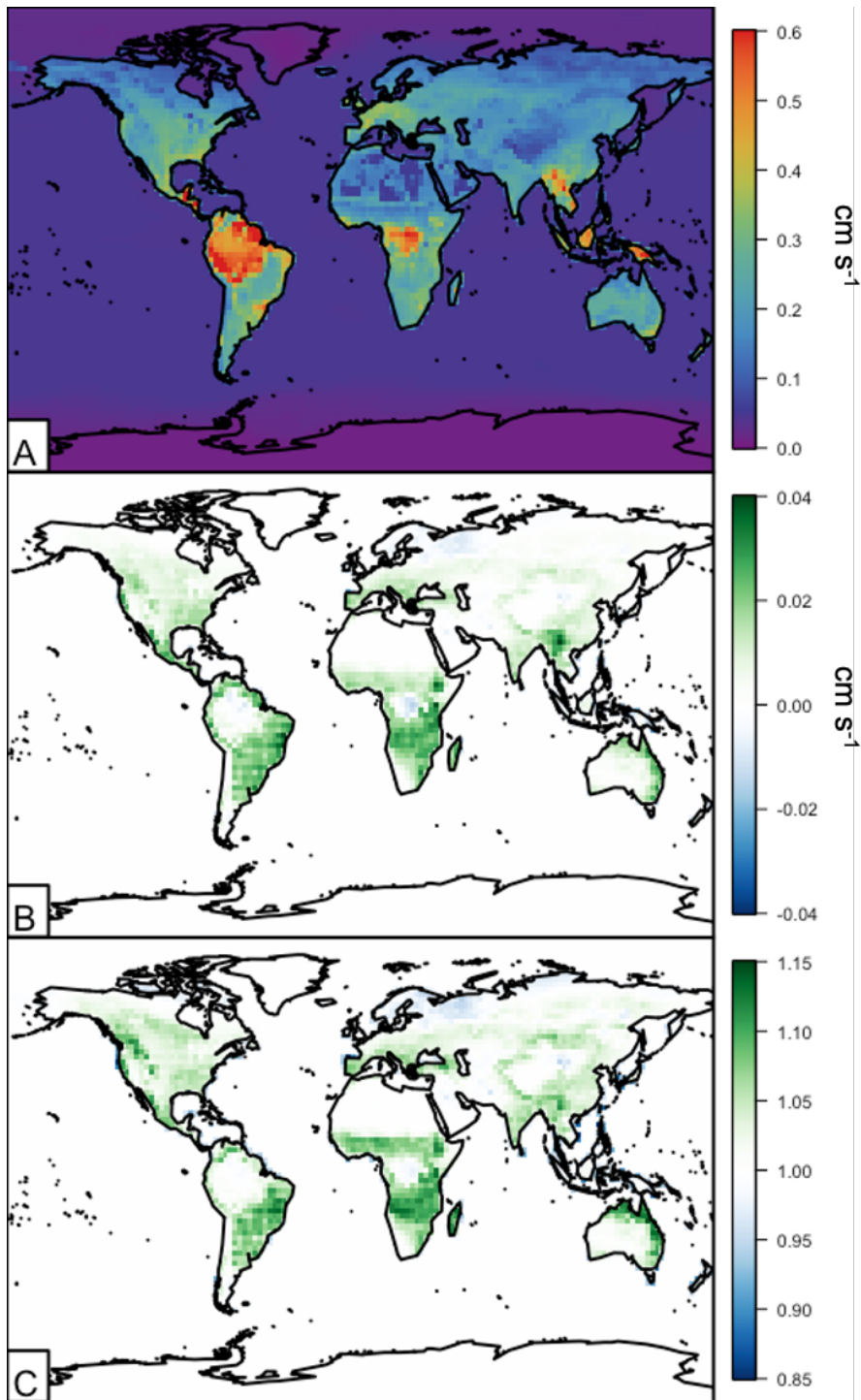



Figure 12. Annual average (2012) ozone dry deposition velocities simulated in GEOS-Chem when driven by the surrogate model canopy physics. Panel A shows the annual average dry deposition velocities (cm s⁻¹). Panel B shows the difference between the surrogate model and base version of simulated dry deposition velocities (cm s⁻¹). Panel C shows the relative difference between the surrogate model and base version of simulated dry deposition velocities (surrogate/base model).

675

The small overall changes to surface-atmosphere exchange processes associated with the updated canopy scheme produce only a modest impact on simulated atmospheric composition. We describe the changes to surface ozone here, as an

680   illustrative example.

