Supporting Information for "Accurate Simulation of Both Sensitivity and Variability for Amazonian Photosynthesis: Too Much to Ask"

Sarah M. Gallup[1], Ian T. Baker[2], John L. Gallup[3], Natalia Restrepo-Coupe[4], Katherine D. Haynes[2], Nicholas M. Geyer[2]and A. Scott Denning[1,2]

[1]Graduate Degree Program in Ecology, Colorado State University, Fort Collins, Colorado, USA

[2]Department of Atmospheric Science, Colorado State University, Fort Collins, Colorado, USA

[3]Department of Economics, Portland State University, Portland, Oregon, USA

[4]

Contents

Text S1: Difference between the variances of outcomes and of predictions

Let $y_i = h(\mathbf{x}_i, \beta) + \varepsilon_i$ be the equation of an outcome y_i to be predicted, where $\mathbf{x}_i$ is a $K \times 1$ vector of explanatory variables, β is a $K \times 1$ vector of parameters, ε_i is a random error

Corresponding author: Sarah Gallup, sgallup@colostate.edu

term with mean 0 and variance σ^2 and $i = 1, \ldots, n$ is an index of observations. Let $h(\cdot, \cdot)$ be a known function, which if it were linear would be $h(\mathbf{x}_i, \beta) = \beta_1 + \beta_2 x_{i2} + \cdots + \beta_K x_{Ki}$.

Consider the prediction of a specific outcome, y_i corresponding to the explanatory variables $\mathbf{x}_i$. If we have a consistent estimator $\hat{\beta}$ of the parameters β, then the predicted value of y_i is $\hat{y}_i = h(\mathbf{x}_i, \hat{\beta})$. Since $\hat{\beta}$ is consistent, as the sample size n becomes large its limit in probability is precisely β, so in the limit $h(\mathbf{x}_i, \hat{\beta})$ becomes indistinguishable from $h(\mathbf{x}_i, \beta)$. However, even in the probability limit, $\hat{y}_i \xrightarrow{p} E[y_i|\mathbf{x}] \neq y_i$ because $y_i = h(\mathbf{x}_i, \beta) + \varepsilon_i$ also has the random error term ε_i.

The variance of y_i has two parts, $V[y_i] = V[h(\mathbf{x}_i, \beta)] + V[\varepsilon_i]$, assuming $\mathbf{x}_i$ is uncorrelated with ε_i, as typically is necessary for $\hat{\beta}$ to be consistent. In large samples, $V[\hat{y}_i] \xrightarrow{p} V[h(\mathbf{x}_i, \beta)]$, but the variance of y_i is larger:

$$V[y_i] = V[h(\mathbf{x}_i, \beta)] + \sigma^2$$

due to the presence of the random error term ε_i in y_i.

Text S2: Methods Details

2.1 Process Models

The Multi-scale synthesis and Terrestrial Model Intercomparison Project [MsTMIP; *Huntzinger et al.*, 2014; *Wei et al.*, 2014] isolates land model GPP structural responsiveness from output differences due to varying inputs. Variants of four of the models participate in IPCC's forecasts. Comparing runs based on standardized drivers is important for the Amazon because its rainfall differs strongly across ESMs [*Ahlström et al.*, 2017; *Huntingford et al.*, 2013; *Jupp et al.*, 2010; *Li et al.*, 2006; *Poulter et al.*, 2010]. MsTMIP did not prescribe how modelers should distribute monthly meteorology into the shorter time steps at which many models run. Forcing all models with the same weather does omit feedbacks between weather and GPP [*Gloor et al.*, 2013; *Harper et al.*, 2014] at times scales longer than a single month.

All participating MsTMIP models provided outputs of GPP, respiration, and closely-derived net ecosystem productivity. While MsTMIP invited additional variables, their absence for at least a varying third of models hampers comparative analysis. Runs represent each model's configuration in about 2014. A subsequent update of CLM, for example, specifically addressed previously excessive modeled tropical GPP [*Oleson and Lawrence*, 2013, p. 9].

Weather reanalyses are less certain for the tropics than for midlatitudes [*Clark and Clark*, 2011; *Li et al.*, 2006; *Malhi and Wright*, 2004]. The only striking outliers in the MsTMIP meteorology were retained. From 4000 to 8,597 mm of rain in January, 2000 is assigned to 56 half-degree grid cells. Otherwise the highest monthly rainfall anywhere in the study area in any month is 2,431 mm. Following convention for the wet tropics [*Saleska et al.*, 2003], dry season is defined as months when long-term mean precipitation is below 100 mm, or less than the approximate maximum plants can metabolize in real time [*Aragão et al.*, 2007].

SiB4 meteorology and land cover drivers were developed in conjunction with the new model version and differ slightly from MsTMIP's.

2.2 Statistical Models

MsTMIP models plus SiB4 are referred to as process models, since each simulates the biological determination of GPP. In contrast, data assimilation estimates of global GPP, labeled as statistical models, simulate retrospectively from remotely sensed inputs. Ideally they are sufficiently accurate to benchmark process models [*Jung et al.*, 2011; *Zhang et al.*, 2017]. They are driven not by MsTMIP weather but by closely-related weather reanalyses.

Global satellite inputs temper Fluxcom's and extrapolate cleaned eddy covariance flux estimates. Fluxcom is widely used as reference GPP globally [for example, *Anav et al.*, 2015; *Bonan et al.*, 2011; *Collier et al.*, 2018; *Jung et al.*, 2019; *Malavelle et al.*, 2019; *Mystakidis et al.*, 2016; *Piao et al.*, 2013; *Tramontana et al.*, 2016; *Xu et al.*, 2015]. We use the half-degree resolution product from the multivariate adaptive regression splines algorithm.

Wecann is similar in both concept and results to Fluxcom. With additional input from GOME-2's solar-induced fluorescence, Wecann fits tower sensible heat, latent heat, and GPP slightly better than does Fluxcom [*Alemohammad et al.*, 2017]. Wecann's one degree resolution is coarser than MsTMIP's. We attribute each value to four half-degree cells, and note below adjustments made to avoid artificially narrowed confidence intervals.

The third statistical model, vegetation photosynthesis model (VPM) is a light-use efficiency model. VPM applies deliberately few and non-fitted numeric constants to temperature reanalysis and multiple MODIS and SPOT satellite products [*Xiao et al.*, 2005; *Zhang et al.*, 2017]. For a test year in North America, VPM provided the median estimate compared to six other global GPP models [*Zhang et al.*, 2016]. Being even more heavily dependent on satellite data than is Fluxcom, VPM is likely to be less accurate in the cloudy tropics than elsewhere, and less accurate for the tropical wet season than for the dry season.

2.3 Study Boundaries

Selecting EBF tiles within cells is not workable because MsTMIP models' GPP estimates are not available for individual PFTs. To assess GPP that is representative for each cell's vegetation (Fig. 1) requires that cell values should be an average only across land area. The MsTMIP models [*Chapin et al.*, 2006], SiB4, and WeCann [*personal communication, Alemohammad, 2020*] give GPP as a mean value across both land and water. VPM [*Zhang et al.*, 2017] and Fluxcom [*Jung et al.*, 2019] provide GPP as means for a cell's land area

only. All GPP datasets except VPM and Fluxcom were adjusted by the cell's water fraction in the MsTMIP PFT map. In summary metrics, months are treated as if they are equally long.

2.4 Eddy Covariance Fluxes

While there is some utility in simply comparing models, knowing their true accuracy is far more useful. The statistical models are candidate reference data sets, but despite circularity issues addressed below, we wish to assess their accuracy as well. For the Amazon, GPP from individual eddy covariance towers is the only remaining option. ECs measure exchange between the land surface and the atmosphere of CO_2 and other gasses that vegetation affects. From measurements related to net ecosystem exchange, the large and opposing contributions of GPP and ecosystem respiration are modeled.

EC GPP was further limited to the study period starting in 2000, cutting off a few months each at sites CAX, K34, and RJA. The merged EC dataset offers eleven GPP algorithm options. Consistent with *Baker et al.* [2013], we use "GEP_model." The two ECs in the Tapajos National Forest are about a dozen kilometers apart in stands with different logging histories. Due to their related synoptic weather and seasonality, for this study the K67 and K83 sites are best considered as pseudo-replicates.

Six sites in the South American rainforest cannot fully represent the region's range in either plant productivity or other criteria. For most study models the six cells that contain a flux tower site encompass less than a third of the model's central 98% of range in mean annual GPP across the Amazon. However, EC site cells typically fall on both sides of a model's median GPP. EC cells also tend to have high GPP (Text S3), which is useful because similarly annual high productivity is uncommon at other flux towers globally whose tendencies might otherwise help constrain modeling of the tropics.

Text S3: Representativeness of EC Sites

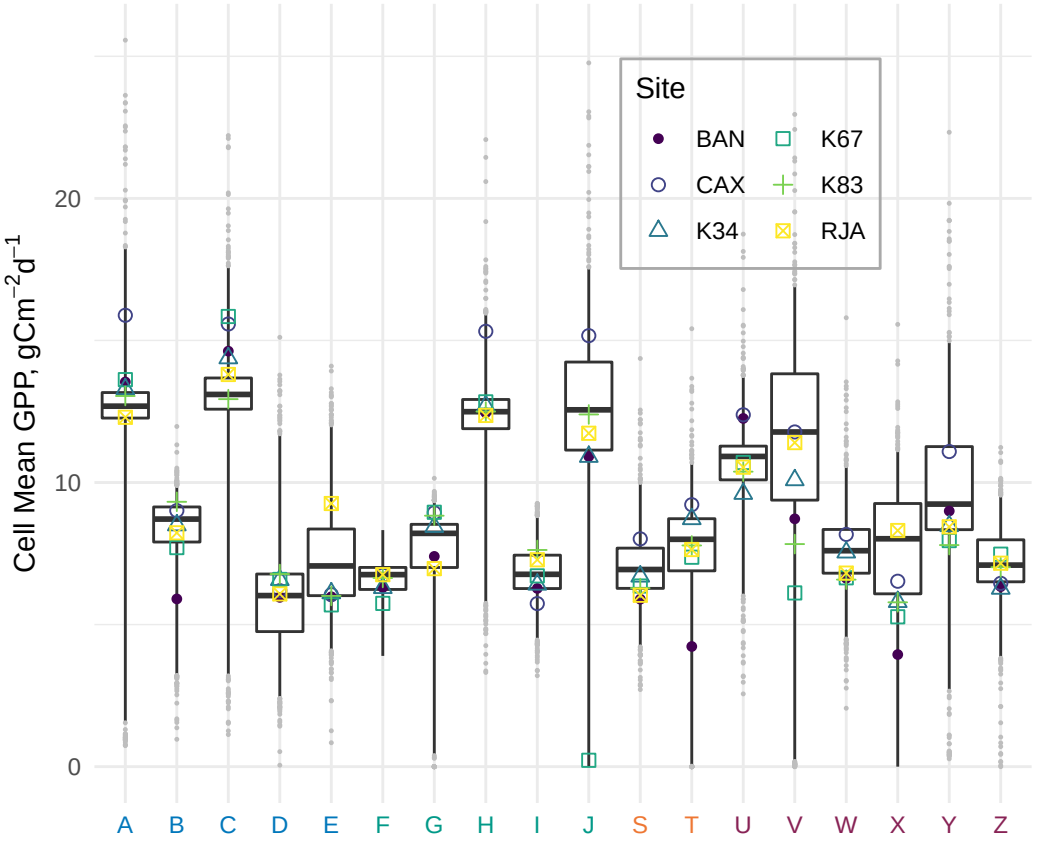


Figure S1: Distribution of cell-level GPP averaged across all study months and the entire study area. A horizontal bar marks a model's median GPP. Boxes encompass the central 50% of a model's cells as ranked by mean GPP, outside of which all grey points indicate all but the most extreme 2% of cells. Colored symbols mark EC cells. Especially for the some of the more responsive models, toward the left, the EC sites are in cells with above-average productivity.

