



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. Heald¹, Alex B. Guenther³

5 ¹Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

²Now at: Pacific Northwest National Laboratory, Richland, WA, USA

³Department of Earth System Science, University of California Irvine, Irvine, CA, USA

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

25 ~40%, though annual global isoprene emissions remain largely consistent (within 5%). These changes to surface-atmosphere exchange lead to modest changes in



surface ozone concentrations of ± 1 ppbv. The use of this surrogate canopy model drives emissions of isoprene and concentrations of surface ozone closer to observationally constrained values, without any noticeable impact on computational demand. 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
50 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
55 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
60 (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
65 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 through various parameterizations. (e.g. Guenther et al., 2006; Wesely, 1989; Zhang



70 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
parameterizations commonly assume that the temperature of leaves is equal to the
75 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
top of the canopy, to having some sort of reduced complexity multiplicative factor
80 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
canopy approximations can lead to regional differences of greater than 20% relative
85 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
processes related to canopy energy and radiative balance are not represented in
90 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 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

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 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 (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 temperature of 10°C. Leaf-level PAR is computed as a function of incoming radiation



115 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
important to note that differing canopy model choice and parameter selection can
125 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.
H. et al., 2018; Geron et al., 2016; Guenther et al., 2006). Additionally, the MEGAN
130 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
models without the need for substantial model architecture development.

135 **3. Surrogate Model Development**



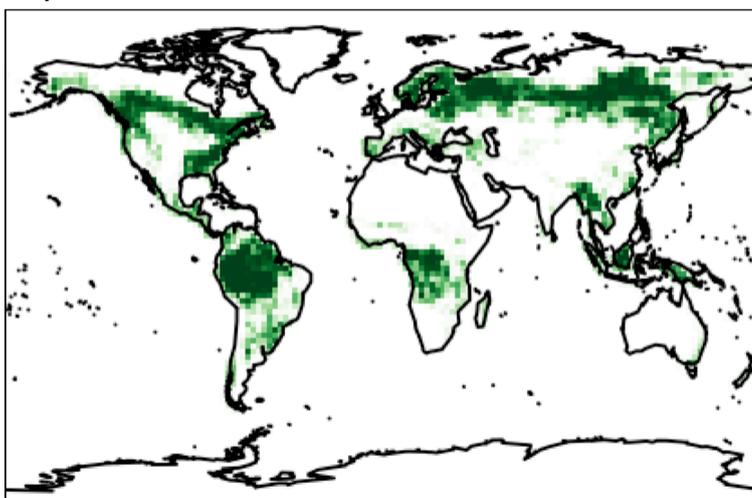
To begin the surrogate model development, we first use a variable selection approach to evaluate and rank which of the suite of model input variables are most important for the simulation of both leaf level PAR and temperature. To do this, we use a machine learning regression method for model simplification and
140 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). While LASSO was originally developed as a complete regression method, we follow
145 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 simulated canopy physics from the MEGAN3 canopy model at the global scale for the
150 year 2012. Input meteorology is from MERRA-2 assimilated meteorological fields at $2^{\circ} \times 2.5^{\circ}$ 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 $2^{\circ} \times 2.5^{\circ}$ horizontal resolution and a monthly timescale. This input data is identical to that
155 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



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

a) Forests



b) Crops, Grasses, and Shrubs

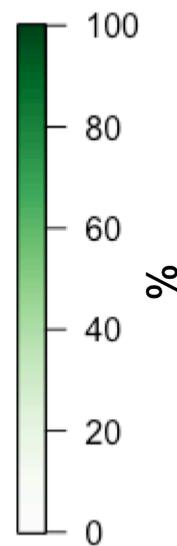
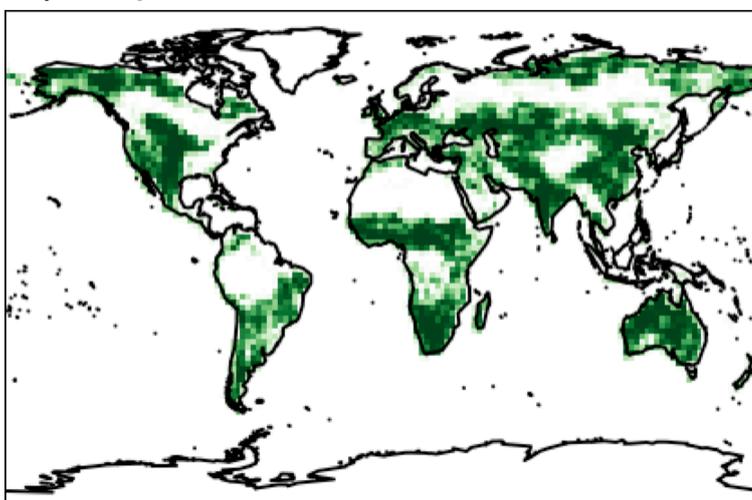




Figure 1. The percent of each $2^\circ \times 2.5^\circ$ gridbox occupied by each vegetation class
165 used in this work. Panel a) is forested vegetation, and panel b) is crops, grasses, and
shrubland.

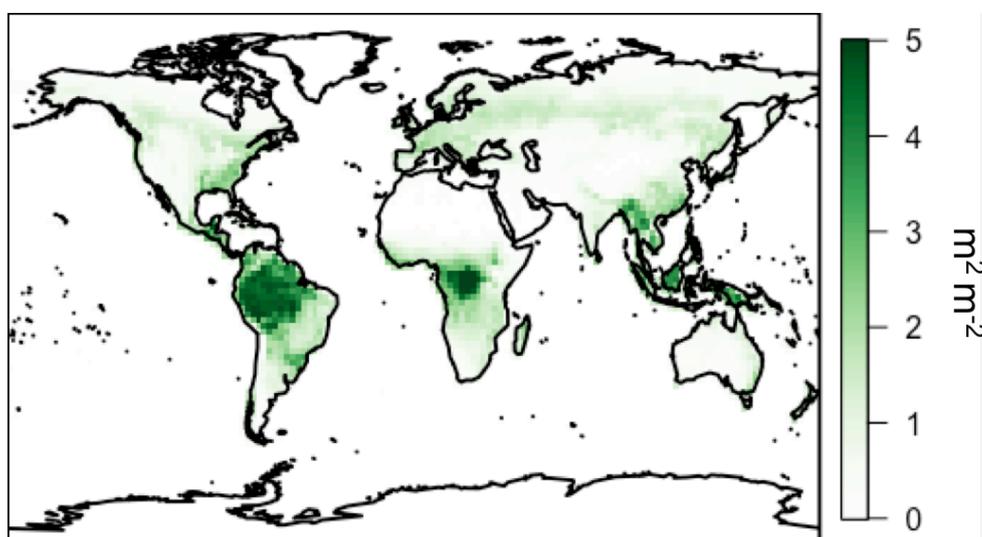


Figure 2. Annual Average LAI from MODIS for the year 2012.

The LASSO importance rankings are remarkably consistent for both sunlit and
170 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
175 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 covary with air temperature, and how the rankings are derived separately for sunlit
180 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.

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

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

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 speed to the regression only improves the performance of the model by less than
190 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 ($\text{PAR}_{i,\text{leaf}}$, $\mu\text{mol m}^{-2} \text{s}^{-1}$), 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):

$$(2) \text{PAR}_{i,\text{leaf}} = \text{PAR}_{\text{toc}} * \exp(C_i + D_i * \text{LAI}).$$

195 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.