The annual average spatial difference in surface ozone between a simulation using the canopy physics described here and the base version of GEOS-Chem is shown in Figure 13. These changes are all generally quite small; all are within 10% of the base simulated annual averages. The changes are generally within ±1 ppbv, with the

685   largest absolute changes over regions with the largest changes in isoprene emissions. The distribution of differences largely follows well-known NO_x-VOC ozone formation patterns. The NO_x limited regions of the world, in particular the remote tropics, show an inverse relationship with isoprene emissions. This is consistent with ozone titration by isoprene in the presence of low NO_x. The largest

690   changes in ozone over the VOC limited regimes of India and China correspond directly to the changes in isoprene emissions changes, with enhanced isoprene emissions over the Indian subcontinent increasing ozone concentrations, and the decrease in isoprene emissions over China leading to a decline in ozone. The overall influence of the changes in ozone dry deposition velocity is fairly negligible. Even

695   the regions where the dry deposition velocity change is the largest (e.g. the Amazon) are dominated by the shift in isoprene emissions.

In total, the changes in surface ozone concentrations slightly ameliorate known biases. There is a persistent high bias (~10 ppbv) across chemical transport models

in simulating surface ozone concentrations over the continental mid-latitudes
700 (Travis et al., 2016). The addition of the new canopy physics parameterization very
modestly reduces this bias in GEOS-Chem (approximately 8 ppbv) by about 1 ppbv,
driving simulated ozone closer to observations.

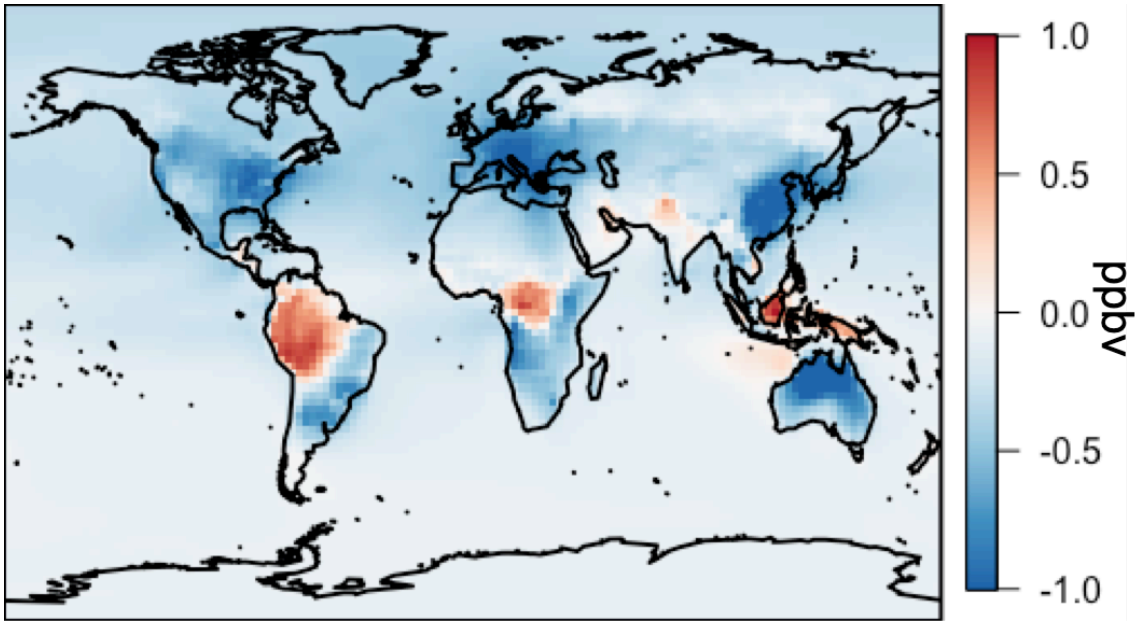



**Figure 13. Annual spatial average surface ozone difference (ppbv) between the updated model version
705 with surrogate canopy physics and the base version of GEOS-Chem (surrogate-base).**

6. Implementation of MEGAN3 Emission Factors

In addition to improved process representation, the canopy surrogate model
presented here allows for the direct application of new emission factors generated
using the MEGAN3 Emission Factor Processor (https://bai.ess.uci.edu/megan),
710 which allows users to generate emission factors from various input datasets. While
the focus of this work is on the impact of representing canopy physics, we include
here a description of this full implementation of MEGAN3 in the GEOS-Chem model
for completeness. We calculate landscape-average emission factor distributions

using the global growth form and ecotype distributions, the emission type

715   speciation, and the leaf level emission factor database available from the MEGAN3