From the perspective of the models being contrasted, how representative are the EC sites of the entire Amazon? One basis of comparison is how much of the range in mean GPP for the 11-year study period the six cells that contain a flux site capture compared to the range across all EBF cells. EC sites are compared to the basin's range as defined by its 1st and 99th percentile values. The outlier is model J, for which EC sites span 85% of the watershed's range in GPP due largely to its near-zero GPP for K67. Among the remaining models, mean GPP for cells with ECs cover from 9% and 51% of the range of the central 98% of Amazon rainforest cells. The median among the models in span that ECs represent

is 29%. For most of the study models, the six ECs represent less than a third of the range in mean grid-cell GPP.

On the other hand, the EC sites represent a portion of the rainforest GPP range that is especially valuable to match. There are disproportionately few flux towers in the exceptionally productive tropics than there are in some of the world's less productive biomes. If plant responsiveness across the global range of environmental driver values is mostly continuous although non-linear, then ECs elsewhere may help constrain modeling of the low end of the rainforest productivity spectrum. By this criterion, the most useful EBF flux towers are at the most productive sites.

The six EC cells do have a higher mean GPP than is typical of the Amazon basin. The median cell-level annual GPP across all the models that the most productive of the sites represents as a percentile of each model's central 98% of study cells is 85%, with a range of 50% to 99% . In most but not all models, the most productive flux site cell is CAX. For the least rather than most productive of the EC cells, usually the cell that contains BAN, model percentiles range from 4% and 49% with a median of 17%. While the cells with eddy covariance data cover only a limited portion of the Amazon basin's range in mean annual GPP, they are in relatively productive sites for which closely-related alternative data are least plentiful.

Text S4: Non-Linearities in GPP Responses to Rain and Light

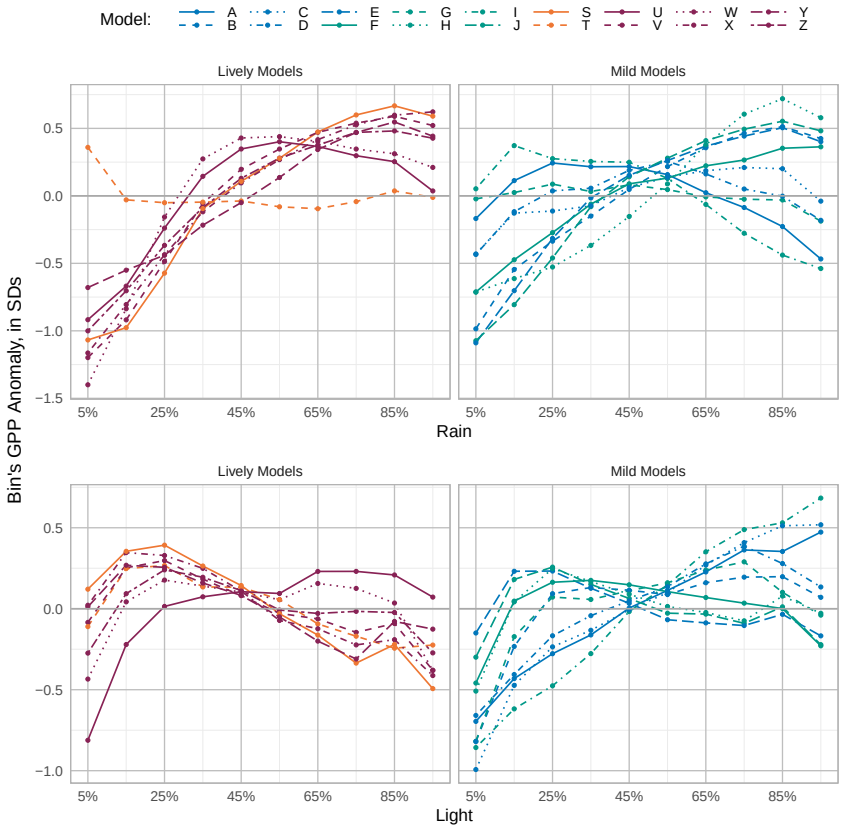


Figure S2: Non-linearity in modeled GPP's responses to rain and light, parallel to the main paper's Fig. 6 for temperature.

In Fig. S2, GPP is shown as z-score relative to a particular model and cell's mean across the study period. Shared MsTMIP driver data is binned by deciles. Mild models are in the left panels, and lively models on the right. Nearly all models simulate GPP as falling in the very wettest months although in only a few mild models is it below average. In the driest 10% of months GPP is below average in all but two models.

GPP increases with rain at the driest deciles and falls in the very wettest months in all but one model. One difference from temperature is that responses to rain for individual mild models are more nearly linear, and models diverge from each other for the driest months in approximately reverse rank as they do at the wettest. Model T, a statistical model, is an exception, with no discernable trend in response to varying rain. The inflection point at which GPP switches from increasing to falling with more rain ranges among models from the sec-

ond to ninth deciles of current month's rain. The drop begins at drier levels in most of the mild models. For some of the lively models, more rain corresponds to higher GPP until about the top two deciles. The disparities represent disagreement about what moisture is optimal for EBF, although do not reveal how modeled soil moisture mediates these responses within many of the models.

Radiation's responsiveness too is almost uniformly curved, consistent with a classic light response curve [*Baker et al.*, 2019]. In the darkest months nearly all models simulate low GPP. Consistent with differences that *Rogers et al.* [2017] noted, the flex points of models' light saturation divide into two groups, one slightly below 200 Wm^{-2} and the other near 220 Wm^{-2}. Some of the lively models show strong drops in plant productivity in especially bright months.

Text S5: Months of Modeled GPP's Seasonal Peaks at EC Sites

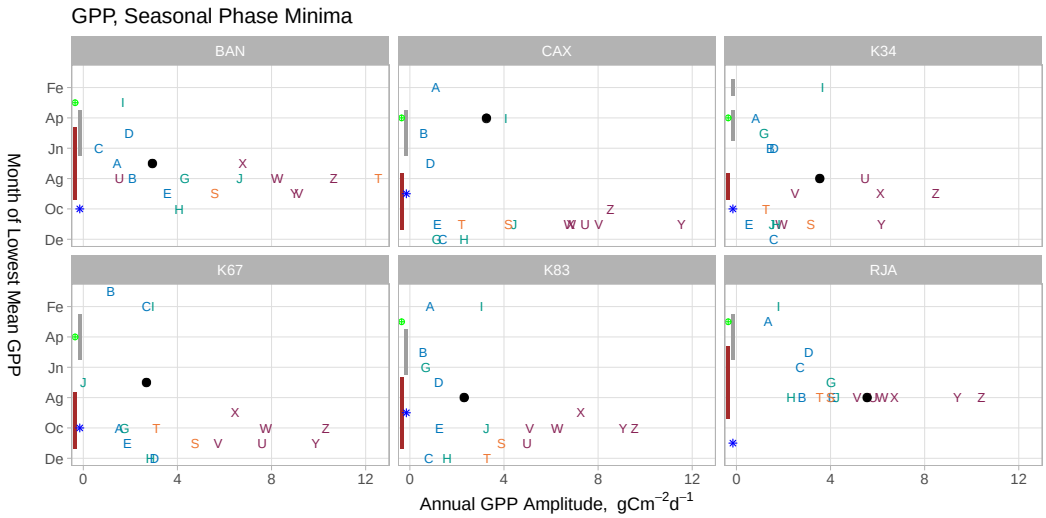


Figure S3: For each site, the month of lowest mean monthly GPP for each mode. The x-axis and symbol colors rank models by extent of seasonality. Grey bars on each site's y-axis indicate the three months with the least light, while a blue star marks the brightest month. Brown bars show the dry season. The wettest month is indicated with a green dot. Mild models are more likely to simulate minimum GPP during a dark month, while lively models' lowest GPP typically occurs during the dry season.

Fig. S3 shows that models tend to fall into the same groups for seasonal phase as they do for seasonal amplitude, both reflecting their relative responsiveness to drivers. Mild and lively models have nearly identical mean timing differences between EC and modeled GPP, of 2.9 and 2.3 months respectively. The model groups differ in the direction of differences. Models with little seasonality tend to simulate the year's lowest GPP before or early in the dry season. For every lively model except Model I, GPP is lowest at every site either during the dry season or in the first month afterward (Fig. S3). Most of the lively models' minima occur late in the dry season when modeled soil moisture presumably is lowest. Again excepting Model I, no lively model simulates minimum GPP during the three darkest months for any site.

Patterns are similar for month of highest rather than of lowest mean GPP (Fig. S4). EC GPP at all sites but CAX peak 2-5 months after the last dry season month. Half of the models, a mix of mild and lively, match CAX' timing to the extent of peaking during its four-

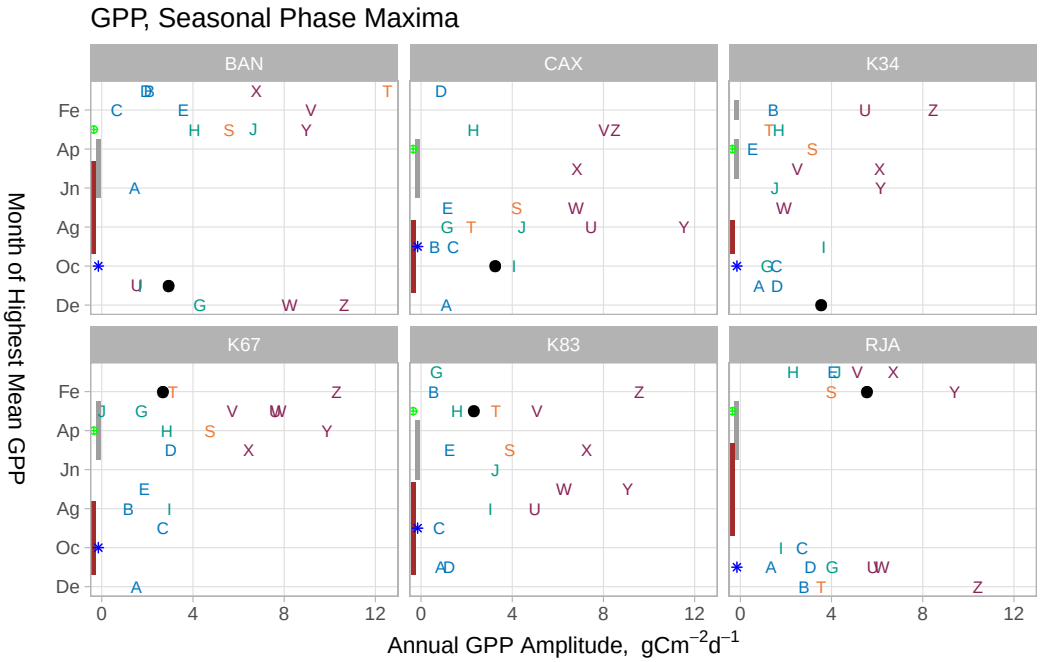


Figure S4: Parallel to Fig. S3 of month of lowest GPP at each site, but showing month of highest GPP. While the peak and trough of the seasonal phase is not consistently offset by six months, overall patterns of model responses are similar between lowest and highest GPP timing.

month dry season. Peak month is most accurately modeled for RJA, where neither the EC tower nor any model reaches maximum GPP during the dry season.