We fit equations 1 and 2 for all layers of the canopy and for sunlit and shaded leaves, 200 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 because this more parsimonious model performs quite well (see Section 3.1) 205 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

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

210 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 MEGAN code, (see Guenther et al., 2006, 2012), without any model simplification. The sunlit fraction is calculated as follows:



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

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

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 leaf (the vector [0.05, 0.23, 0.5, 0.77, 0.95]). Consistent with the MEGAN3 parent
220 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, Leaf PAR, and Sunlit Leaf fraction), we are able to implement more physically
225 realistic parameterizations for biosphere-atmosphere interactions.

3.1. Surrogate Model Performance

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

3.1.1 Temperature

The surrogate model simulated annual canopy average leaf temperature
230 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
235 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.

240 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.

245 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
250 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
255 the total magnitude.

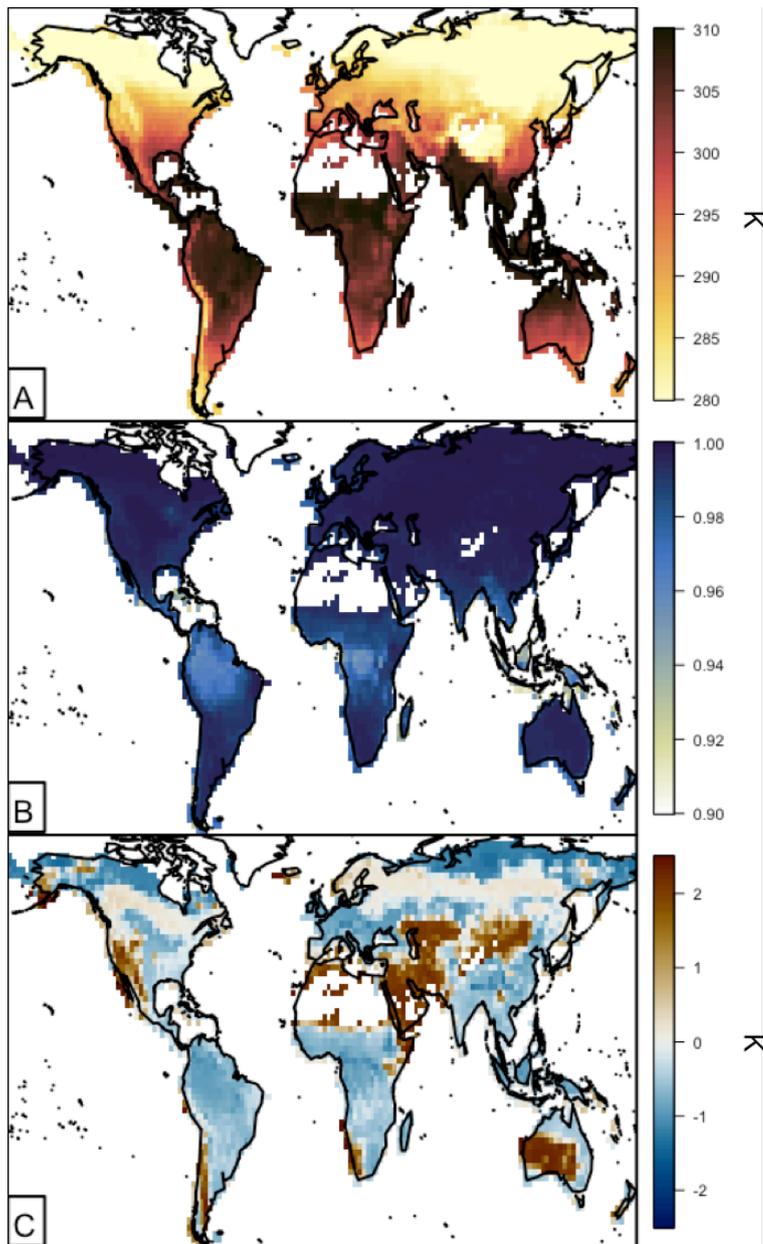


Figure 3. Annual canopy average spatial temperature surrogate model performance for 2012. Panels are as follows: A) Annual average surrogate model leaf level temperature (Kelvin), B) R^2 between the surrogate and the full model leaf level temperature, C) Annual average leaf level temperature bias (surrogate-full model, K)

260

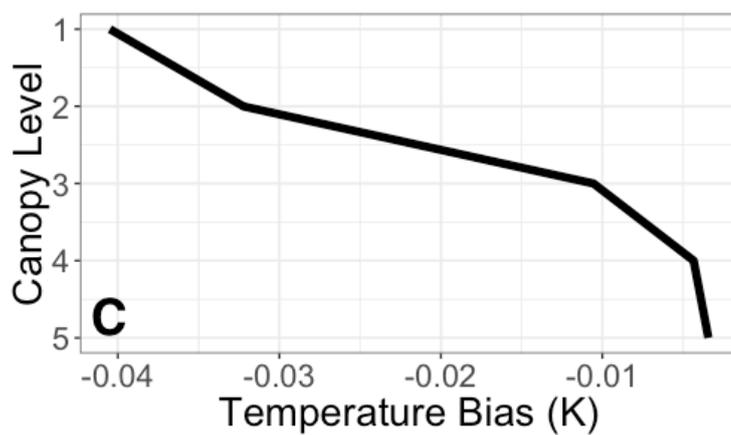
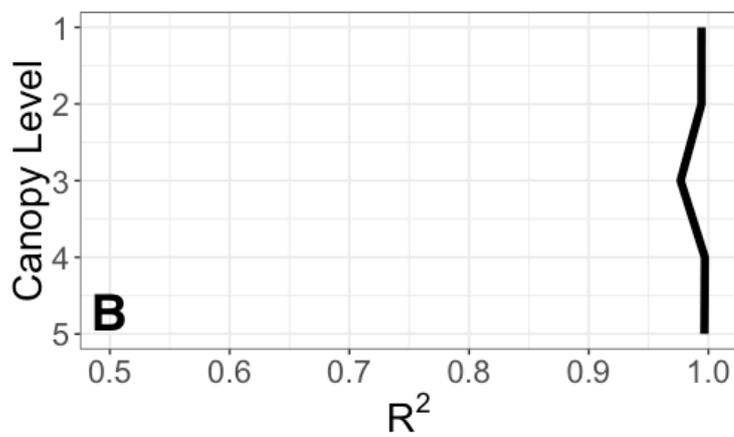
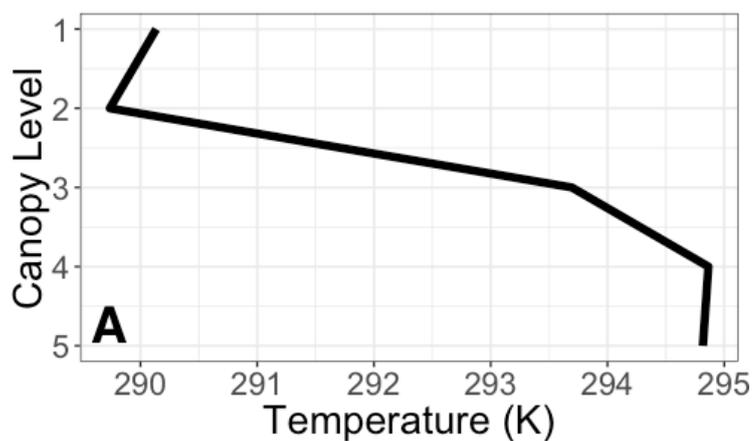


The vertical profile of annual average leaf temperature is shown in Figure 4a. The broad shape of the vertical distribution is consistent everywhere. The upper canopy layers are cooler than the lower canopy layers, as an insulating effect from air temperature occurs within the canopy. The higher order variability (e.g. small differences within 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, LAI, and wind speed. However, the generally consistent behavior of this higher order variability allows for it to be reproduced in the simplified surrogate model.