Emission Factor Preprocessor. All MEGAN3 Emissions Factor Preprocessor options

are kept at their default values (i.e. confidence rating J = 0, and 20 total species

classes). The landcover and emissions data are the same as that used for MEGAN2.1

except that the landcover updates described by Yu et al. 2017 were used for the

720   contiguous US. The updated landcover is based on high resolution (30-m) PFT and

detailed vegetation types and is expected to more accurately represent the

landcover distributions in this region. The spatial distribution of the MEGAN3

activity factors is shown in Figure 14. It is important to note that these new

emission factors are input at the leaf level with units on a per LAI basis, as opposed

725   to the canopy-scale factors used in previous versions of MEGAN (applied earlier in

this manuscript), which makes direct comparisons of emission factor magnitudes

infeasible. This canopy to leaf level change ultimately has the consequence of

removing the need for the normalization factor in the activity factor calculation (see

Section 4.1). When these emission factors are scaled to the same units as in

730   MEGAN2.1 (i.e. the per LAI basis in the MEGAN3 emission factors is accounted for)

using the MODIS LAI product applied in this work, the resulting emission factors are

relatively similar (within ±75%) though the MEGAN3 emission factors are lower

than those used with MEGAN2.1 in GEOS-Chem. Generally, more than half of all

changes are within 1000 μg m^{-2} hr^{-1}, and 90% of all emission factors are within 3000

735   μg m^{-2} hr^{-1}. This comparison is not exact, due to the fact that the MODIS LAI product

used here is different from the input vegetation files used to create the original

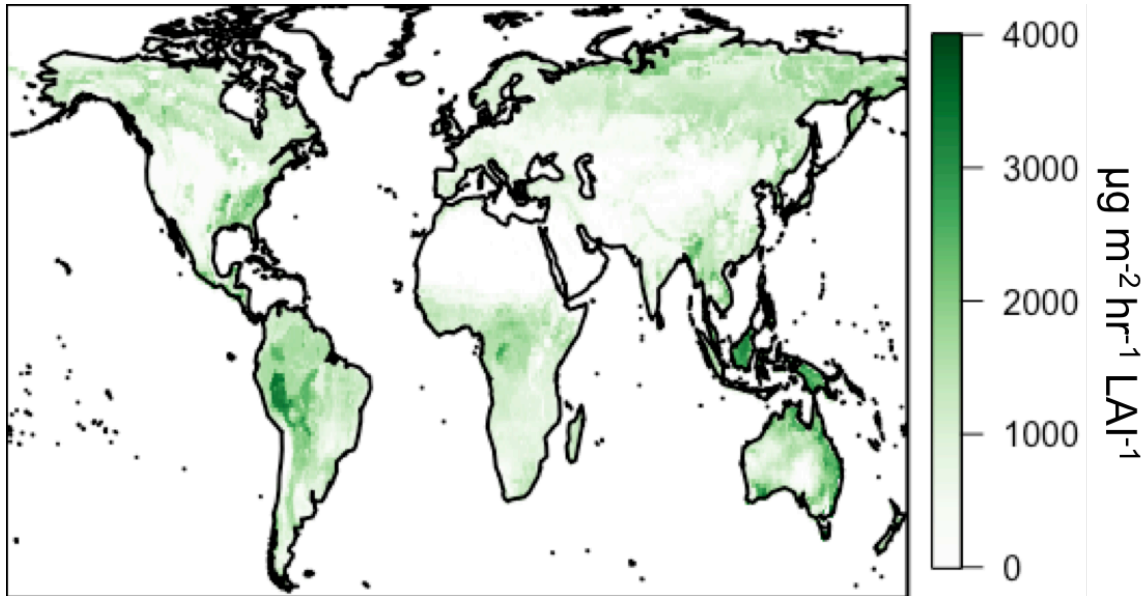
MEGAN2.1 emission factors. Despite the differences in absolute magnitude, the spatial pattern in emission factors in Figure 14 are very similar to those used earlier in this work (Figure 8), with a spatial R^2 of 0.91.

740

Figure 14. MEGAN3 isoprene emission factors (μg m^{-2} hr^{-1} LAI^{-1}).

We implement the MEGAN3 emission factors in GEOS-Chem v12.3.0 using the canopy surrogate model activity factor formulation. For clarity, we refer to "MEGAN3$_{Full}$" as the emissions implementation in GEOS-Chem v12.3.0 using the

745 MEGAN3 leaf-level emission factors and MEGAN3 activity factors with canopy physics calculated following the canopy surrogate model described in Section 3. The annual isoprene emissions simulated using MEGAN3$_{Full}$ are higher than using the MEGAN2.1 canopy-scale factors in GEOS-Chem (as in both MEGAN$_{Canopy}$ and MEGAN$_{PCEEA}$), but more in line with previous work (Guenther et al. 2012).