The wet tropics have a modest annual cycle in both leaf area index and the photosynthetic capacity of average leaves [*Albert et al.*, 2018; *Borchert et al.*, 2002; *Dahlin et al.*, 2017; *Doughty and Goulden*, 2008; *Goulden et al.*, 2004; *Samanta et al.*, 2012; *Wilson et al.*, 2001; *Wu et al.*, 2016] but see [*Morton et al.*, 2016]. Seasonal rainforest leaf phenology is thus far absent from most process models of GPP *Albert et al.* [2019]. While there is at least speculative logic for the timing of each site's maximum plant stress, the dominant mechanism appears to vary across sites. At Tapajos, sites K34 and K83, the lowest EC GPP occurs early in the dry season (Fig. S4) and corresponds to the annual peak of leaf exchange. The timing at K67 and BAN also is reasonably consistent with a leaf demography hypothesis. RJA reaches its lowest EC GPP late in its stark dry season. CAX's minimum is during a dark month in the middle of its mildly wetter season. The timing of rainforest leaf exchange may respond to a continuum of water v. light limitation even if instantaneous GPP does not *Albert*

et al. [2019]. The mismatches suggest that adding tropical leaf seasonality could improve the accuracy of modeled GPP.

Text S6: Cumulative rainfall's influence on EC GPP

Although site intercepts and current weather in simple regressions explain on average 64% of modeled GPP's variance, on average about a third remains unexplained. For the ECs, the correlation of soil moisture with GPP is -0.31. Lively models' lower GPP than ECs during the dry season suggests that modeled rainforest plants experience more severe water stress than real plants. Process models simulate and track soil moisture, often at multiple depths. Cumulative water deficit may be a key variable omitted from the descriptive regressions. Soil moisture output is not available for enough models, but an indirect indicator of its effects is the strength of connection between modeled GPP and cumulative rain over recent months. The added explanatory power of cumulative rainfall is one way to characterize the strength of a site's hydrologic memory.

Testing soil moisture modeling directly requires reasonable reference GPP across the basin's spectrum of annual precipitation. Local plants logically adapt to the degree of drought they experience episodically [*Corlett*, 2016], and satellite data imply that sensitivity to tropical drought is spatially heterogeneous [*Bonal et al.*, 2016; *Feldpausch et al.*, 2016]. Soil moisture varies markedly also at fine scales, making it difficult to measure [*Broedel et al.*, 2017; *Huang et al.*, 2016] or model [*Parazoo et al.*, 2014] for even an EC footprint. Benchmarking modeled soil moisture across the Amazon is therefore particularly challenging.

It would be helpful if instead accumulated rain were a rough proxy for soil moisture. Rain summed over periods ranging from only the most recent month to the entire last year explain greatly varying portions of individual sites' GPP variability. Each point in Fig. S5, represents a regression of GPP on MsTMIP temperature, light and accumulated rain, optimized for a single site except the summary line for all sites. Dots and solid connecting lines mark regressions whose rain coefficients are statistically significant at $p \leq 0.05$. Dashed lines pass through r^2 of regressions whose rain coefficients fail the significance test. The maximum predictive power of a full year for all sites is of little practical consequence. A full year's cumulative rain is significant at only one site.

For each site, the point for one month of lag in Fig. S5 shows how much of the variation in GPP current weather alone explains. The difference from each site's peak value indicates how much more information rain history can add to current month's weather. The legend lists each site's maximum fit improvement due to cumulative rain. As with current

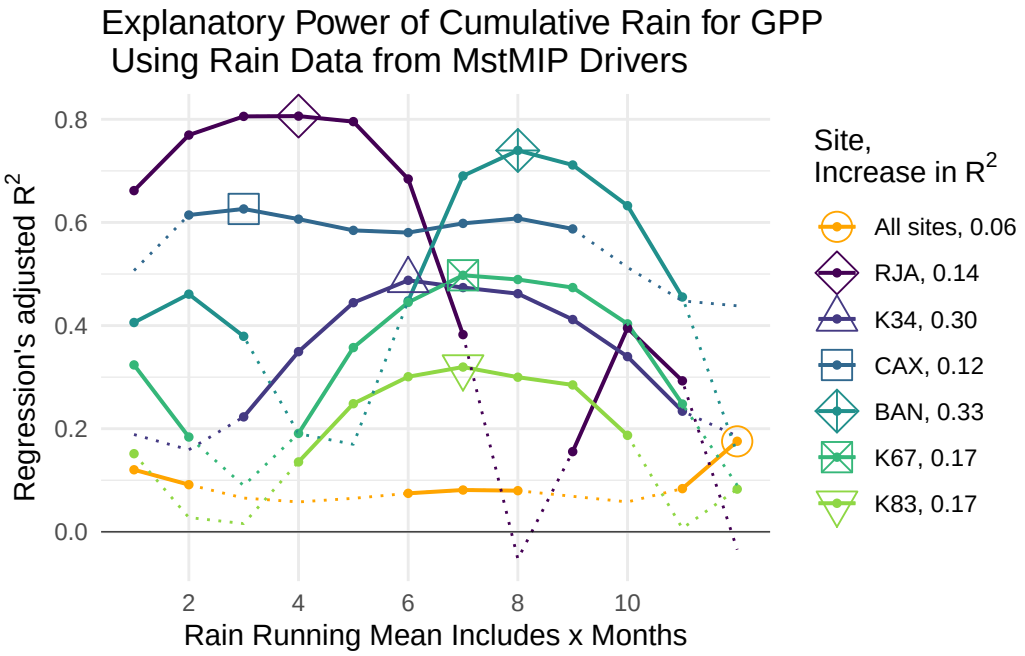


Figure S5: For each site, the rain accumulation duration best predicts EC GPP. The x-axis shows the number of months before and including the current month that are totaled. The y-axis indicates the portion of variability explained by regressing each rain accumulation period plus current month's light and temperature on GPP. Large open symbols indicate each site's accumulation duration with the most explanatory power. The increase in r^2 listed in the legend equals the difference between the optimal formula and one that uses only current month's rain. The month with the most explanatory power varies so much that no single duration explains more of the variability across all sites than can current month's rain alone.

month predictors of GPP, weather measured near each EC has slightly less explanatory power (Fig. S6).

We attempted to predict each site's best rain lag period. The negative coefficient on annual average rain, -0.010 months of optimal lag per mm increase, implies that a longer lag and possibly greater soil moisture retention capacity or deep rootedness exist at relatively dry sites. Correlations are very low for latitude, dry season length, and mean rain during the three driest months. Site mean annual precipitation explained 73% of the variation in best lag length.

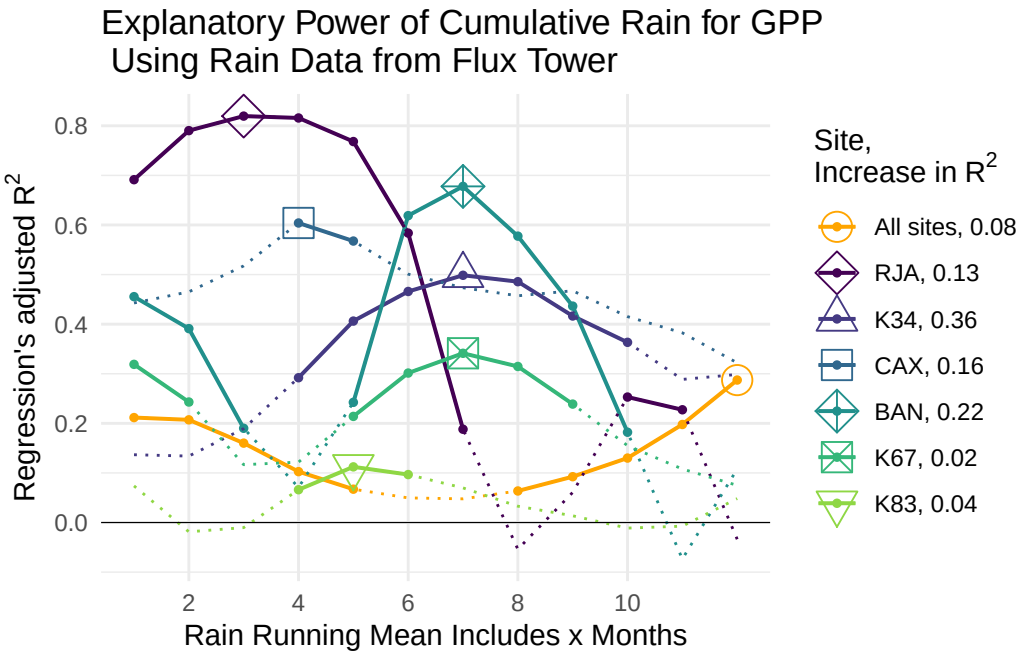


Figure S6: Parallel to Fig. S5 of rain lag duration versus explanatory power, using EC meteorology rather than MsTMIP weather reanalysis rain. The sharpest differences in results for the two meteorology sources are the worse fits of EC rain for K67 and K83. EC rain does slightly better than MsTMIP rain for K34 and CAX.

Highlighting the difficulties in developing modeling equations that are reasonable for all sites are differences in Fig. S5's site-specific light and temperature coefficients. Particularly for light, responsiveness is typically two to eight times stronger at individual ECs than when calculated across all sites simultaneously. Light at one or more rain lags is significant at only three sites, one only at cumulations longer than 9 months. That light nonetheless is significant for the regression across all sites for every lag period option suggests that light may be partly a surrogate for omitted drivers correlated with latitude. Temperature differs more across sites, significant in regressions for 2 to 7 inconsistent groupings of the 12 possible lag periods.

Figure S7: Mean and Variance of GPP for Each Site

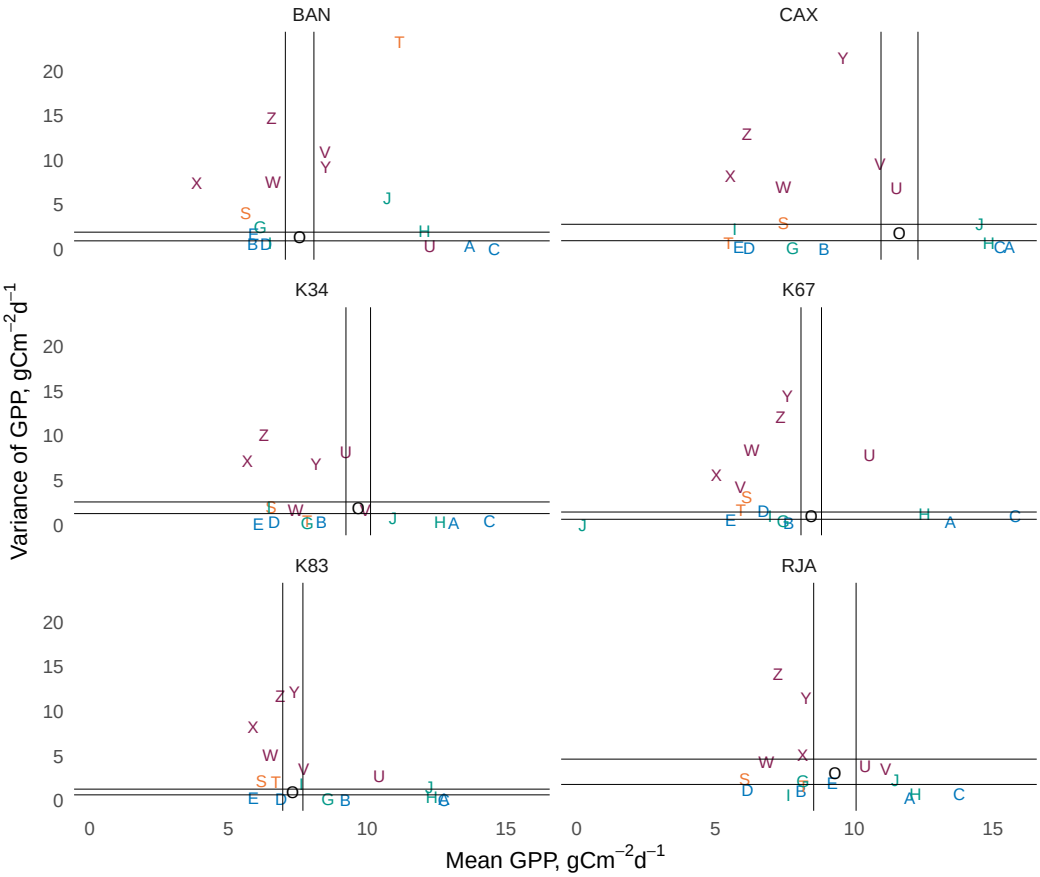


Figure S7: For each eddy covariance site, comparison of EC estimates to the means and variances in GPP estimates from each statistical or process model. Vertical and horizontal lines bracketing Model "O", EC estimates, are 99th percentile confidence bounds. For most models, both mean and variance fall outside the confidence bounds, with some models higher and some lower. For some models, their GPP variability exceeds that of EC estimates even more markedly than does mean GPP.