265

270 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, 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.

275





280 **Figure 4. Vertical profile average temperature performance of the surrogate model for 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

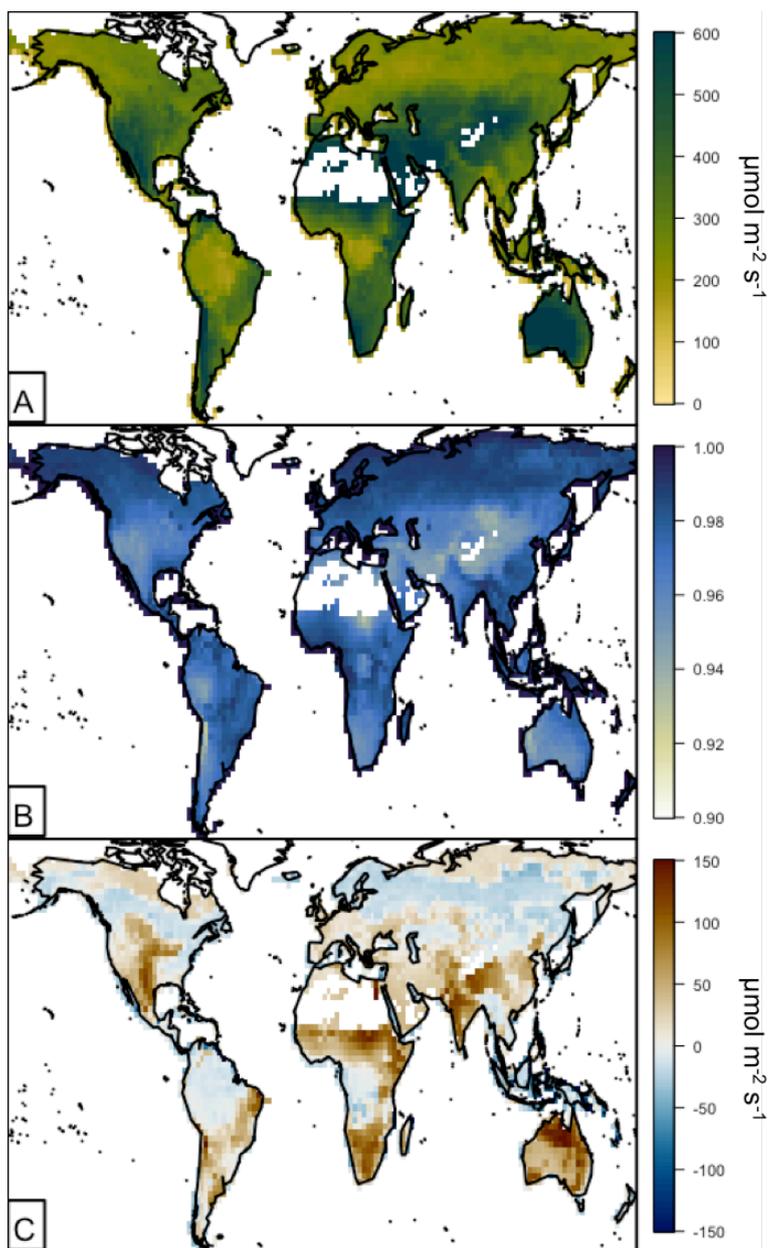
The annual canopy average leaf-level PAR for 2012 is shown spatially in Figure 5a.

285 In Figure 5, canopy temperature averages are calculated using the same method as for canopy temperature. Annual average PAR varies from $\sim 200 \mu\text{mol m}^{-2} \text{s}^{-1}$ to greater than $600 \mu\text{mol m}^{-2} \text{s}^{-1}$. 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
290 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
295 northern boreal forests. The low 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 are low in part due to relatively high LAI, but also due to reduced incoming PAR in the winter months when the solar angle is low.

300 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
305 shading and other canopy physical processes are greatly reduced. The worst model
 R^2 performance is over regions with the highest LAI, where the elevated importance
of shading and resulting complexity in the PAR 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
310 average model biases are generally within $\pm 40 \mu\text{mol m}^{-2} \text{s}^{-1}$, with a few more
extreme values reaching $\pm 200 \mu\text{mol m}^{-2} \text{s}^{-1}$. The surrogate model is broadly biased
high over regions with lower LAI and slightly low over regions with high LAI. In a
relative sense, these changes are nearly all within 10-15%, with a maximum
normalized mean bias of 0.4.



315

Figure 5. Canopy spatial PAR surrogate model performance for 2012. Panels are as follows: A) Annual average surrogate model leaf level PAR ($\mu\text{mol m}^{-2} \text{s}^{-1}$), B) R^2 between the leaf level PAR simulated using the surrogate and the full model, C) Annual average leaf level PAR bias (surrogate-full model, $\mu\text{mol m}^{-2} \text{s}^{-1}$).



320 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 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
325 than 5. The low LAI range represents ~40% of all vegetation throughout the year, the middle range represents nearly 60%, and the high LAI range contains only a small fraction of all vegetation (~1%).

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
330 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 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
335 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 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
340 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



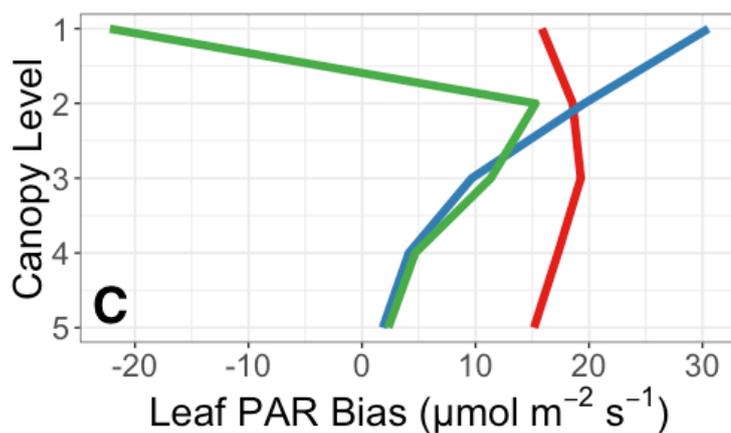
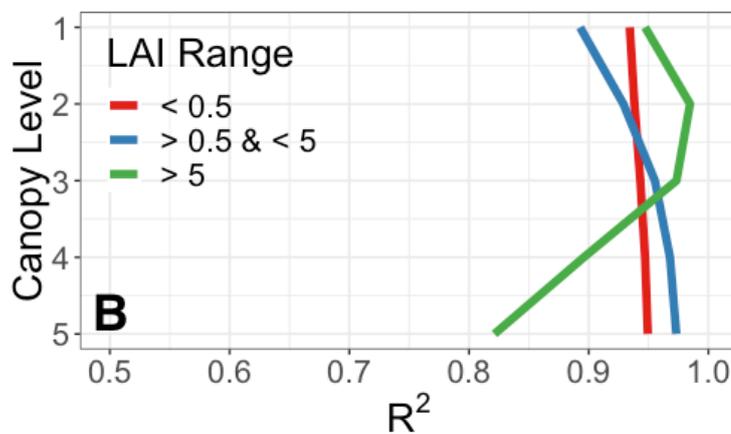
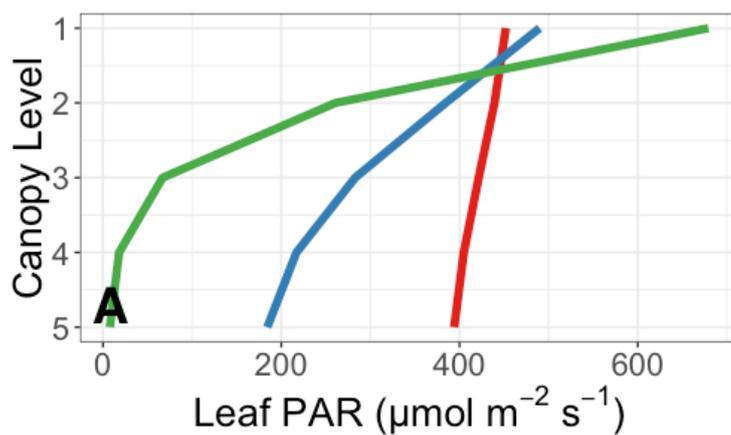
layer may intercept light and shade leaves lower within that same layer. This has the effect of reducing the layer average leaf-level PAR as a function of leaf geometries and LAI, and explains why the highest canopy layer average leaf-level PAR is not the same as the average PAR incident on the top of the canopy.