750 Specifically, annual total isoprene emissions for 2012 are ~570 Tg C yr^{-1} in MEGAN3$_{Full}$ which is a factor of 1.6 larger than those configurations discussed

earlier in this manuscript. The largest contribution to these differences is not the differences in emission factor maps, but is instead the removal of the normalization factor of 0.21, which additionally removes the need for the somewhat arbitrary

755 choice of "standard conditions" for emissions (see Section 4.1). This 570 Tg yr^{-1} emissions total is much more similar to the magnitude of global emissions from versions of MEGAN2.1 implemented outside of the GEOS-Chem model (535-578 Tg yr^{-1}) given by Guenther et al. (2012), and within the stated uncertainty range for MEGAN isoprene emissions (Guenther et al., 2012). These annual average isoprene

760 emissions using MEGAN3$_{Full}$ are shown in Figure 15 below. In general the spatial pattern in the emissions in Figure 15 matches those from the MEGAN$_{Canopy}$ configuration (Figure 9), with an R^2 of ~0.8.

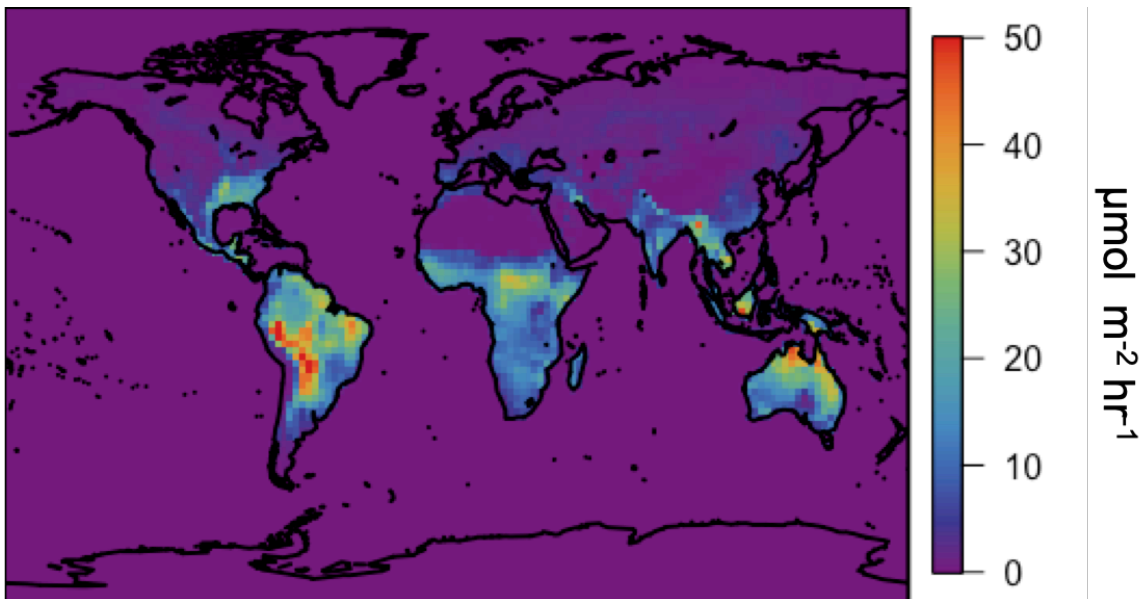



Figure 15. Annual average (2012) isoprene emissions simulated in GEOS-Chem driven by the surrogate
765 **model canopy physics and the MEGAN3 emission and activity factors.**

Since the isoprene emissions calculated using the MEGAN3$_{Full}$ algorithm are so much larger than those used in previous versions of GEOS-Chem, they alter the

composition of the atmosphere significantly. For example, annual average surface

ozone concentrations in the Southeast U.S. increase by nearly 5 ppbv relative to the

770   base version of GEOS-Chem v12.3.0 (which uses MEGAN$_{PCEEA}$), exacerbating the

existing model bias further (Travis et al., 2016). However, MEGAN3$_{Full}$ represents a

more up-to-date and physical characterization of biogenic emissions. Future work

reconciling the differences between these bottom up isoprene emissions estimates

and top down constraints from measurements of composition (e.g. Kaiser et al.,

775   2018) is needed.

7. Conclusions

We describe a novel method for simulating canopy physics relevant to atmospheric

chemistry at very low computational cost. Our surrogate canopy model is based on

the detailed canopy model in the MEGAN3 codebase, and simplified using a

780   statistical learning technique for the determination of variable importance. This

updated scheme allows for improved physical process representation of biosphere-

atmosphere interactions, including a full implementation of the MEGAN3 emissions

scheme activity factors and a more direct implementation of the light and LAI

dependence of dry deposition.