Comparing GPP's mean and variance for individual EC sites to each model relative to the mean rather than by the absolute variance further contradicts the possibility that high variance is simply due to high absolute GPP. Six models' variance at one or more sites exceeds 100% of the site's mean (Fig. 2). These six models are among the eight whose overall variance is larger than overall EC variance. Site-level variance for the model with the highest overall variance ranges from 160 to 226% of the site's mean modeled GPP. In contrast, the

model with the lowest variance as a percent of mean modeled GPP ranges from 1% to 8% at individual sites.

258 **Figure S8: Correlations of EC GPP with Process and Statistical Models**

model	BAN	CAX	K34	K67	K83	RJA	All
A	-0.07	0.17	0.45	-0.07	0.18	0.07	0.36
B	0.69	0.03	0.74	-0.42	0.50	0.77	0.39
C	0.21	-0.32	-0.22	0.09	-0.05	0.37	0.25
D	0.64	0.41	0.67	-0.36	0.42	0.12	0.03
E	0.40	-0.60	-0.16	-0.52	-0.03	0.87	0.19
G	0.82	0.12	0.43	-0.08	0.65	0.85	0.37
H	0.02	-0.54	-0.26	-0.48	-0.10	0.89	0.34
I	0.33	0.68	-0.04	-0.45	-0.19	0.11	-0.19
J	0.34	-0.61	-0.27	0.30	-0.18	0.87	0.20
S	0.25	-0.70	-0.19	-0.38	0.01	0.80	0.17
T	0.49	-0.59	0.12	-0.28	0.12	0.85	-0.05
U	0.89	-0.45	0.28	-0.38	-0.01	0.75	0.02
V	0.41	-0.55	-0.14	-0.34	0.01	0.87	0.34
W	0.61	-0.53	0.02	-0.34	0.06	0.76	0.18
X	0.58	-0.74	-0.09	-0.19	0.02	0.78	0.13
Y	0.05	-0.68	-0.59	-0.50	-0.16	0.75	-0.03
Z	0.81	-0.50	0.32	0.19	0.31	0.77	0.21
Mean	0.44	-0.32	0.06	-0.25	0.09	0.66	0.17

Figure S8: Simple correlations between a process or statistical model's GPP at a particular site to its EC estimate. Letter colors correspond to models' seasonal amplitude relative to that of the ECs, with green and blue models milder and orange and red models livelier. The mean months that a EC operated, or number of paired values per site correlation, is 43. The column on the far right, for all sites, is the variance calculated on all pairs of EC GPP with the other model, across all sites and months, not the mean of the six site-level variances. A small squared correlation suggests *prima facie* that there is only random connection between a model and EC estimate.

Figure S9: Seasonal Cycle Amplitudes for Each EC Site

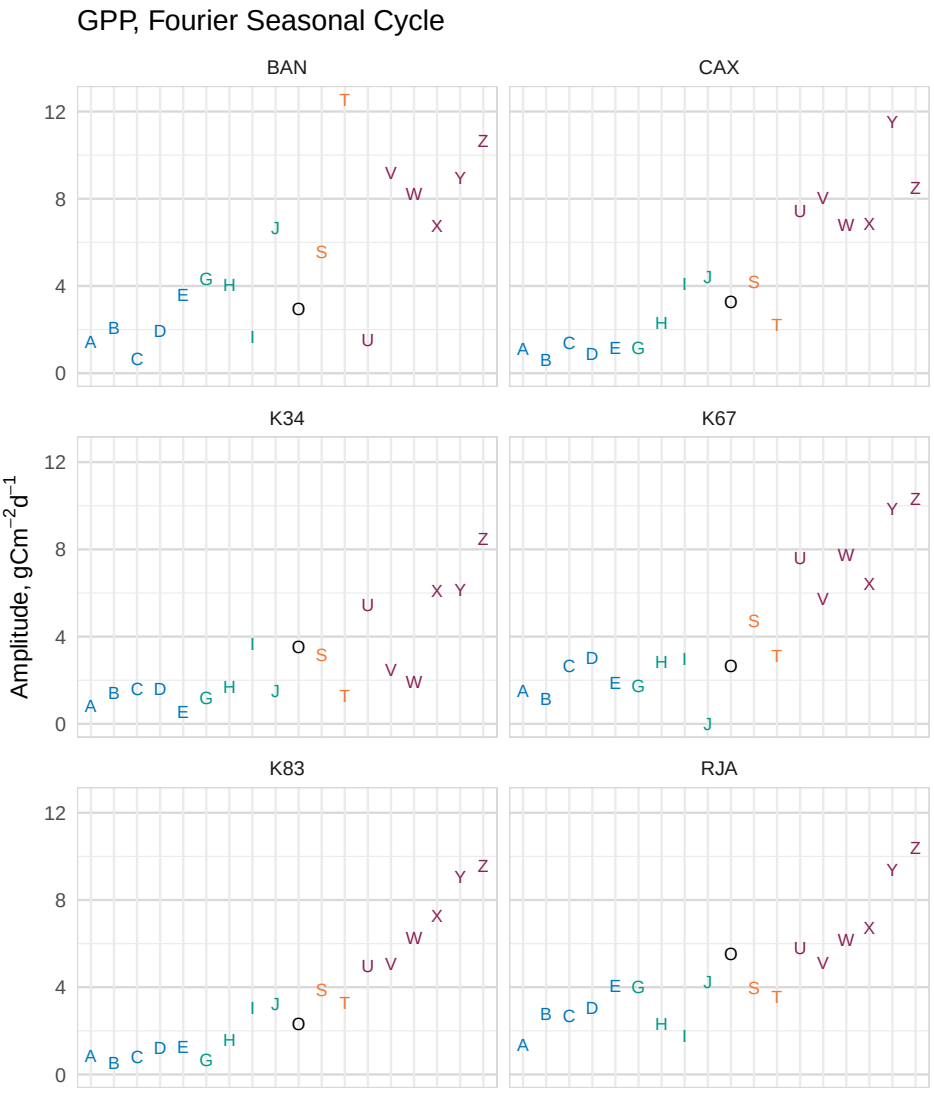


Figure S9: Parallel to the main paper's Fig. 3 of seasonal amplitudes, showing each site separately. Steadily increasing values from left to right indicate moderate similarity in site-level amplitude ranking and the ranking of mean amplitudes.

Figure S10: Yearly Mean GPP by Model

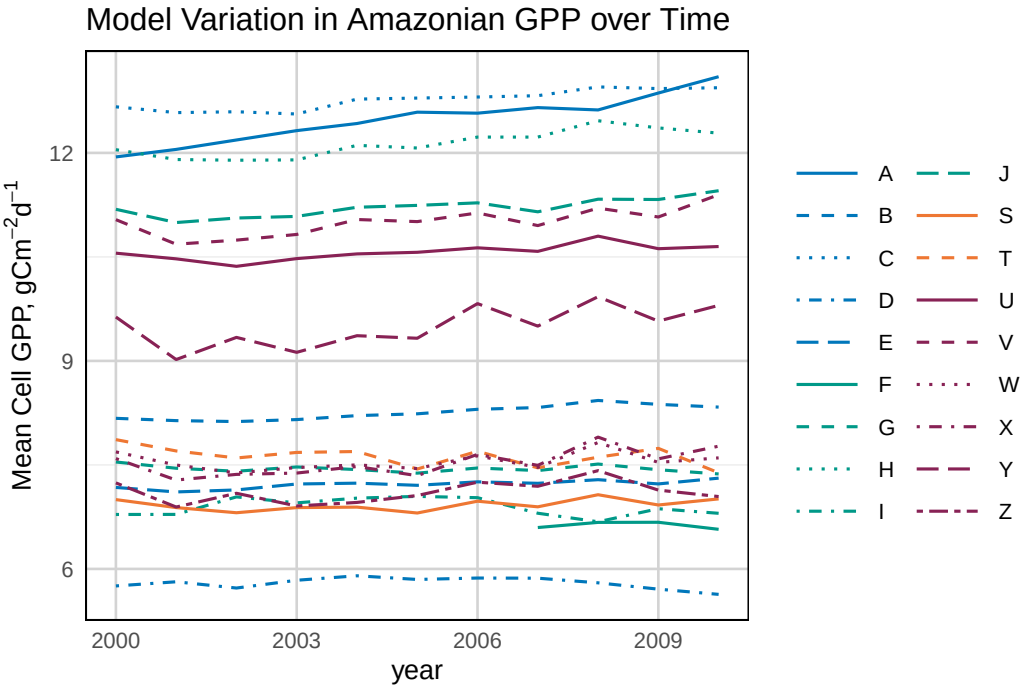


Figure S10: Yearly mean GPP for Amazon rainforest model cells. Interannual variability of individual models is much smaller than the differences among models in a particular year. Even driven by identical climate inputs, different models simulate consequentially different GPP. In every year the mean reconstructed GPP for the Amazon is over twice as high for the highest three models as for the lowest two. The differences mean that ESMs' predictions of tropical GPP several reflect not only substantial differences in meteorological predictions but also in tropical GPP's model structure and parameterization.