Figures 6b and 6c summarize the statistical performance of the surrogate model 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). While the surrogate model struggles to capture this lower canopy variability, the ultimate influence on the canopy-scale bias is generally low.

The vertical distribution of that bias is shown in Figure 6c. Broadly, the absolute 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 $\sim 1\%$ 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 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 $50 \mu\text{mol m}^{-2} \text{s}^{-1}$ in the top canopy layer). For both the high and middle LAI ranges,



365 the absolute magnitude of the PAR bias decreases throughout the canopy, and the
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-
370 level PAR, the resulting difference between small numbers leads to a relatively large
normalized mean bias of ~ 0.3 .





375 **Figure 6. Vertical profile average PAR surrogate model performance as a function of LAI for 2012. Panel A shows the vertical average surrogate model leaf-level PAR ($\mu\text{mol m}^{-2} \text{s}^{-1}$) 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 ($\mu\text{mol m}^{-2} \text{s}^{-1}$) 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 ($\gamma_{\text{PAR}} \geq 0.9$) 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 $\sim 5\%$, with a median bias of -2% .

380
385
390

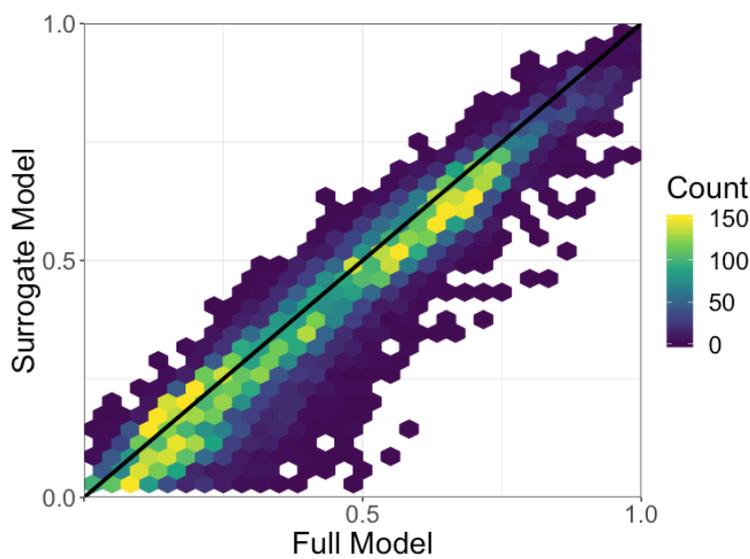


Figure 7. The annual grid box average fraction of light saturated leaves as simulated by the full and surrogate models throughout the canopy for the year 2012. The colorbar represents the number of observations in a given hex. The 1:1 line is shown in black.

395 Ultimately, this assessment demonstrates that the surrogate model reproduces the
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
400 order variability in absolute magnitude of the variables is missing from this
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-



405 chem.org). GEOS-Chem is a computational model for simulating atmospheric
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,
410 respectively. Identically to the canopy model input data, we use LAI values from the
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
415 (CEDS, Hoesly et al., 2018). Regional emissions over the United States, Africa, and
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
420 considering gas-phase chemical impacts.

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
425 “activity factors” representing standard processes that govern biogenic emissions
(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, 430 leaf-level PAR, leaf age, leaf area index, soil moisture, and atmospheric CO₂ 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 435 resolution of 0.25°x0.3125° to match the GEOS-Chem resolution of 2°x2.5°.

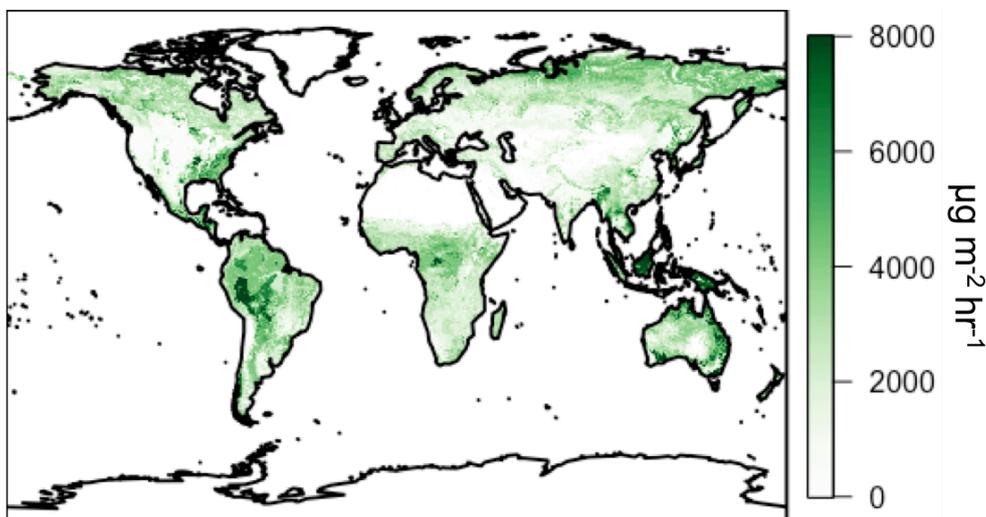


Figure 8. Base Isoprene emission factors used in this work.

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 440 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
445 includes response to the LAI, PAR incident to the top of the canopy, and the solar
zenith angle.