785   When implemented into a chemical transport model, this canopy scheme impacts

the spatial distribution of isoprene emissions, but maintains the global total to

within 5%. Consistent with prior work (Kaiser et al., 2018), isoprene emissions are

reduced over the Southeast United States, with local absolute changes that can

exceed 30%. This difference in surface-atmosphere exchange ultimately has a

790   modest impact on surface ozone, with absolute annual average changes generally

less than 1ppbv, though it does drive ozone concentrations closer to observed

values. The surrogate model additionally allows for integrating new leaf level

emission factor maps into GEOS-Chem, which we show leads to substantial changes

in biogenic emissions.

795   In a rapidly changing earth system, it is critical to represent chemical, biological, and

physical processes with as high fidelity as possible. Surrogate models that allow for

rapid implementation of computationally expensive processes can play a key role in

representing these processes. The work presented in this manuscript represents a

step toward further explicit descriptions of biosphere-atmosphere interactions in

800   models of atmospheric chemistry. Future work should include more detailed

observational constraints and characterization of in-canopy chemical reactions,

turbulent exchange, and biological processes, and their resulting impacts on

abundances of trace gases in the atmosphere.

Author Contributions

805   CLH and SJS designed the study. SJS developed and implemented the surrogate

model and performed simulations and analysis. ABG developed MEGAN3 model

code. All authors contributed to the manuscript preparation.

Code Availablity

The MEGAN3 and GEOS-Chem model code are available at:

810   https://bai.ess.uci.edu/megan/data-and-code, and

Acknowledgements

815   This study was supported by the U.S. National Science Foundation (AGS 1564495)

and by NASA Headquarters under the NASA Earth and Space Science Fellowship

Program (NNX16AN92H).

Competing Interests

The authors declare no competing interests.

820   **References**

Ashworth, K., Chung, S. H., Griffin, R. J., Chen, J., Forkel, R., Bryan, A. M. and Steiner, A. L.: FORest Canopy Atmosphere Transfer (FORCAsT) 1.0: a 1-D model of biosphere–atmosphere chemical exchange, Geoscientific Model Development, 8(11), 3765–3784, doi:https://doi.org/10.5194/gmd-8-3765-2015, 2015.

825   Ashworth, K., Chung, S. H., McKinney, K. A., Liu, Y., Munger, J. W., Martin, S. T. and Steiner, A. L.: Modelling bidirectional fluxes of methanol and acetaldehyde with the FORCAsT canopy exchange model, Atmos. Chem. Phys., 16(24), 15461–15484, doi:10.5194/acp-16-15461-2016, 2016.

Baldocchi, D. D., Hicks, B. B. and Camara, P.: A canopy stomatal resistance model for 830   gaseous deposition to vegetated surfaces, Atmospheric Environment (1967), 21(1), 91–101, doi:10.1016/0004-6981(87)90274-5, 1987.

Bey, I., Jacob, D. J., Yantosca, R. M., Logan, J. A., Field, B. D., Fiore, A. M., Li, Q., Liu, H. Y., Mickley, L. J. and Schultz, M. G.: Global modeling of tropospheric chemistry with assimilated meteorology: Model description and evaluation, J. Geophys. Res., 835   106(D19), 23073–23095, doi:10.1029/2001JD000807, 2001.

Chen W. H., Guenther A. B., Wang X. M., Chen Y. H., Gu D. S., Chang M., Zhou S. Z., Wu L. L. and Zhang Y. Q.: Regional to Global Biogenic Isoprene Emission Responses to

Changes in Vegetation From 2000 to 2015, Journal of Geophysical Research: Atmospheres, 123(7), 3757–3771, doi:10.1002/2017JD027934, 2018.

840   Committee on the Future of Atmospheric Chemistry Research, Board on Atmospheric Sciences and Climate, Division on Earth and Life Studies and National Academies of Sciences, Engineering, and Medicine: The Future of Atmospheric Chemistry Research: Remembering Yesterday, Understanding Today, Anticipating Tomorrow, National Academies Press, Washington, D.C., 2016.

845   Engel, R. K., Moser, L. E., Stubbendieck, J. and Lowry, S. R.: Yield Accumulation, Leaf Area Index, and Light Interception of Smooth Bromegrass1, Crop Science, 27(2), 316, doi:10.2135/cropsci1987.0011183X002700020039x, 1987.

Geddes, J. A., Heald, C. L., Silva, S. J. and Martin, R. V.: Land cover change impacts on atmospheric chemistry: simulating projected large-scale tree mortality in the United
850   States, Atmos. Chem. Phys., 16(4), 2323–2340, doi:10.5194/acp-16-2323-2016, 2016.