Figures S11 and S12: GPP Responses to Environmental Drivers Across the Entire Amazon Basin

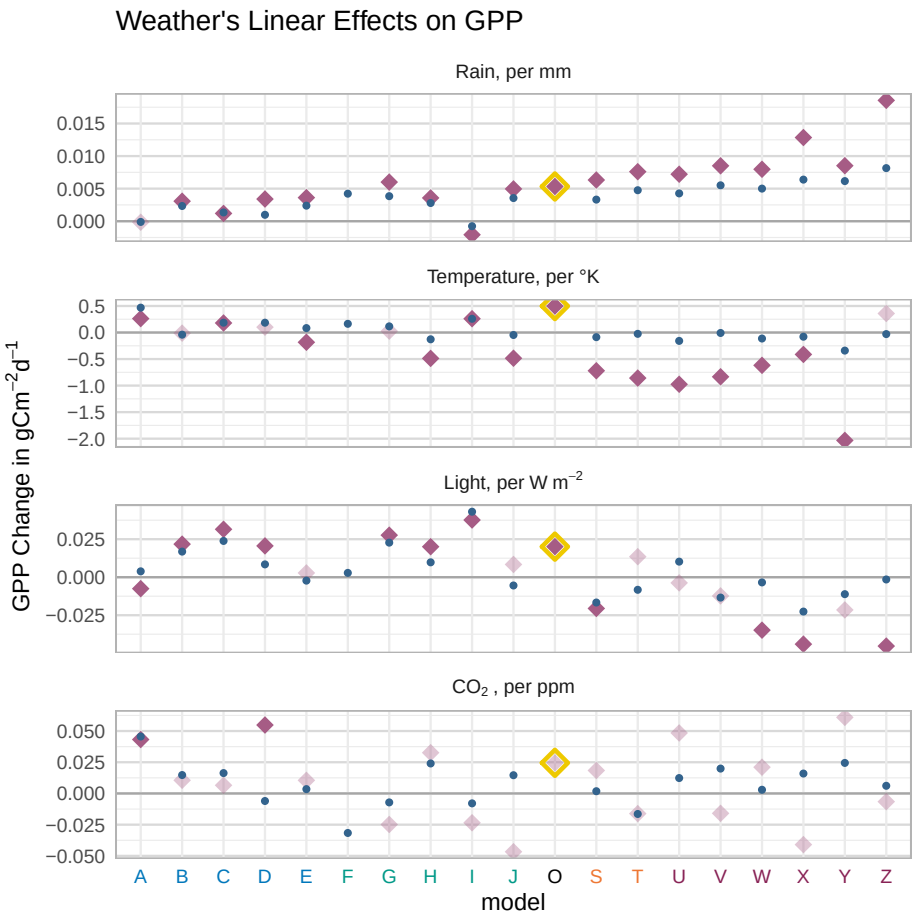


Figure S11: Fig. 5 shows each model's slope of GPP for EC cells only with driver units in z-scores. This figure summarizes tendencies also across the entire basin, displays slopes per unit value of each driver, and includes slopes for CO_2. The large purple diamonds are for the EC cells only while the blue dots are for the entire Amazon. EC estimates are highlighted with a yellow background. Site slopes with probability ≤0.05 are semi-transparent. With hundreds of cells, p≤0.05 for all basin-level predictors. Model responsiveness at EC sites generally mimics their basinwide responsiveness.

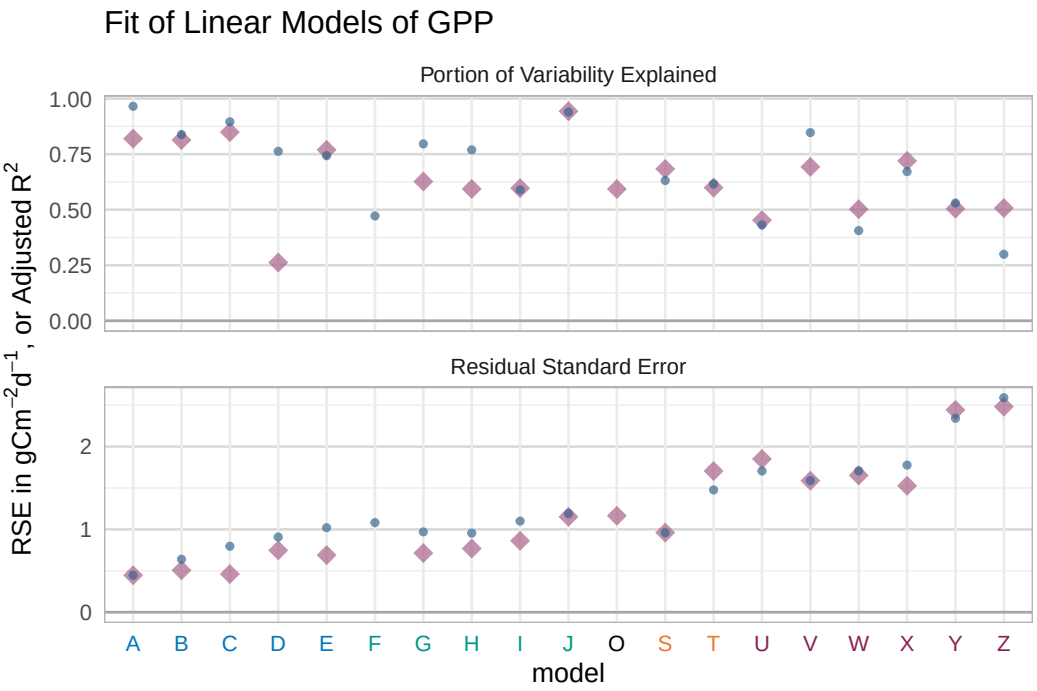


Figure S12: Please see caption for previous figure. Fits for Model F are not directly comparable because they summarize a 1° spatial grid while all other models have ½°, or approximately four times as many cells in the Amazon.

Figure S13: Maps of Driver Responsiveness by Model

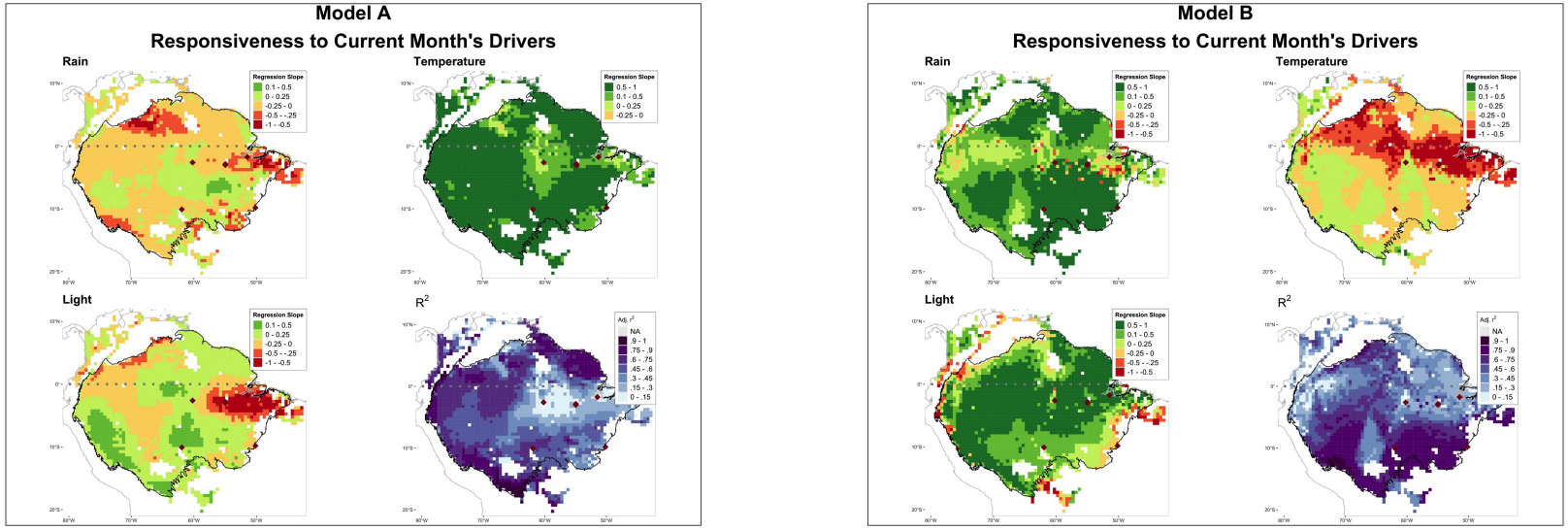
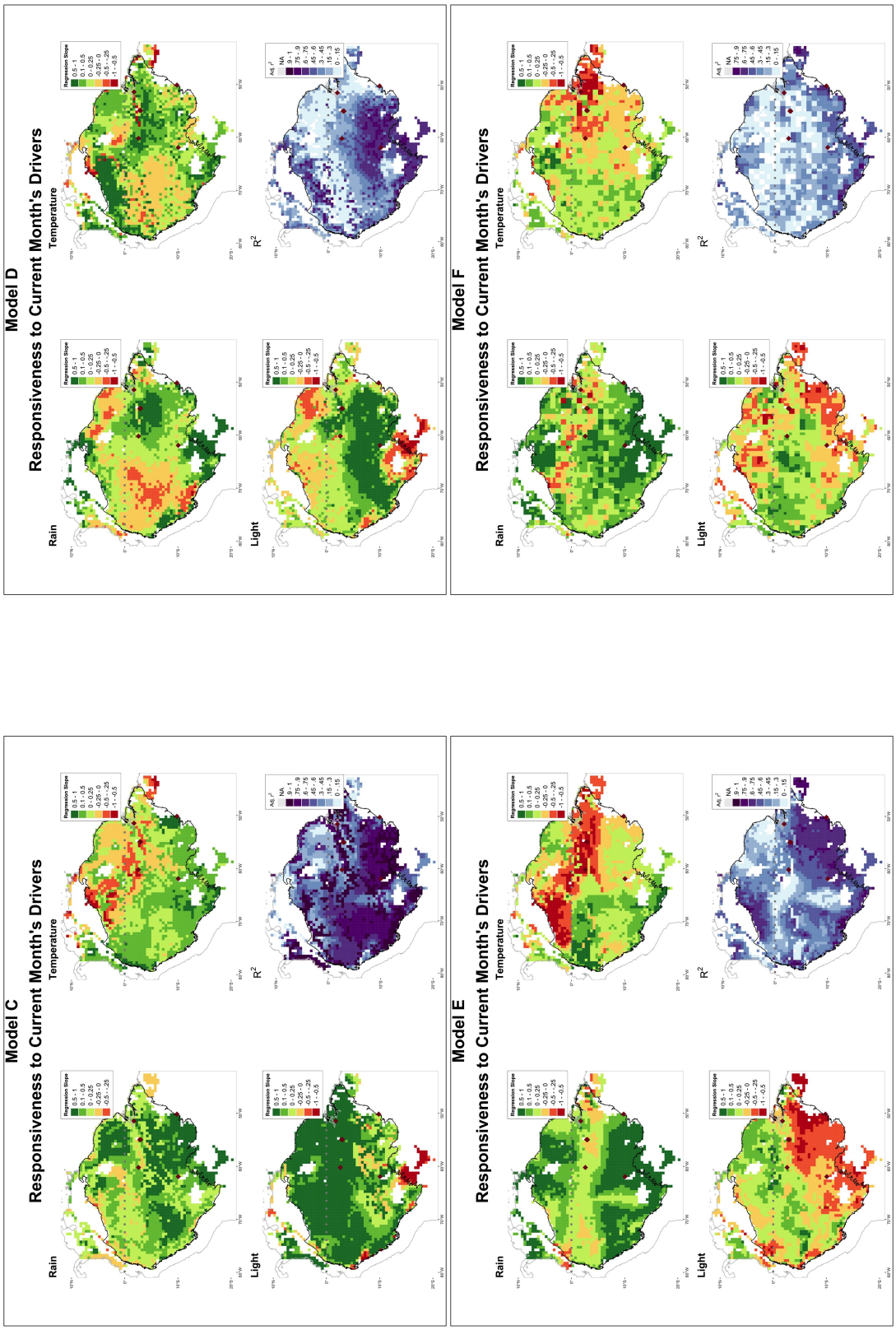
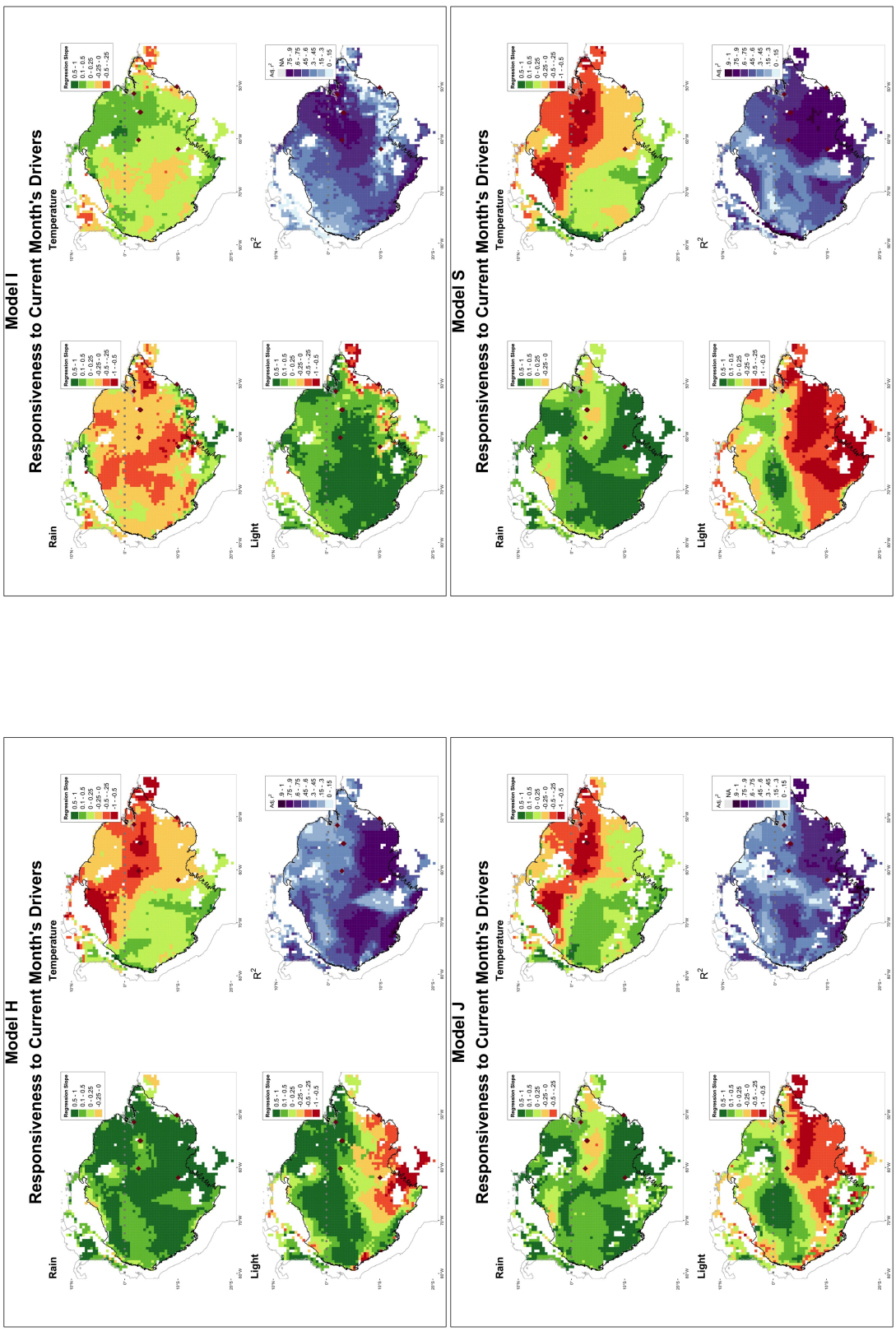


