Supplement

Study Site

Bartlett Experimental Forest (44° 17´N, 71° 03´W) is a US Forest Service research forest located outside of Bartlett, NH in the White Mountains (Lee et al., 2018). Species composition is typical of northern hardwood forests and consists predominantly of *Acer rubrum* (red maple), *Fagus grandifolia* (American beech), *Betula papyrifera* (paper birch), and *Tsuga canadensis* (eastern hemlock). Climate is also typical of central New England with short summers (20 °C) and long cold winters (-8 °C). The site is generally moist, receiving approximately 1300 mm/yr of precipitation. Soils are sandy loam Spodosols and can become saturated during spring snowmelt.

An eddy-covariance tower (26.5m) was installed in November 2003 at a lowland site (272m) within the experimental forest (Richardson et al., 2007). Topography near the eddy-covariance tower is flat to gently sloping but larger hills (1-3 km distant) surround the site. Canopy height is 19m with a mean stand age of approximately 100 yr. The eddy-covariance system consists of a LI-6262 CO_2/H_2O infra-red gas analyser (LiCor, Lincoln, NE) and SAT-211/3K 3-axis sonic anemometer (Applied Technologies, Longmont, Colo.). Measurements were made at 5 Hz and fluxes were estimated every 30 minutes. The meteorological data used in this analysis were derived from measurements made at the eddy-covariance tower for years 2005-2006. These include air temperature above the canopy (22.3 m), soil temperature, relative humidity, precipitation, above canopy PAR and wind speed.

The Bartlett tower footprint contains twelve vegetation inventory plots that follow the Forest Inventory and Analysis (FIA) design consisting of four circular 10 m radius subplots: one central and three evenly spaced at a radius of 36.5 meters. Vegetation plots were established in May 2004 and used to initialize ED2. Bradford et al. (2010) provided soil carbon and live aboveground biomass estimates for Bartlett which we used to initialize SIPNET.

Soil respiration measurements were made manually in each plot (n=12) at permanently installed rings that are 10cm in diameter using a soil CO2 flux chamber (LiCOR 6400-9). Soil temperature and moisture were measured concurrently using a soil temperature probe and a TDR probe. During 2006, soil respiration censuses were made approximately every 4-5 days from day 138 to day 325 for a total of 39 chamber censuses.

SIPNET Model

The simplified Photosynthesis and Evapotranspiration model (SIPNET) is a simple ecosystem model which can be used to interpret carbon water exchange between vegetation and the atmosphere. SIPNET has been developed from the PnET family of models (Aber & Federer 1992) to facilitate model comparisons to flux towers (Braswell et al., 2005; Sacks et al., 2007). SIPNET runs at a half-hourly time step. It represents relatively few processes (has two vegetation carbon pools, a single aggregated soil carbon pool, and a simple soil moisture sub-model), making it easier to evaluate which data contributes how much to the parameterization of each process. As a result of this setup, SIPNET is a fast model (~ 5.5 sec per MCMC iteration in PEcAn including model execution, and writing and reading model outputs), which makes it suitable for application of bruteforce methods.

Forest inventory data collected in the tower footprint were used to set initial conditions in SIPNET. We fitted Bayesian models using the allometric equations available in the literature (Jenkins et al., 2004) to estimate the aboveground biomass at Bartlett through PEcAn's allometry module (Dietze et al., *in prep.*). These values were in agreement with live aboveground biomass estimates by Bradford et al. (2010) whose soil carbon pool estimates were also used to set the initial values in our SIPNET runs (Table S1).

Table S1. Initial state values used for SIPNET runs.

Pool	Value	Units
Above- and below-ground woody biomass	9600	gC / m^2 ground area
Initial leaf area	0	m^2 leaves / m^2 ground area
Litter biomass	200	gC / m^2 ground area
Soil biomass	1600	gC / m^2 ground area

Table S2. The prior and posterior distributions of the constrained SIPNET parameters.

Parameter	Prior	Posterior (Emulator)	Posterior (Bruteforce)
SOM Respiration rate	unif(0.003, 0.6)	weibull(1.62, 0.13)	norm(0.1, 0.009)
Soil Respiration Q10	unif(1.4, 5.0)	lnorm(0.697, 0.24)	lnorm(0.39, 0.046)
Soil WHC	unif(0.1, 36.0)	lnorm(2.95, 0.31)	lnorm(2.7, 0.035)
Half saturation PAR	unif(4.0, 27.0)	weibull(3.74, 17.5)	lnorm(2.8, 4.5e-02)
dVPDSlope	unif(0.01, 0.25)	weibull(2.26, 7e-02)	norm(0.08, 2.6e-03)
Seasonal leaf growth	unif(0.0, 252.0)	norm(150.6, 46.8)	norm(145, 10.8)
psnTOpt	unif(5.0, 40.0)	weibull(12.07, 35.7)	weibull(336, 39.9)
Leaf turnover rate	unif(0.03, 10.0)	norm(5.14, 1.9)	lnorm(1.64, 5e-02)

Table S3. Calibrated SIPNET parameters and the 'true' values used to produce the synthetic data.