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C. A., Darmenov, A., Bosilovich, M. G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty, A., da Silva, A. M., Gu, W., Kim, G.-K., Koster, R.,
855   Lucchesi, R., Merkova, D., Nielsen, J. E., Partyka, G., Pawson, S., Putman, W., Rienecker, M., Schubert, S. D., Sienkiewicz, M. and Zhao, B.: The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2), J. Climate, 30(14), 5419–5454, doi:10.1175/JCLI-D-16-0758.1, 2017.

Geron, C., Daly, R., Harley, P., Rasmussen, R., Seco, R., Guenther, A., Karl, T. and Gu, L.:
860   Large drought-induced variations in oak leaf volatile organic compound emissions during PINOT NOIR 2012, Chemosphere, 146, 8–21, doi:10.1016/j.chemosphere.2015.11.086, 2016.

Giglio, L., Randerson, J. T. and van der Werf, G. R.: Analysis of daily, monthly, and annual burned area using the fourth-generation global fire emissions database
865   (GFED4), J. Geophys. Res. Biogeosci., 118(1), 317–328, doi:10.1002/jgrg.20042, 2013.

Goldstein, A. H., McKay, M., Kurpius, M. R., Schade, G. W., Lee, A., Holzinger, R. and Rasmussen, R. A.: Forest thinning experiment confirms ozone deposition to forest canopy is dominated by reaction with biogenic VOCs, Geophysical Research Letters,
870   31(22), doi:10.1029/2004GL021259, 2004.

Goudriaan, J. and Laar, H. H. van: Modelling potential crop growth processes: textbook with exercises, Kluwer, Dordrecht., 1994.

Goudriaan, J. and Monteith, J. L.: A Mathematical Function for Crop Growth Based on Light Interception and Leaf Area Expansion, Ann Bot, 66(6), 695–701, 1990.

875   Guenther, A., Baugh, B., Brasseur, G., Greenberg, J., Harley, P., Klinger, L., Serça, D. and Vierling, L.: Isoprene emission estimates and uncertainties for the central African EXPRESSO study domain, J. Geophys. Res., 104(D23), 30625–30639, doi:10.1029/1999JD900391, 1999.

Guenther, A., Karl, T., Harley, P., Wiedinmyer, C., Palmer, P. I. and Geron, C.:
880   Estimates of global terrestrial isoprene emissions using MEGAN (Model of Emissions of Gases and Aerosols from Nature), Atmos. Chem. Phys., 6(11), 3181–3210, doi:10.5194/acp-6-3181-2006, 2006.

Guenther, A. B., Jiang, X., Heald, C. L., Sakulyanontvittaya, T., Duhl, T., Emmons, L. K. and Wang, X.: The Model of Emissions of Gases and Aerosols from Nature version
885   2.1 (MEGAN2.1): an extended and updated framework for modeling biogenic emissions, Geosci. Model Dev., 5(6), 1471–1492, doi:10.5194/gmd-5-1471-2012, 2012.

Harley, P., Guenther, A. and Zimmerman, P.: Effects of light, temperature and canopy position on net photosynthesis and isoprene emission from sweetgum
890   (Liquidambar styraciflua) leaves, Tree Physiology, 16(1–2), 25–32, doi:10.1093/treephys/16.1-2.25, 1996.

Hastie, T., Tibshirani, R. and Friedman, J.: The Elements of Statistical Learning, Springer New York Inc., New York, NY, USA., 2001.

Hoesly, R. M., Smith, S. J., Feng, L., Klimont, Z., Janssens-Maenhout, G., Pitkanen, T.,
895   Seibert, J. J., Vu, L., Andres, R. J., Bolt, R. M., Bond, T. C., Dawidowski, L., Kholod, N., Kurokawa, J.-I., Li, M., Liu, L., Lu, Z., Moura, M. C. P., O'Rourke, P. R. and Zhang, Q.: Historical (1750–2014) anthropogenic emissions of reactive gases and aerosols from the Community Emissions Data System (CEDS), Geosci. Model Dev., 11(1), 369–408, doi:10.5194/gmd-11-369-2018, 2018.

900   Hudman, R. C., Moore, N. E., Mebust, A. K., Martin, R. V., Russell, A. R., Valin, L. C. and Cohen, R. C.: Steps towards a mechanistic model of global soil nitric oxide emissions: implementation and space based-constraints, Atmos. Chem. Phys., 12(16), 7779–7795, doi:10.5194/acp-12-7779-2012, 2012.

IPCC: Climate Change 2013: The Physical Science Basis. Contribution of Working
905   Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA., 2013.