We modify the MEGAN implementation in GEOS-Chem to allow for the
representation of canopy physics described in Section 3. In order to properly scale all
emission factors to the plant canopy using a canopy model, a normalization factor
450 must be applied at a set of standard environmental and ecological conditions
(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 $1500 \mu\text{mol m}^{-2} \text{s}^{-1}$, and a 10%/80%/10% split of growing,
455 mature, and senescent leaves (Guenther et al., 2012; Kaiser et al., 2018). We
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 these
previous standard conditions (e.g. Current Temperature = 298.5K, Current PAR =
460 $1500 \mu\text{mol m}^{-2} \text{s}^{-1}$) are met to within 10%, we calculate the 24-hour average prior
meteorological conditions from the 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 $740 \mu\text{mol m}^{-2} \text{s}^{-1}$, respectively. These standard conditions result in a
465 normalization factor of 0.21 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 small deviations from these standard conditions (± 3 K or $\pm 20 \mu\text{mol m}^{-2} \text{s}^{-1}$) can lead to changes in the
470 normalization factor of nearly 20%. 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 the parameter choices made in this normalization processes.

475 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 of MEGAN in the base version of GEOS-Chem, activity factors are calculated
480 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
485 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}):

490 (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 $\gamma_{PCEE A}$ activity factors as “MEGAN_{PCEE A}” and those using the γ_{Canopy} activity factors as “MEGAN_{Canopy}”. We use the canopy physics surrogate model described in Section 3 to calculate the leaf temperature and PAR in the MEGAN_{Canopy} implementation.

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 CO₂ inhibition) are the same in both the MEGAN_{Canopy} and MEGAN_{PCEE A}.



4.2 Dry Deposition

Dry Deposition in GEOS-Chem v12.3.0 is calculated through a resistor-in-series
510 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
transfer velocity known as the “dry deposition velocity”. A recent assessment of the
dry deposition velocity parameterization in GEOS-Chem found that biases in
515 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).

Prior to this work, canopy effects were not directly considered in GEOS-Chem dry
deposition, and only approximated in bulk using a polynomial decomposition
520 scheme (Wang et al., 1998) that jointly calculated both a multiplicative factor ($1 +$
 $b/\text{PAR}_{\text{leaf}}$, $b = 50 \mu\text{mol m}^{-2} \text{s}^{-1}$) to the stomatal resistance from Baldocchi et al. (1987)
based on leaf-level PAR and a normalization 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,
525 and then explicitly 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 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
530 stomatal resistance by a factor of 0.6.



5. Surrogate Model Integration into GEOS-Chem

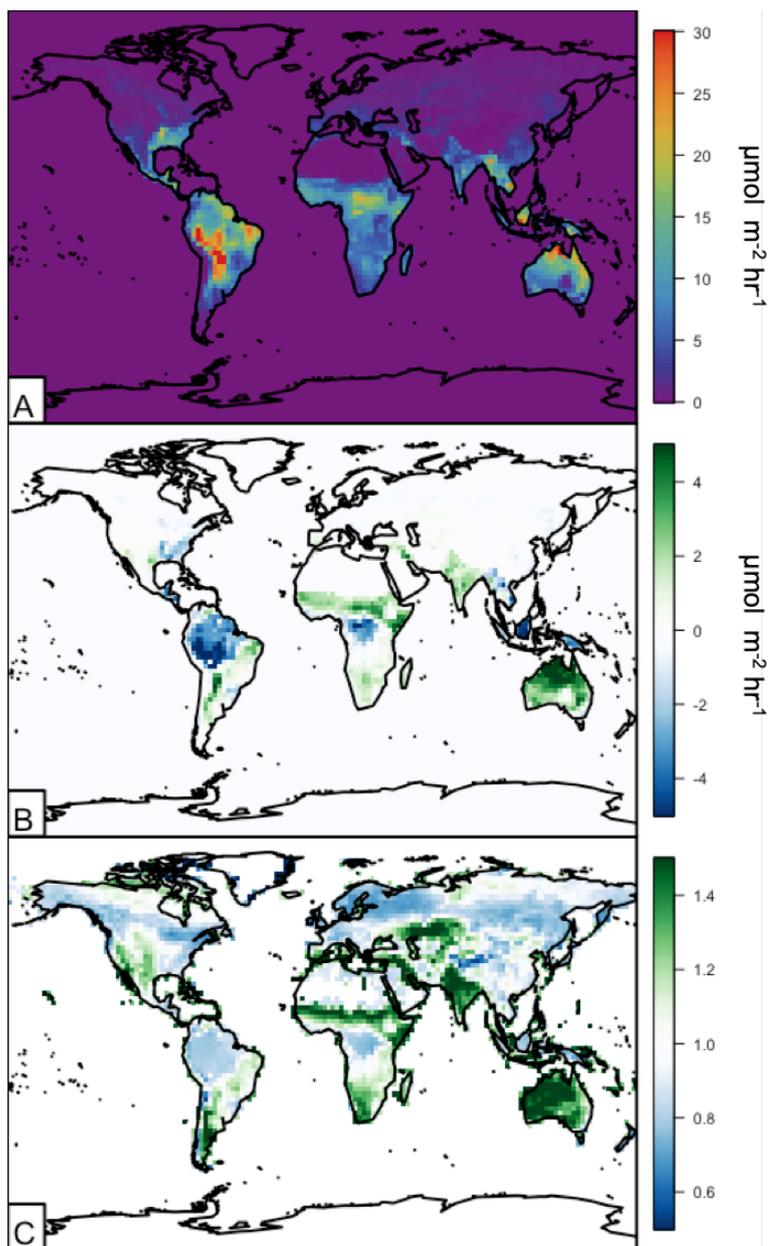
Implementing the updated canopy surrogate in a global model directly impacts the surface-atmosphere exchanges processes of biogenic emissions and dry deposition, which together influence the chemical composition of the atmosphere. In this section we outline the changes to both surface processes, focusing on isoprene emissions and ozone dry deposition, followed by the changes to surface level ozone concentrations in the GEOS-Chem model.

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 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 $\sim 350 \text{ Tg C yr}^{-1}$.

The annual average differences in the simulated isoprene emissions following implementation of the surrogate canopy model ($\text{MEGAN}_{\text{canopy}} - \text{MEGAN}_{\text{PCEEA}}$) 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