Figure S13: Spatial patterns in driver slopes and regression's portion of GPP variability explained for each model. Each model's four panels are arranged in the same locations as the panels of Fig. 5. For the EC sites in Fig. 5, the regression form allows each site to have a different intercept while driver slopes are forced to be identical at every site. For the model maps, in contrast, every cell's slopes as well as intercept are calculated individually.

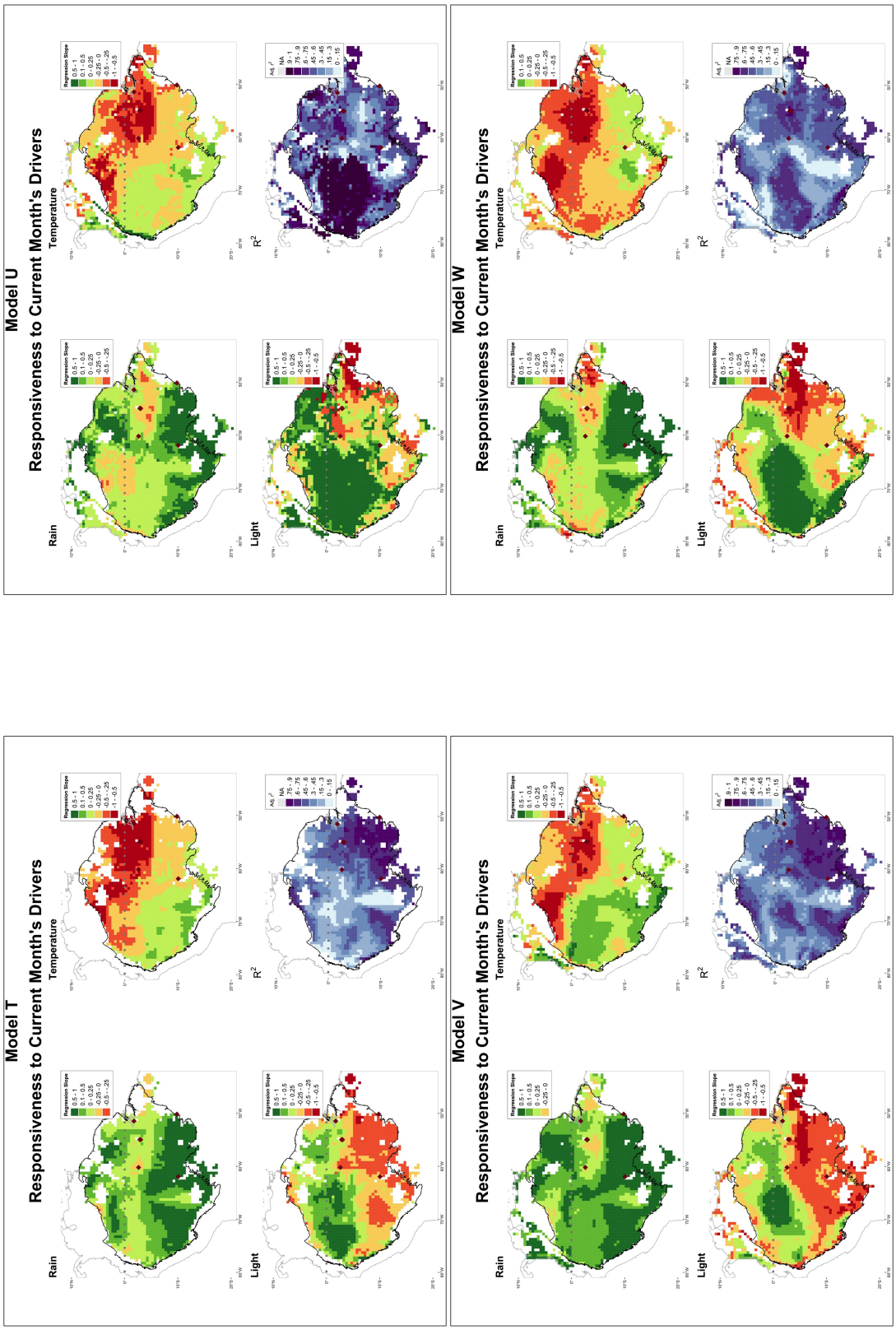
All but Models D and I respond most weakly to rain in an east-west swath south of the equator. There, twice-yearly overcrossing of the Intertropical Convergence Zone may reduce the distinctiveness of monthly mean rain. For the especially wet northwest, models respond similarly to rain, differing mostly in the extent but not general location of places where GPP increases most in the rainiest months. Temperature's spatial pattern is similar, but the central band has stronger rather than weaker response, and extends farther to the north on the west end of the basin. GPP is least responsive near the basin's periphery for models with weak seasonal cycles. For strongly seasonal models, instead there is a sharp tendency to switch from a positive response to more light north of 5°S to negative, or lower GPP with more light, farther south.



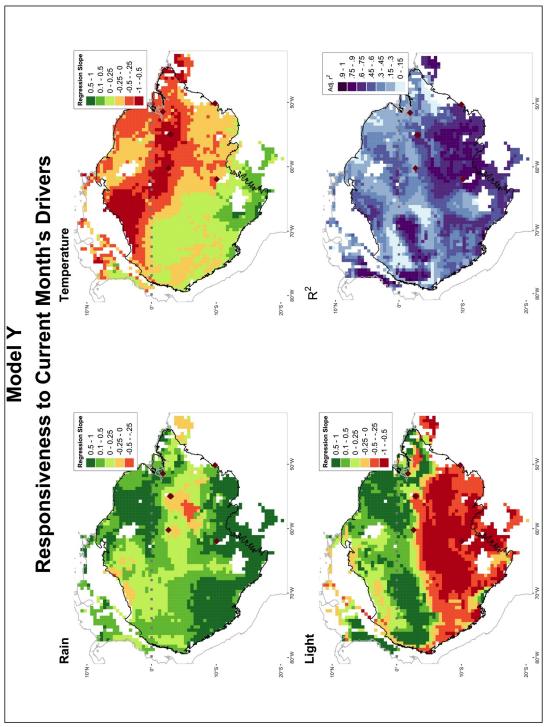
Please see caption below first figure in this section.

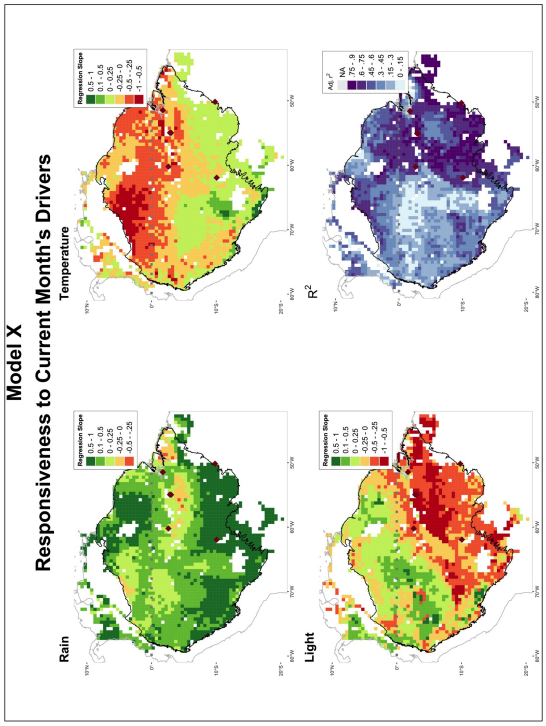


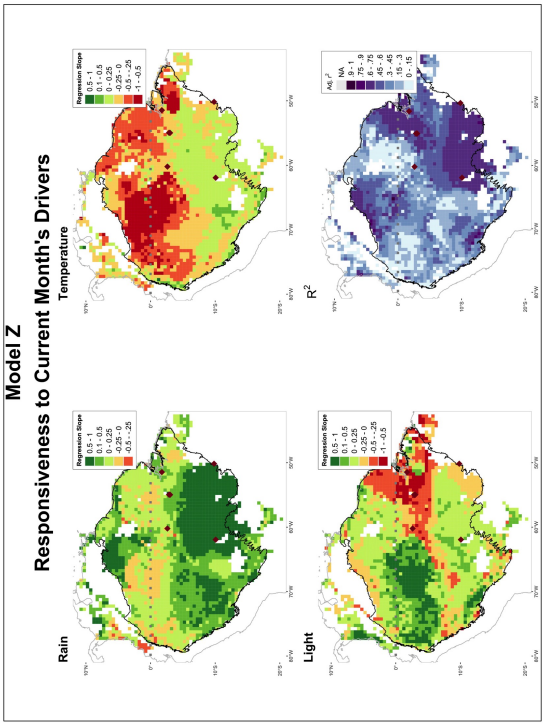
Please see caption below first figure in this section.