Kaiser, J., Jacob, D. J., Zhu, L., Travis, K. R., Fisher, J. A., González Abad, G., Zhang, L., Zhang, X., Fried, A., Crounse, J. D., St. Clair, J. M. and Wisthaler, A.: High-resolution
910   inversion of OMI formaldehyde columns to quantify isoprene emission on ecosystem-relevant scales: application to the southeast US, Atmos. Chem. Phys., 18(8), 5483–5497, doi:10.5194/acp-18-5483-2018, 2018.

Keenan, T. F., Grote, R. and Sabaté, S.: Overlooking the canopy: The importance of canopy structure in scaling isoprenoid emissions from the leaf to the landscape, Ecological Modelling, 222(3), 737–747, doi:10.1016/j.ecolmodel.2010.11.004, 2011.

Lamb, B., Pierce, T., Baldocchi, D., Allwine, E., Dilts, S., Westberg, H., Geron, C., Guenther, A., Klinger, L., Harley, P. and Zimmerman, P.: Evaluation of forest canopy models for estimating isoprene emissions, J. Geophys. Res., 101(D17), 22787–22797, doi:10.1029/96JD00056, 1996.

Lelieveld, J. and Dentener, F. J.: What controls tropospheric ozone?, J. Geophys. Res., 105(D3), 3531–3551, doi:10.1029/1999JD901011, 2000.

Leuning, R., Kelliher, F. M., Pury, D. G. G. D. and Schulze, E.-D.: Leaf nitrogen, photosynthesis, conductance and transpiration: scaling from leaves to canopies, Plant, Cell & Environment, 18(10), 1183–1200, doi:10.1111/j.1365-3040.1995.tb00628.x, 1995.

Li, M., Zhang, Q., Kurokawa, J.-I., Woo, J.-H., He, K., Lu, Z., Ohara, T., Song, Y., Streets, D. G., Carmichael, G. R., Cheng, Y., Hong, C., Huo, H., Jiang, X., Kang, S., Liu, F., Su, H. and Zheng, B.: MIX: a mosaic Asian anthropogenic emission inventory under the international collaboration framework of the MICS-Asia and HTAP, Atmos. Chem. Phys., 17(2), 935–963, doi:10.5194/acp-17-935-2017, 2017.

Makar, P. A., Fuentes, J. D., Wang, D., Staebler, R. M. and Wiebe, H. A.: Chemical processing of biogenic hydrocarbons within and above a temperate deciduous forest, Journal of Geophysical Research: Atmospheres, 104(D3), 3581–3603, doi:10.1029/1998JD100065, 1999.

Makar, P. A., Staebler, R. M., Akingunola, A., Zhang, J., McLinden, C., Kharol, S. K., Pabla, B., Cheung, P. and Zheng, Q.: The effects of forest canopy shading and turbulence on boundary layer ozone, Nature Communications, 8, 15243, doi:10.1038/ncomms15243, 2017.

Mao, J., Paulot, F., Jacob, D. J., Cohen, R. C., Crounse, J. D., Wennberg, P. O., Keller, C. A., Hudman, R. C., Barkley, M. P. and Horowitz, L. W.: Ozone and organic nitrates over the eastern United States: Sensitivity to isoprene chemistry, J. Geophys. Res. Atmos., 118(19), 2013JD020231, doi:10.1002/jgrd.50817, 2013.

Marais, E. A. and Wiedinmyer, C.: Air Quality Impact of Diffuse and Inefficient Combustion Emissions in Africa (DICE-Africa), Environ. Sci. Technol., 50(19), 10739–10745, doi:10.1021/acs.est.6b02602, 2016.

Millet, D. B., Guenther, A., Siegel, D. A., Nelson, N. B., Singh, H. B., Gouw, J. A. de, Warneke, C., Williams, J., Eerdekens, G., Sinha, V., Karl, T., Flocke, F., Apel, E., Riemer, D. D., Palmer, P. I. and Barkley, M.: Global atmospheric budget of acetaldehyde: 3-D model analysis and constraints from in-situ and satellite observations, Atmospheric

950   Chemistry and Physics, 10(7), 3405–3425, doi:https://doi.org/10.5194/acp-10-3405-2010, 2010.

Myneni, R. B., Hoffman, S., Knyazikhin, Y., Privette, J. L., Glassy, J., Tian, Y., Wang, Y., Song, X., Zhang, Y., Smith, G. R., Lotsch, A., Friedl, M., Morisette, J. T., Votava, P., Nemani, R. R. and Running, S. W.: Global products of vegetation leaf area and
955   fraction absorbed PAR from year one of MODIS data, Remote Sensing of Environment, 83(1–2), 214–231, doi:10.1016/S0034-4257(02)00074-3, 2002.