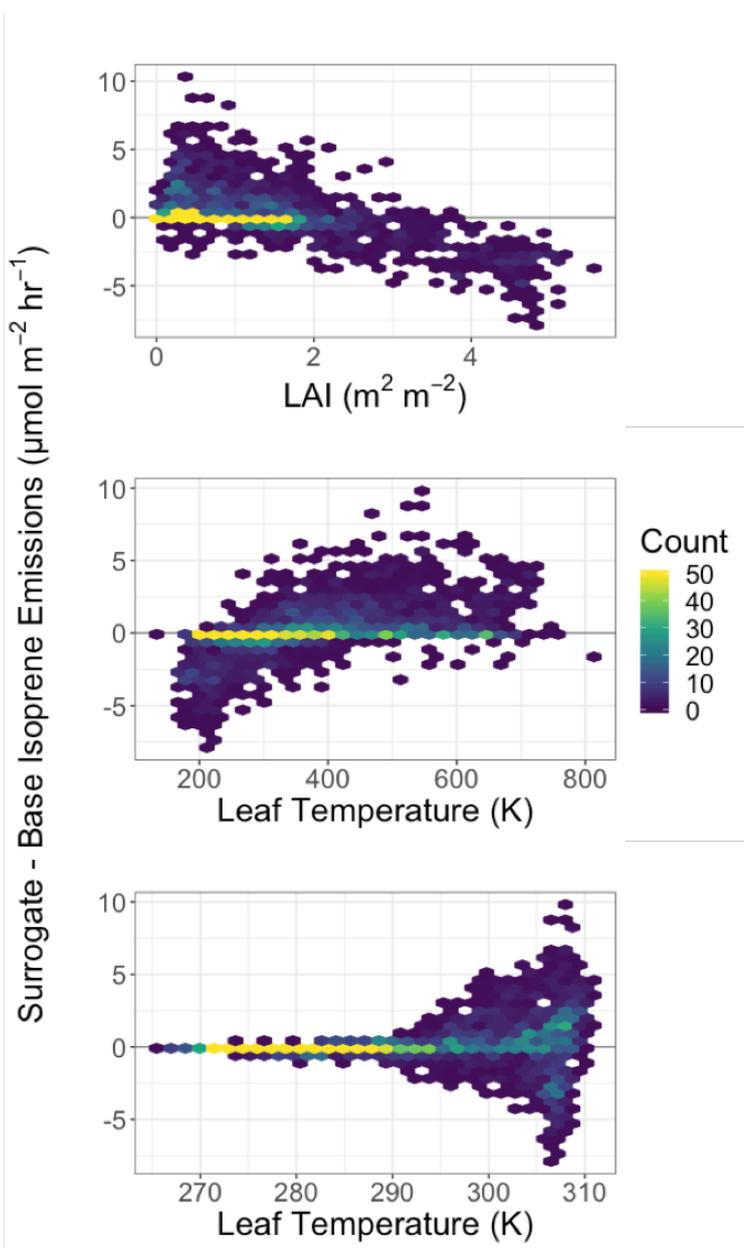
more apparent. This relative change ($\text{MEGAN}_{\text{Canopy}}/\text{MEGAN}_{\text{PCEEA}}$) is shown in Figure 9c. While there are relatively modest decreases in tropical and boreal forests, the emissions increase in heavily cropped Indian Subcontinent and sub-Saharan Africa show the largest relative change. Though the spatial variability in relative difference is substantial, the annual global isoprene emissions from the canopy model are within 5% of the original model version ($\sim 340 \text{ Tg C yr}^{-1}$). 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 ($\text{MEGAN}_{\text{PCEEA}}$) over a direct canopy model implementation in the MEGAN framework (as in the surrogate model application, $\text{MEGAN}_{\text{Canopy}}$).



565 **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).



570 It is not possible to directly parse the individual process contributions to the total
emissions changes due to the 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,
575 leaf-level PAR, and leaf temperature (Figure 10) indicates that the 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
580 shown in Figure 10. The forested areas with the largest decreases in 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.
585 Overall, these results indicate that the representation of canopy radiative physics is
more important than thermodynamically resolving the difference between air and
leaf temperature for simulating biogenic emissions in the MEGAN framework.



590

Figure 10. Difference in annual average isoprene emissions between the surrogate canopy model (MEGAN_{canopy}) and the base simulation (MEGAN_{PCEEA}) ($\text{atoms C cm}^{-2} \text{s}^{-1}$, see Figure 9B) as a function of LAI, Leaf Level PAR ($\mu\text{mol m}^{-2} \text{s}^{-1}$), 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.



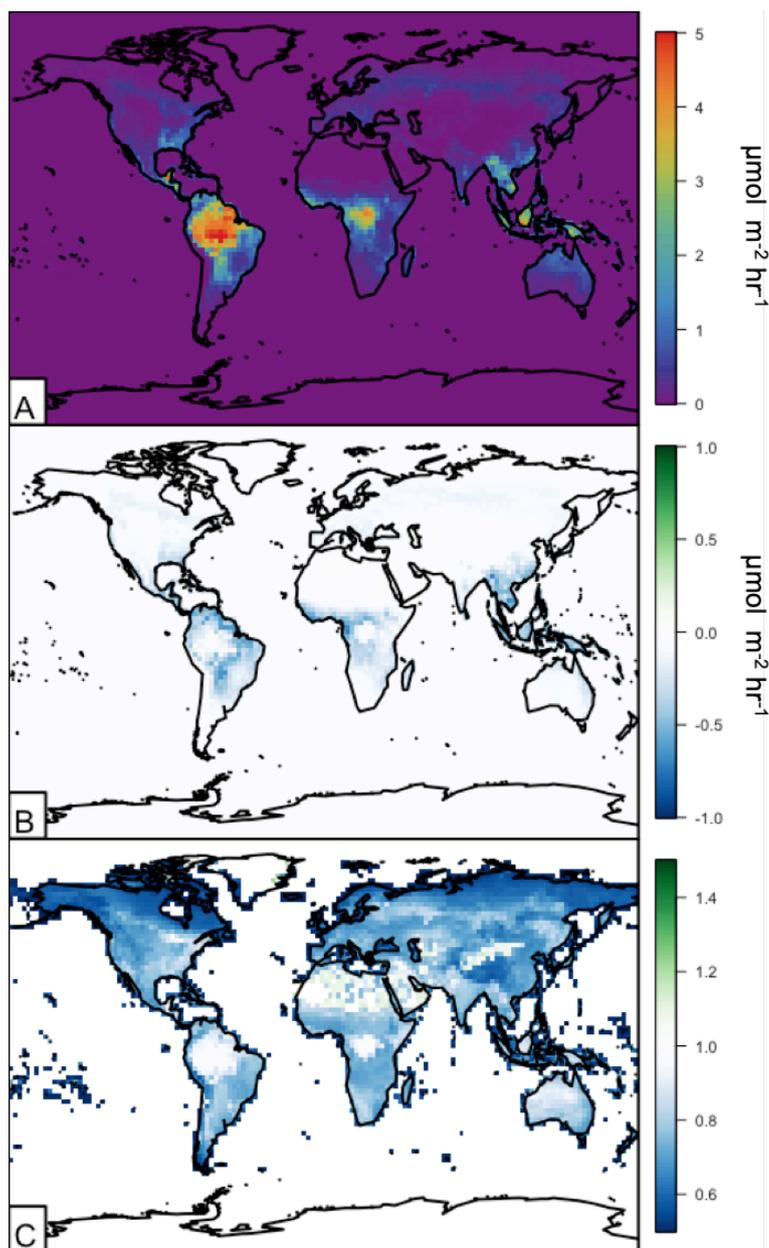
There are few spatial constraints on isoprene emissions that can act as independent validation data for the new model framework. However, recent work over the
595 southeast United States (Kaiser et al., 2018; Travis et al., 2016; Yu et al., 2016) indicates that the base version of GEOS-Chem used here (v12.3.0), which uses MEGAN_{PCEEAA}, 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
600 model into better agreement with these observational constraints.

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
605 emissions are shown in Figure 11a and are highest over the densely vegetated 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
610 and relative changes to monoterpene emissions due to the canopy surrogate model (MEGAN_{Canopy} - MEGAN_{PCEEAA}) 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
615 (Guenther et al., 2012). There is a fairly constant 20-30% decrease across regions



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

620 substantial, are well within the stated uncertainty ranges in monoterpene emissions of the MEGAN model (300-400%, Guenther et al., 2012).