Please see caption below first figure in this section.







Please see caption below first figure in this section.

263 **Figure S14: Phase of Site-Level Seasonality**

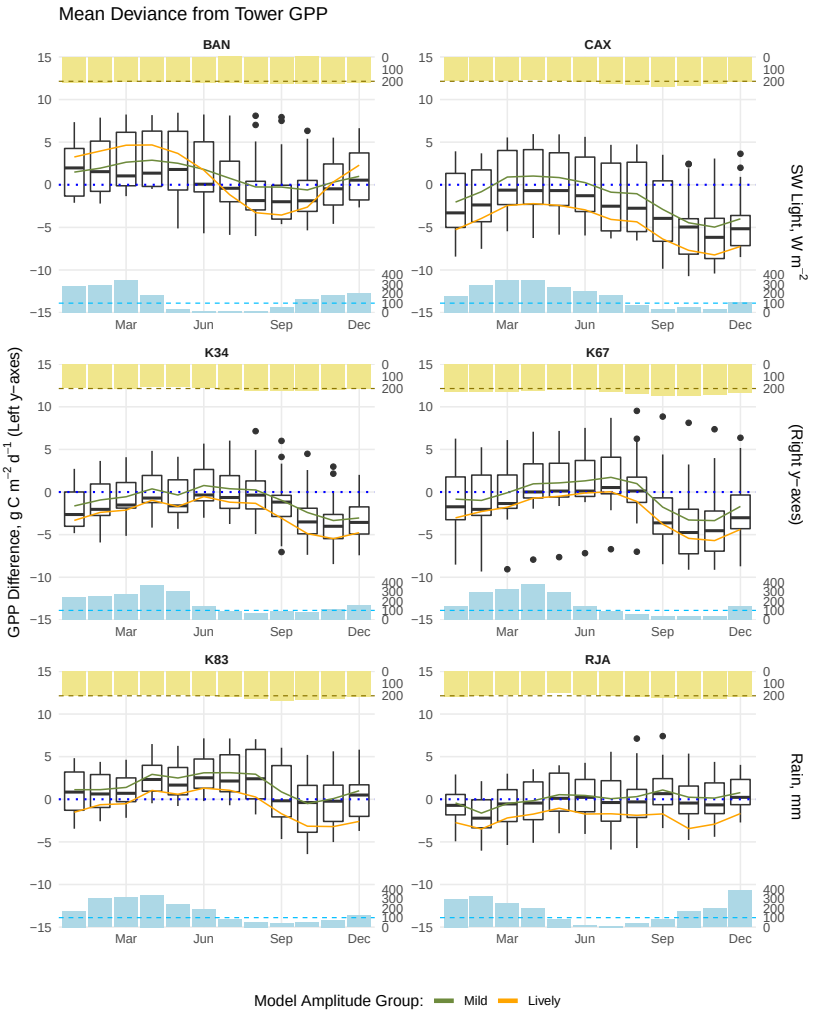


Figure S14: Parallel to the main paper's Fig. 7 for each site separately, showing deviance in modeled GPP from EC estimates. Blue and yellow bars at the top of each panel show mean monthly insolation and rainfall. Relative variations in light are small, and the line at 200 W m^{-2} is an arbitrary visual reference. The dotted line for rain defines dry season months. At most sites, divergences are largest for lively models late in the dry season.

References

Ahlström, A., J. G. Canadell, G. Schurgers, M. Wu, J. A. Berry, K. Guan, and R. B. Jackson (2017), Hydrologic resilience and Amazon productivity, *Nature Communications*, *8*(1), doi:10.1038/s41467-017-00306-z.

Albert, L. P., J. Wu, N. Prohaska, P. B. de Camargo, T. E. Huxman, E. S. Tribuzy, V. Y. Ivanov, R. S. Oliveira, S. Garcia, M. N. Smith, R. C. O. Junior, N. Restrepo-Coupe, R. da Silva, S. C. Stark, G. A. Martins, D. V. Penha, and S. R. Saleska (2018), Age-dependent leaf physiology and consequences for crown-scale carbon uptake during the dry season in an Amazon evergreen forest, *New Phytologist*, *219*(3), 870–884, doi:10.1111/nph.15056.

Albert, L. P., N. Restrepo-Coupe, M. N. Smith, J. Wu, C. Chavana-Bryant, N. Prohaska, T. C. Taylor, G. A. Martins, P. Ciais, J. Mao, M. A. Arain, W. Li, X. Shi, D. M. Ricciuto, T. E. Huxman, S. M. McMahon, and S. R. Saleska (2019), Cryptic phenology in plants: Case studies, implications, and recommendations, *Global Change Biology*, *25*(11), 3591–3608, doi:10.1111/gcb.14759.

Alemohammad, S. H., B. Fang, A. G. Konings, F. Aires, J. K. Green, J. Kolassa, D. Miralles, C. Prigent, and P. Gentine (2017), Water, Energy, and Carbon with Artificial Neural Networks (WECANN): A statistically based estimate of global surface turbulent fluxes and gross primary productivity using solar-induced fluorescence, *Biogeosciences*, *14*(18), 4101–4124, doi:10.5194/bg-14-4101-2017.

Anav, A., P. Friedlingstein, C. Beer, P. Ciais, A. Harper, C. Jones, G. Murray-Tortarolo, D. Papale, N. C. Parazoo, P. Peylin, S. Piao, S. Sitch, N. Viovy, A. Wiltshire, and M. Zhao (2015), Spatiotemporal patterns of terrestrial gross primary production: A review, *Reviews of Geophysics*, *53*(3), 785–818, doi:10.1002/2015RG000483.

Aragão, L. E. O. C., Y. Malhi, R. M. Roman-Cuesta, S. Saatchi, L. O. Anderson, and Y. E. Shimabukuro (2007), Spatial patterns and fire response of recent Amazonian droughts, *Geophysical Research Letters*, *34*(7), doi:10.1029/2006GL028946.

Baker, I., A. Harper, H. da Rocha, A. Denning, A. Araújo, L. Borma, H. Freitas, M. Goulden, A. Manzi, S. Miller, A. Nobre, N. Restrepo-Coupe, S. Saleska, R. Stöckli, C. von Randow, and S. Wofsy (2013), Surface ecophysiological behavior across vegetation and moisture gradients in tropical South America, *Agricultural and Forest Meteorology*, *182-183*, 177–188, doi:10.1016/j.agrformet.2012.11.015.

Baker, I. T., A. Denning, D. A. Dazlich, A. B. Harper, M. D. Branson, D. A. Randall, M. C. Phillips, K. D. Haynes, and S. M. Gallup (2019), Surface-Atmosphere Coupling Scale, the Fate of Water, and Ecophysiological Function in a Brazilian Forest, *Journal of Advances in Modeling Earth Systems*, *11*(8), 2523–2546, doi:10.1029/2019MS001650.

Bonal, D., B. Burban, C. Stahl, F. Wagner, and B. Hérault (2016), The response of tropical rainforests to drought—lessons from recent research and future prospects, *Annals of Forest Science*, *73*(1), 27–44, doi:10.1007/s13595-015-0522-5.

Bonan, G. B., P. J. Lawrence, K. W. Oleson, S. Levis, M. Jung, M. Reichstein, D. M. Lawrence, and S. C. Swenson (2011), Improving canopy processes in the Community Land Model version 4 (CLM4) using global flux fields empirically inferred from FLUXNET data, *Journal of Geophysical Research*, *116*(G2), G02,014, doi:10.1029/2010JG001593.

Borchert, R., G. Rivera, and W. Hagnauer (2002), Modification of Vegetative Phenology in a Tropical Semi-deciduous Forest by Abnormal Drought and Rain, *Biotropica*, *34*(1), 27–39, doi:10.1111/j.1744-7429.2002.tb00239.x.

Broedel, E., J. Tomasella, L. A. Cândido, and C. von Randow (2017), Deep soil water dynamics in an undisturbed primary forest in central Amazonia: Differences between normal years and the 2005 drought, *Hydrological Processes*, *31*(9), 1749–1759, doi:10.1002/hyp.11143.

Chapin, F. S., G. M. Woodwell, J. T. Randerson, E. B. Rastetter, G. M. Lovett, D. D. Baldocchi, D. A. Clark, M. E. Harmon, D. S. Schimel, R. Valentini, C. Wirth, J. D. Aber, J. J. Cole, M. L. Goulden, J. W. Harden, M. Heimann, R. W. Howarth, P. A. Matson, A. D. McGuire, J. M. Melillo, H. A. Mooney, J. C. Neff, R. A. Houghton, M. L. Pace, M. G. Ryan, S. W. Running, O. E. Sala, W. H. Schlesinger, and E.-D. Schulze (2006), Reconciling Carbon-cycle Concepts, Terminology, and Methods, *Ecosystems*, *9*(7), 1041–1050, doi:10.1007/s10021-005-0105-7.

Clark, D. A., and D. B. Clark (2011), Assessing Tropical Forests' Climatic Sensitivities with Long-term Data, *Biotropica*, *43*(1), 31–40, doi:10.1111/j.1744-7429.2010.00654.x.

Collier, N., F. M. Hoffman, D. M. Lawrence, G. Keppel-Aleks, C. D. Koven, W. J. Riley, M. Mu, and J. T. Randerson (2018), The International Land Model Benchmarking (ILAMB) System: Design, Theory, and Implementation, *Journal of Advances in Modeling Earth Systems*, pp. 1–24, doi:10.1029/2018MS001354.

Corlett, R. T. (2016), The Impacts of Droughts in Tropical Forests, *Trends in Plant Science*, *21*(7), 584–593, doi:10.1016/j.tplants.2016.02.003.

Dahlin, K. M., D. D. Ponte, E. Setlock, and R. Nagelkirk (2017), Global patterns of drought deciduous phenology in semi-arid and savanna-type ecosystems, *Ecography*, *40*(2), 314–323, doi:10.1111/ecog.02443.

Doughty, C. E., and M. L. Goulden (2008), Seasonal patterns of tropical forest leaf area index and CO_2 exchange, *Journal of Geophysical Research*, *113*, doi:10.1029/2007JG000590.

Feldpausch, T. R., O. L. Phillips, R. J. W. Brienen, E. Gloor, J. Lloyd, G. Lopez-Gonzalez, A. Monteagudo-Mendoza, Y. Malhi, A. Alarcón, E. Álvarez Dávila, P. Alvarez-Loayza, A. Andrade, L. E. O. C. Aragao, L. Arroyo, G. A. Aymard C., T. R. Baker, C. Baraloto, J. Barroso, D. Bonal, W. Castro, V. Chama, J. Chave, T. F. Domingues, S. Fauset, N. Groot, E. Honorio Coronado, S. Laurance, W. F. Laurance, S. L. Lewis, J. C. Licona, B. S. Marimon, B. H. Marimon-Junior, C. Mendoza Bautista, D. A. Neill, E. A. Oliveira, C. Oliveira dos Santos, N. C. Pallqui Camacho, G. Pardo-Molina, A. Prieto, C. A. Quesada, F. Ramírez, H. Ramírez-Angulo, M. Réjou-Méchain, A. Rudas, G. Saiz, R. P. Salomão, J. E. Silva-Espejo, M. Silveira, H. ter Steege, J. Stropp, J. Terborgh, R. Thomas-Caesar, G. M. F. van der Heijden, R. Vásquez Martinez, E. Vilanova, and V. A. Vos (2016), Amazon forest response to repeated droughts, *Global Biogeochemical Cycles*, *30*(7), 964–982, doi:10.1002/2015GB005133.