Myneni, R. B., Yang, W., Nemani, R. R., Huete, A. R., Dickinson, R. E., Knyazikhin, Y., Didan, K., Fu, R., Negron Juarez, R. I., Saatchi, S. S., Hashimoto, H., Ichii, K., Shabanov, N. V., Tan, B., Ratana, P., Privette, J. L., Morisette, J. T., Vermote, E. F., Roy, D. P., Wolfe,
960   R. E., Friedl, M. A., Running, S. W., Votava, P., El-Saleous, N., Devadiga, S., Su, Y. and Salomonson, V. V.: Large seasonal swings in leaf area of Amazon rainforests, Proceedings of the National Academy of Sciences, 104(12), 4820–4823, doi:10.1073/pnas.0611338104, 2007.

Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V. N.,
965   Underwood, E. C., D'amico, J. A., Itoua, I., Strand, H. E., Morrison, J. C., Loucks, C. J., Allnutt, T. F., Ricketts, T. H., Kura, Y., Lamoreux, J. F., Wettengel, W. W., Hedao, P. and Kassem, K. R.: Terrestrial Ecoregions of the World: A New Map of Life on EarthA new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity, BioScience, 51(11), 933–938, doi:10.1641/0006-
970   3568(2001)051[0933:TEOTWA]2.0.CO;2, 2001.

Safieddine, S. A., Heald, C. L. and Henderson, B. H.: The Global Non-Methane Reactive Organic Carbon Budget: A Modeling Perspective, Geophys. Res. Lett., 2017GL072602, doi:10.1002/2017GL072602, 2017.

Silva, S. J. and Heald, C. L.: Investigating Dry Deposition of Ozone to Vegetation, J.
975   Geophys. Res. Atmos., 2017JD027278, doi:10.1002/2017JD027278, 2018.

Silva, S. J., Heald, C. L., Geddes, J. A., Austin, K. G., Kasibhatla, P. S. and Marlier, M. E.: Impacts of current and projected oil palm plantation expansion on air quality over Southeast Asia, Atmos. Chem. Phys., 16(16), 10621–10635, doi:10.5194/acp-16-10621-2016, 2016.

980   Spitters, C. J. T.: Separating the diffuse and direct component of global radiation and its implications for modeling canopy photosynthesis Part II. Calculation of canopy photosynthesis, Agricultural and Forest Meteorology, 38(1), 231–242, doi:10.1016/0168-1923(86)90061-4, 1986.

Travis, K. R., Jacob, D. J., Fisher, J. A., Kim, P. S., Marais, E. A., Zhu, L., Yu, K., Miller, C.
985   C., Yantosca, R. M., Sulprizio, M. P., Thompson, A. M., Wennberg, P. O., Crounse, J. D., St. Clair, J. M., Cohen, R. C., Laughner, J. L., Dibb, J. E., Hall, S. R., Ullmann, K., Wolfe, G. M., Pollack, I. B., Peischl, J., Neuman, J. A. and Zhou, X.: Why do models overestimate

surface ozone in the Southeast United States?, Atmos. Chem. Phys., 16(21), 13561–13577, doi:10.5194/acp-16-13561-2016, 2016.

990   Unger, N.: Human land-use-driven reduction of forest volatiles cools global climate, Nature Clim. Change, 4(10), 907–910, doi:10.1038/nclimate2347, 2014.

Wang, Y., Jacob, D. J. and Logan, J. A.: Global simulation of tropospheric O_3-NO_x-hydrocarbon chemistry: 1. Model formulation, Journal of Geophysical Research: Atmospheres, 103(D9), 10713–10725, doi:10.1029/98JD00158, 1998.

995   Wesely, M. L.: Parameterization of surface resistances to gaseous dry deposition in regional-scale numerical models, Atmospheric Environment (1967), 23(6), 1293–1304, doi:10.1016/0004-6981(89)90153-4, 1989.

Yu, K., Jacob, D. J., Fisher, J. A., Kim, P. S., Marais, E. A., Miller, C. C., Travis, K. R., Zhu, L., Yantosca, R. M., Sulprizio, M. P., Cohen, R. C., Dibb, J. E., Fried, A., Mikoviny, T.,
1000   Ryerson, T. B., Wennberg, P. O. and Wisthaler, A.: Sensitivity to grid resolution in the ability of a chemical transport model to simulate observed oxidant chemistry under high-isoprene conditions, Atmos. Chem. Phys., 16(7), 4369–4378, doi:10.5194/acp-16-4369-2016, 2016.

Zhang, L., Brook, J. R. and Vet, R.: A revised parameterization for gaseous dry
1005   deposition in air-quality models, Atmos. Chem. Phys., 3(6), 2067–2082, doi:10.5194/acp-3-2067-2003, 2003.