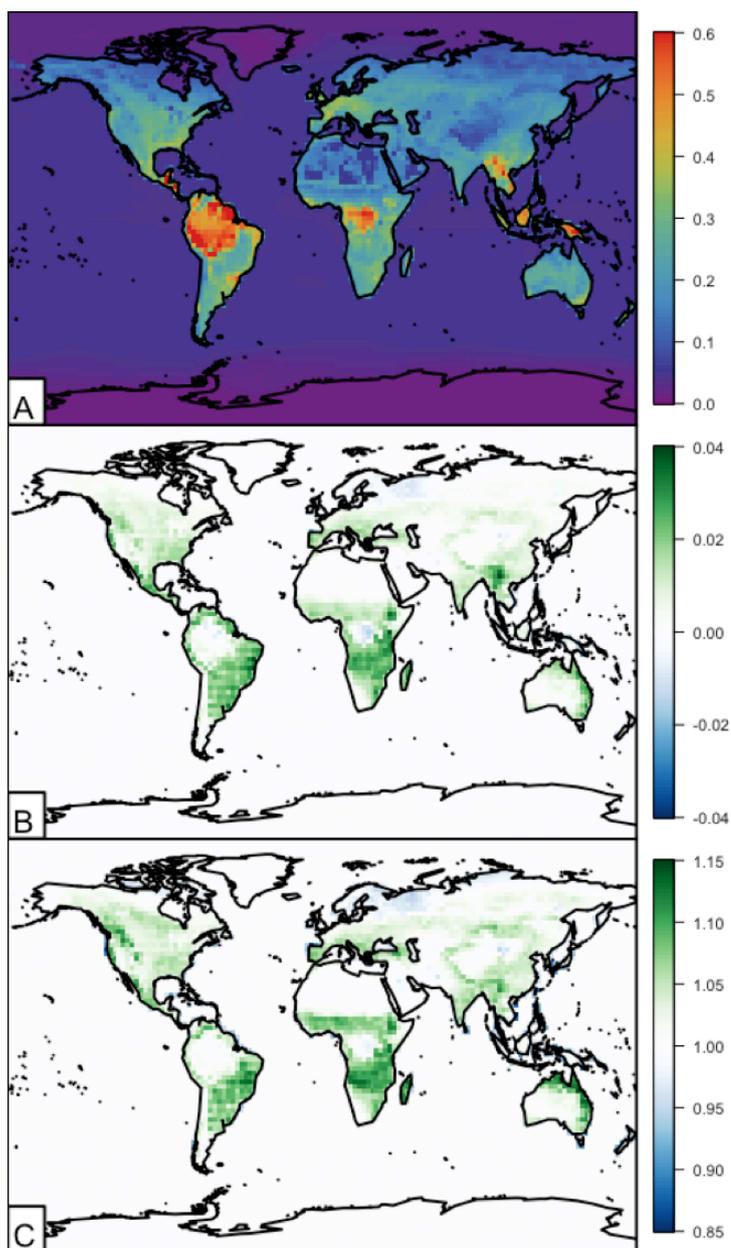
625

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).



Changes in simulated ozone dry deposition velocities in 2012 are summarized in
630 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 ($\sim 0.004 \text{ cm s}^{-1}$). The annual
635 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 $\pm 5\%$, or $\pm 0.01 \text{ cm s}^{-1}$, with a maximum change of 15%,
(0.04 cm s^{-1}). Relative changes track most strongly with broadleaf and coniferous
forested areas, consistent with those regions being most sensitive to stomatal
640 deposition (Silva and Heald, 2018).



645

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).



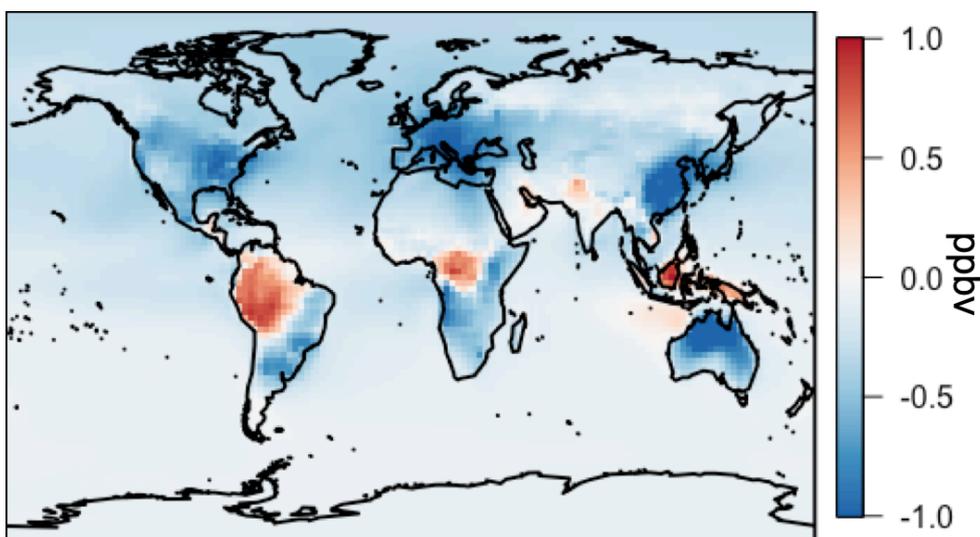
The small overall changes to surface-atmosphere exchange processes associated with the updated canopy scheme produce only a modest impact on simulated atmospheric composition. We describe the changes to surface ozone here, as an
650 illustrative example.

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
655 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
660 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
665 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
670 (Travis et al., 2016). The addition of the new canopy physics parameterization very
modestly reduces this bias by about ~ 1 ppbv, driving simulated ozone closer to
observations.



675 **Figure 13.** Annual spatial average surface ozone difference (ppbv) between the updated model version
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>),
680 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
685 speciation, and the leaf level emission factor database available from the MEGAN3
Emission Factor Preprocessor. All MEGAN3 Emissions Factor Preprocessor options
we 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 for some updates for the contiguous US. The spatial distribution of the
690 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 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
695 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 MEGAN2.1 using the MODIS LAI product used in this work, the resulting
emission factors are relatively similar (within $\pm 75\%$) though the MEGAN3 emission
factors are lower than those used with MEGAN2.1 in GEOS-Chem. Generally, more
700 than half of all changes are within $1000 \mu\text{mol m}^{-2} \text{hr}^{-1}$, and 90% of all emission
factors are within $3000 \mu\text{mol m}^{-2} \text{hr}^{-1}$. This comparison is not exact, due to the fact
that the MODIS LAI product used here is likely 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
705 similar to those used earlier in this work (Figure 8), with a spatial R^2 of 0.91.

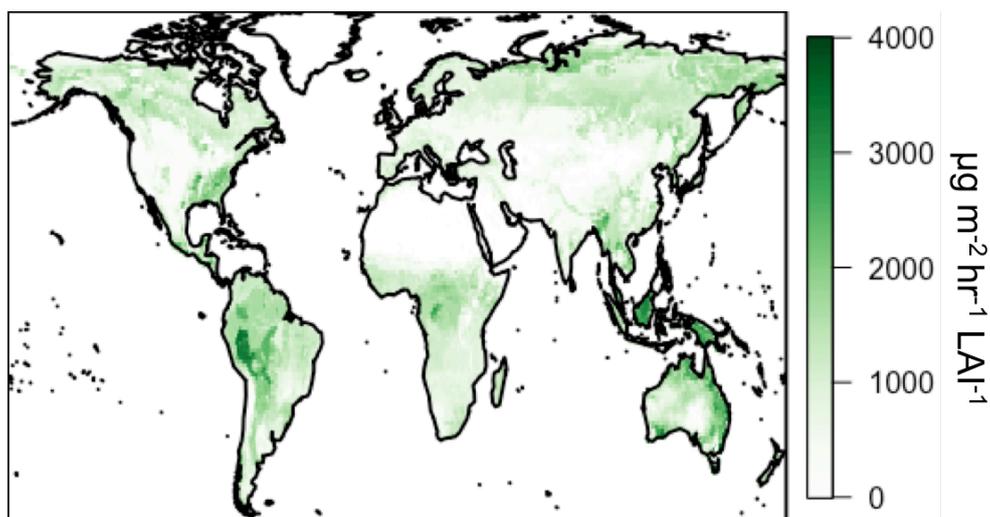


Figure 14. MEGAN3 isoprene emission factors ($\mu\text{g m}^{-2} \text{hr}^{-1} \text{LAI}^{-1}$).