Gloor, M., R. J. W. Brienen, D. Galbraith, T. R. Feldpausch, J. Schöngart, J.-L. Guyot, J. C. Espinoza, J. Lloyd, and O. L. Phillips (2013), Intensification of the Amazon hydrological cycle over the last two decades, *Geophysical Research Letters*, *40*(9), 1729–1733, doi:10.1002/grl.50377.

Goulden, M. L., S. D. Miller, H. R. da Rocha, M. C. Menton, H. C. de Freitas, A. M. e Silva Figueira, and C. A. D. de Sousa (2004), Diel and Seasonal Patterns of Tropical Forest CO_2 Exchange, *Ecological Applications*, *14*(sp4), 42–54, doi:10.1890/02-6008.

Harper, A., I. T. Baker, A. S. Denning, D. A. Randall, D. Dazlich, and M. Branson (2014), Impact of Evapotranspiration on Dry Season Climate in the Amazon Forest, *Journal of Climate*, *27*(2), 574–591, doi:10.1175/JCLI-D-13-00074.1.

Huang, Y., S. Gerber, T. Huang, and J. W. Lichstein (2016), Evaluating the drought response of CMIP5 models using global gross primary productivity, leaf area, precipitation, and soil moisture data, *Global Biogeochemical Cycles*, *30*(12), 1827–1846, doi:

10.1002/2016GB005480.

Huntingford, C., P. Zelazowski, D. Galbraith, L. M. Mercado, S. Sitch, R. Fisher, M. Lomas, A. P. Walker, C. D. Jones, B. B. B. Booth, Y. Malhi, D. Hemming, G. Kay, P. Good, S. L. Lewis, O. L. Phillips, O. K. Atkin, J. Lloyd, E. Gloor, J. Zaragoza-Castells, P. Meir, R. Betts, P. P. Harris, C. Nobre, J. Marengo, and P. M. Cox (2013), Simulated resilience of tropical rainforests to CO2-induced climate change, *Nature Geoscience*, *6*(4), 268–273, doi:10.1038/ngeo1741.

Huntzinger, D., C. Schwalm, Y. Wei, R. Cook, A. Michalak, K. Schaefer, A. Jacobson, M. Arain, J. Fisher, D. Hayes, M. Huang, S. Huang, A. Ito, H. Lei, C. Lu, F. Maignan, J. Mao, N. Parazoo, C. Peng, S. Peng, B. Poulter, D. Ricciuto, H. Tian, X. Shi, W. Wang, N. Zeng, F. Zhao, and Q. Zhu (2014), NACP MsTMIP: Global 0.5-deg Terrestrial Biosphere Model Outputs in Standard Format, doi:10.3334/ORNLDAAC/1225.

Jung, M., M. Reichstein, H. A. Margolis, A. Cescatti, A. D. Richardson, M. A. Arain, A. Arneth, C. Bernhofer, D. Bonal, J. Chen, D. Gianelle, N. Gobron, G. Kiely, W. Kutsch, G. Lasslop, B. E. Law, A. Lindroth, L. Merbold, L. Montagnani, E. J. Moors, D. Papale, M. Sottocornola, F. Vaccari, and C. Williams (2011), Global patterns of land-atmosphere fluxes of carbon dioxide, latent heat, and sensible heat derived from eddy covariance, satellite, and meteorological observations, *Journal of Geophysical Research: Biogeosciences*, *116*(G3), doi:10.1029/2010JG001566.

Jung, M., S. Koirala, U. Weber, K. Ichii, F. Gans, Gustau-Camps-Valls, D. Papale, C. Schwalm, G. Tramontana, and M. Reichstein (2019), The FLUXCOM ensemble of global land-atmosphere energy fluxes, *Scientific Data*, *6*(1), 74.

Jupp, T. E., P. M. Cox, A. Rammig, K. Thonicke, W. Lucht, and W. Cramer (2010), Development of probability density functions for future South American rainfall, *New Phytologist*, *187*(3), 682–693, doi:10.1111/j.1469-8137.2010.03368.x.

Li, W., R. Fu, and R. E. Dickinson (2006), Rainfall and its seasonality over the Amazon in the 21st century as assessed by the coupled models for the IPCC AR4, *Journal of Geophysical Research: Atmospheres*, *111*(D2), doi:10.1029/2005JD006355.

Malavelle, F. F., J. M. Haywood, L. M. Mercado, G. A. Folberth, N. Bellouin, S. Sitch, and P. Artaxo (2019), Studying the impact of biomass burning aerosol radiative and climate effects on the Amazon rainforest productivity with an Earth system model, *Atmospheric Chemistry and Physics*, *19*(2), 1301–1326, doi:10.5194/acp-19-1301-2019.

Malhi, Y., and J. Wright (2004), Spatial patterns and recent trends in the climate of tropical rainforest regions, *Philosophical Transactions of the Royal Society B: Biological Sciences*, *359*(1443), 311–329, doi:10.1098/rstb.2003.1433.

Morton, D. C., J. Rubio, B. D. Cook, J.-P. Gastellu-Etchegorry, M. Longo, H. Choi, M. Hunter, and M. Keller (2016), Amazon forest structure generates diurnal and seasonal variability in light utilization, *Biogeosciences*, *13*(7), 2195–2206, doi:10.5194/bg-13-2195-2016.

Mystakidis, S., E. L. Davin, N. Gruber, and S. I. Seneviratne (2016), Constraining future terrestrial carbon cycle projections using observation-based water and carbon flux estimates, *Global Change Biology*, *22*(6), 2198–2215, doi:10.1111/gcb.13217.

Oleson, K. W., and D. M. Lawrence (2013), Technical description of version 4.5 of the Community Land Model (CLM).

Parazoo, N. C., K. Bowman, J. B. Fisher, C. Frankenberg, D. B. A. Jones, A. Cescatti, O. Pérez-Priego, G. Wohlfahrt, and L. Montagnani (2014), Terrestrial gross primary production inferred from satellite fluorescence and vegetation models, *Global Change Biology*, *20*(10), 3103–3121, doi:10.1111/gcb.12652.

Piao, S., S. Sitch, P. Ciais, P. Friedlingstein, P. Peylin, X. Wang, A. Ahlström, A. Anav, J. G. Canadell, N. Cong, C. Huntingford, M. Jung, S. Levis, P. E. Levy, J. Li, X. Lin, M. R. Lomas, M. Lu, Y. Luo, Y. Ma, R. B. Myneni, B. Poulter, Z. Sun, T. Wang, N. Viovy, S. Zaehle, and N. Zeng (2013), Evaluation of terrestrial carbon cycle models for their response to climate variability and to CO $_2$ trends, *Global Change Biology*, *19*(7), 2117–2132, doi:10.1111/gcb.12187.

Poulter, B., L. Aragão, U. Heyder, M. Gumpenberger, J. Heinke, F. Langerwisch, A. Rammig, K. Thonicke, and W. Cramer (2010), Net biome production of the Amazon Basin in the 21st century, *Global Change Biology*, *16*(7), 2062–2075, doi:10.1111/j.1365-2486.2009.02064.x.

Rogers, A., B. E. Medlyn, J. S. Dukes, G. Bonan, S. von Caemmerer, M. C. Dietze, J. Kattge, A. D. B. Leakey, L. M. Mercado, Ü. Niinemets, I. C. Prentice, S. P. Serbin, S. Sitch, D. A. Way, and S. Zaehle (2017), A roadmap for improving the representation of photosynthesis in Earth system models, *New Phytologist*, *213*(1), 22–42, doi:10.1111/nph.14283.

Saleska, S. R., S. D. Miller, D. M. Matross, M. L. Goulden, S. C. Wofsy, H. R. da Rocha, P. B. de Camargo, P. Crill, B. C. Daube, H. C. de Frietas, L. Hutyra, M. Keller, V. Krichhoff, M. Menton, J. W. Munger, E. Hammond Pyle, A. H. Rice, and H. Silva (2003), Car-

bon in Amazon Forests: Unexpected Seasonal Fluxes and Disturbance-Induced Losses, *Science*, *302*(5650), 1554–1557, doi:10.1126/science.1091165.

Samanta, A., Y. Knyazikhin, L. Xu, R. E. Dickinson, R. Fu, M. H. Costa, S. S. Saatchi, R. R. Nemani, and R. B. Myneni (2012), Seasonal changes in leaf area of Amazon forests from leaf flushing and abscission, *Journal of Geophysical Research: Biogeosciences*, *117*(G1), doi:10.1029/2011JG001818.

Tramontana, G., M. Jung, C. R. Schwalm, K. Ichii, G. Camps-Valls, B. Ráduly, M. Reichstein, M. A. Arain, A. Cescatti, G. Kiely, L. Merbold, P. Serrano-Ortiz, S. Sickert, S. Wolf, and D. Papale (2016), Predicting carbon dioxide and energy fluxes across global FLUXNET sites with regression algorithms, *Biogeosciences*, *13*(14), 4291–4313, doi:10.5194/bg-13-4291-2016.

Wei, Y., S. Liu, D. Huntzinger, A. Michalak, N. Viovy, W. Post, C. Schwalm, K. Schaefer, A. Jacobson, C. Lu, H. Tian, D. Ricciuto, R. Cook, J. Mao, and X. Shi (2014), NACP MsTMIP: Global and North American Driver Data for Multi-Model Intercomparison, doi:10.3334/ORNLDAAC/1220.

Wilson, K. B., D. D. Baldocchi, and P. J. Hanson (2001), Leaf age affects the seasonal pattern of photosynthetic capacity and net ecosystem exchange of carbon in a deciduous forest, *Plant, Cell & Environment*, *24*(6), 571–583, doi:10.1046/j.0016-8025.2001.00706.x.

Wu, J., L. P. Albert, A. P. Lopes, N. Restrepo-Coupe, M. Hayek, K. T. Wiedemann, K. Guan, S. C. Stark, B. Christoffersen, N. Prohaska, J. V. Tavares, S. Marostica, H. Kobayashi, M. L. Ferreira, K. S. Campos, R. da Silva, P. M. Brando, D. G. Dye, T. E. Huxman, A. R. Huete, B. W. Nelson, and S. R. Saleska (2016), Leaf development and demography explain photosynthetic seasonality in Amazon evergreen forests, *Science*, *351*(6276), 972–976, doi:10.1126/science.aad5068.

Xiao, X., Q. Zhang, S. Saleska, L. Hutyra, P. De Camargo, S. Wofsy, S. Frolking, S. Boles, M. Keller, and B. Moore (2005), Satellite-based modeling of gross primary production in a seasonally moist tropical evergreen forest, *Remote Sensing of Environment*, *94*(1), 105–122, doi:10.1016/j.rse.2004.08.015.

Xu, L., S. S. Saatchi, Y. Yang, R. B. Myneni, C. Frankenberg, D. Chowdhury, and J. Bi (2015), Satellite observation of tropical forest seasonality: Spatial patterns of carbon exchange in Amazonia, *Environmental Research Letters*, *10*(8), 084,005, doi:10.1088/1748-9326/10/8/084005.

458 Zhang, Y., X. Xiao, C. Jin, J. Dong, S. Zhou, P. Wagle, J. Joiner, L. Guanter, Y. Zhang,
459 G. Zhang, Y. Qin, J. Wang, and B. Moore (2016), Consistency between sun-induced
460 chlorophyll fluorescence and gross primary production of vegetation in North America,
461 *Remote Sensing of Environment*, *183*, 154–169, doi:10.1016/j.rse.2016.05.015.
462 Zhang, Y., X. Xiao, X. Wu, S. Zhou, G. Zhang, Y. Qin, and J. Dong (2017), A global moder-
463 ate resolution dataset of gross primary production of vegetation for 2000–2016, *Scientific*
464 *Data*, *4*, 170,165, doi:10.1038/sdata.2017.165.