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

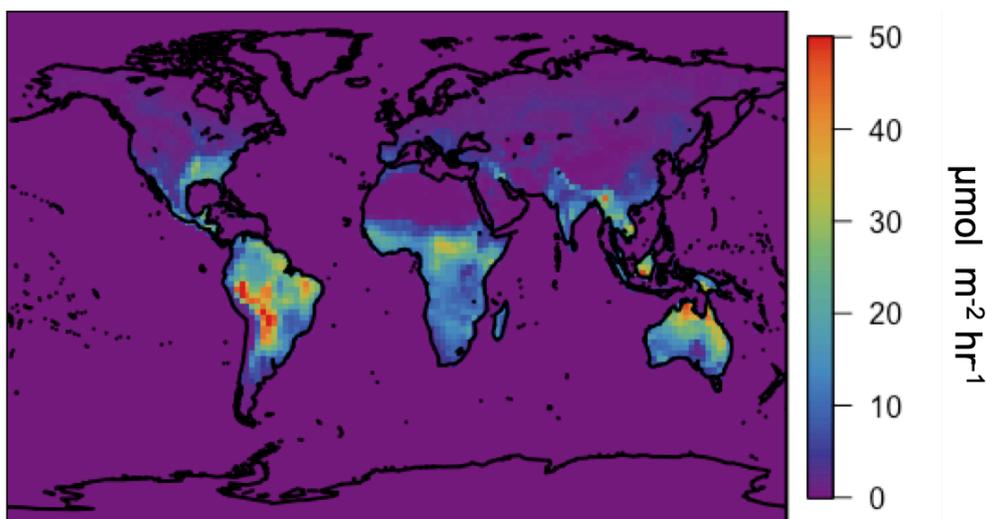
710 “MEGAN3_{Full}” as the emissions implementation in GEOS-Chem v12.3.0 using the MEGAN3 leaf-level emission factors and MEGAN3 activity factors with canopy physics calculated following the canopy surrogate model described in Section 3. The annual isoprene emissions simulated using MEGAN3_{Full} are higher than using the MEGAN2.1 canopy-scale factors in GEOS-Chem (as in both MEGAN_{Canopy} and

715 MEGAN_{PCEEA}), but more in line with previous work (Guenther et al. 2012). Specifically, annual total isoprene emissions for 2012 are $\sim 570 \text{ Tg C yr}^{-1}$ 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

720 factor of 0.21, which additionally removes the need for the somewhat arbitrary



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). These annual average isoprene emissions
725 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 R² of ~0.8.



730 **Figure 15. Annual average (2012) isoprene emissions simulated in GEOS-Chem driven by the surrogate 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
735 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.,
740 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
745 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.

750 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
755 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.

760 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
765 models of atmospheric chemistry. Future work should include more detailed observational constraints and characterization of in-canopy chemical reactions, turbulent exchange, and biological processes.

Author Contributions

CLH and SJS designed the study. SJS developed and implemented the surrogate
770 model and performed simulations and analysis. ABG developed MEGAN3 model code. All authors contributed to the manuscript preparation.

Code Availability

The MEGAN3 and GEOS-Chem model code are available at:
<https://bai.ess.uci.edu/megan/data-and-code>, and
775 <https://zenodo.org/record/2620535>, respectively. The updated GEOS-Chem code containing the canopy model changes is available at:
<https://zenodo.org/record/3614062>.



Acknowledgements

This study was supported by the U.S. National Science Foundation (AGS 1564495)
780 and by NASA Headquarters under the NASA Earth and Space Science Fellowship
Program (NNX16AN92H).

Competing Interests

The authors declare no competing interests.

References

- 785 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.
- Ashworth, K., Chung, S. H., McKinney, K. A., Liu, Y., Munger, J. W., Martin, S. T. and
790 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
gaseous deposition to vegetated surfaces, *Atmospheric Environment* (1967), 21(1),
795 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.*,
106(D19), 23073–23095, doi:10.1029/2001JD000807, 2001.
- 800 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.
- 805 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.



- 810 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.
- 815 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 States, *Atmos. Chem. Phys.*, 16(4), 2323–2340, doi:10.5194/acp-16-2323-2016, 2016.
- 820 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., 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.
- 825 Geron, C., Daly, R., Harley, P., Rasmussen, R., Seco, R., Guenther, A., Karl, T. and Gu, L.: 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.
- 830 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 (GFED4), *J. Geophys. Res. Biogeosci.*, 118(1), 317–328, doi:10.1002/jgrg.20042, 2013.
- 835 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*, 31(22), doi:10.1029/2004GL021259, 2004.
- 840 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.
- 840 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.
- 845 Guenther, A., Karl, T., Harley, P., Wiedinmyer, C., Palmer, P. I. and Geron, C.: 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.



- 850 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 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.
- 855 Harley, P., Guenther, A. and Zimmerman, P.: Effects of light, temperature and canopy position on net photosynthesis and isoprene emission from sweetgum (*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.
- 860 Hoesly, R. M., Smith, S. J., Feng, L., Klimont, Z., Janssens-Maenhout, G., Pitkanen, T., 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.
- 865 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.
- 870 IPCC: *Climate Change 2013: The Physical Science Basis. Contribution of Working 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.
- 875 Kaiser, J., Jacob, D. J., Zhu, L., Travis, K. R., Fisher, J. A., González Abad, G., Zhang, L., Zhang, X., Fried, A., Crouse, J. D., St. Clair, J. M. and Wisthaler, A.: High-resolution 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.
- 880 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.



- 885 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.
- 890 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.
- 895 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.
- 900 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.
- 905 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.
- 910 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 Chemistry and Physics*, 10(7), 3405–3425, doi:https://doi.org/10.5194/acp-10-3405-2010, 2010.
- 915 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 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.
- 920



- 925 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, 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.
- 930 Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V. N., 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 Earth A new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity, *BioScience*, 51(11), 933–938, doi:10.1641/0006-3568(2001)051[0933:TEOTWA]2.0.CO;2, 2001.
- 935 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. Geophys. Res. Atmos.*, 2017JD027278, doi:10.1002/2017JD027278, 2018.
- 940 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.
- 945 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.
- 950 Travis, K. R., Jacob, D. J., Fisher, J. A., Kim, P. S., Marais, E. A., Zhu, L., Yu, K., Miller, C. 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.
- 955 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 O₃-NO_x-hydrocarbon chemistry: 1. Model formulation, *Journal of Geophysical Research: Atmospheres*, 103(D9), 10713–10725, doi:10.1029/98JD00158, 1998.



- 960 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.
- 965 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., 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.
- 970 Zhang, L., Brook, J. R. and Vet, R.: A revised parameterization for gaseous dry deposition in air-quality models, *Atmos. Chem. Phys.*, 3(6), 2067–2082, doi:10.5194/acp-3-2067-2003, 2003.