Parameter	**Definition**	**Units**	**True Values**
SOM Respiration rate	Soil organic matter respiration rate coefficient	Day-1	0.01
Opt. photosynthesis T	Optimum temperature for photosynthesis	Celcius	38.73
Soil Respiration Q10	Scalar determining effect of temperature on soil heterotrophic respiration	ratio	2.59
Soil WHC	Soil water holding capacity	cm	25.75
Seasonal leaf growth	Amount of leaf growth following leaf-out	gC m-2	212.55
Leaf turnover rate	Average turnover rate of leaves	y-1	2.56
Slope-VPD	Slope of VPD-photosynthesis relationship	kPa-1	0.06
Half saturation PAR	Photosynthetically active radiation at which photosynthesis occurs at 1/2 theoretical maximum	Einsteins m-2 day-1	6.46

Ecosystem Demography Model

The Ecosystem Demography model version 2.1 (ED2) is a terrestrial biosphere model that couples plant community dynamics to biogeochemical models of associated soil fluxes of carbon, water, and nitrogen (Moorcroft et al. 2001, Medvigy et al. 2009). ED2 is explicitly designed to scale from the individual to the region and to account for community processes, such as disturbance and resource competition, in a manner analogous to forest gap models. ED2 achieves this with a size and age structured (SAS) approximation to a forest gap model which accounts for the vertical size distribution within a stand/patch and the distribution of different stand ages across the landscape. This hierarchical SAS allows ED to be compared to data operating at multiple scales but in practice this means that a single ED run will simulate a large number of different patches, each with a number of trees of different sizes and species. The resulting computational expenses and complexity of drivers and outputs make ED2 an ideal example of the challenges of model-data fusion. The initialization of vegetation and soil for ED2 was done using the same forest inventory data and soil carbon measurements described for SIPNET. The species occurring in the inventory data were mapped to ED2 PFTs following Dietze and Moorcroft (2011).

Table S4. The PDA prior (meta-analysis posterior) approximated parametric distributions of the targeted ED2 parameters. t.EH: temperate Early Hardwood, t.LC: temperate Late Conifer, t.LH: temperate Late Hardwood, t.NMH: temperate North Mid-Hardwood, t.NP: temperate Northern Pine

Plant Functional Type Physiological Parameters					
	t.EH	t.LC	t.LH	t.NMH	t.NP
stomatal slope	gamma(19.7, 2.97)	weibull(2, 10)	weibull(2, 10)	weibull(2, 10)	weibull(2, 10)
quantum efficiency	gamma(16.6, 279)	norm(0.08, 0.014)	weibull(2.9, 0.07)	lnorm(-3.28, 0.08)	gamma(82, 1.4e+03)
Vcmax	norm(74.9, 9.8)	weibull(1.7, 80)	norm(60.5, 11.9)	gamma(37.8, 0.53)	weibull(2.2, 80)
cuticular conductance	lnorm(9.4, 0.7)	lnorm(9.4, 0.7)	lnorm(9.4, 0.7	norm(9988, 497)	lnorm(9.4, 0.7)
growth respiration f.	beta(4.06, 7.2)	beta(2.63, 6.52)	beta(4.06, 7.2)	beta(2.63, 6.52)	beta(2.63, 6.52)
fine root allocation	gamma(16.59, 23.32)	lnorm(-0.25, 1)	gamma(9.13, 8.22)	gamma(9.44, 8.82)	lnorm(-0.25, 1)
Soil Biogeochemistry (decomposition) parameters					
r_stsc	beta(1, 1)				
decay rate stsc	unif(0.005, 0.75)				
resp tem. increase	unif(0.05, 0.2)				

Table S5. The emulator-PDA approximated parametric posterior distributions of the targeted ED2 parameters.

Plant Functional Type Physiological Parameters					
	t.EH	t.LC	t.LH	t.NMH	t.NP
stomatal slope	lnorm(1.48, 0.13)	gamma(4.01, 1.6)	gamma(4.01, 1.6)	gamma(4.01, 1.6)	gamma(4.01, 1.6)
quantum efficiency	lnorm(-2.8, 0.11)	norm(0.08, 6.3e-03)	gamma(35.8, 541)	lnorm(-3.3, 0.04)	lnorm(-2.8, 0.05)
Vcmax	norm(47.3, 3.45)	gamma(2.83, 1.04)	norm(27.1, 4.17)	norm(42.9, 2.85)	weibull(2.4, 6.4)
cuticular conductance	lnorm(9.85, 0.385)	lnorm(9.85, 0.385)	lnorm(9.85, 0.385)	norm(10308, 273)	lnorm(9.85, 0.385)
growth respiration f.	beta(3.59, 7.47)	beta(2.29, 6.8)	beta(3.59, 7.47)	beta(2.29, 6.8)	beta(2.29, 6.8)
fine root allocation	gamma(30.7, 7.47)	lnorm(-0.3, 0.73)	gamma(16.7, 15.6)	gamma(17.3, 16.8)	lnorm(-0.3, 0.73)
Soil Biogeochemistry (decomposition) parameters					
r_stsc	beta(1, 1.98)				
decay rate stsc	lnorm(-2.97, 1.02)				
resp tem. increase	lnorm(-2.16, 0.28)				

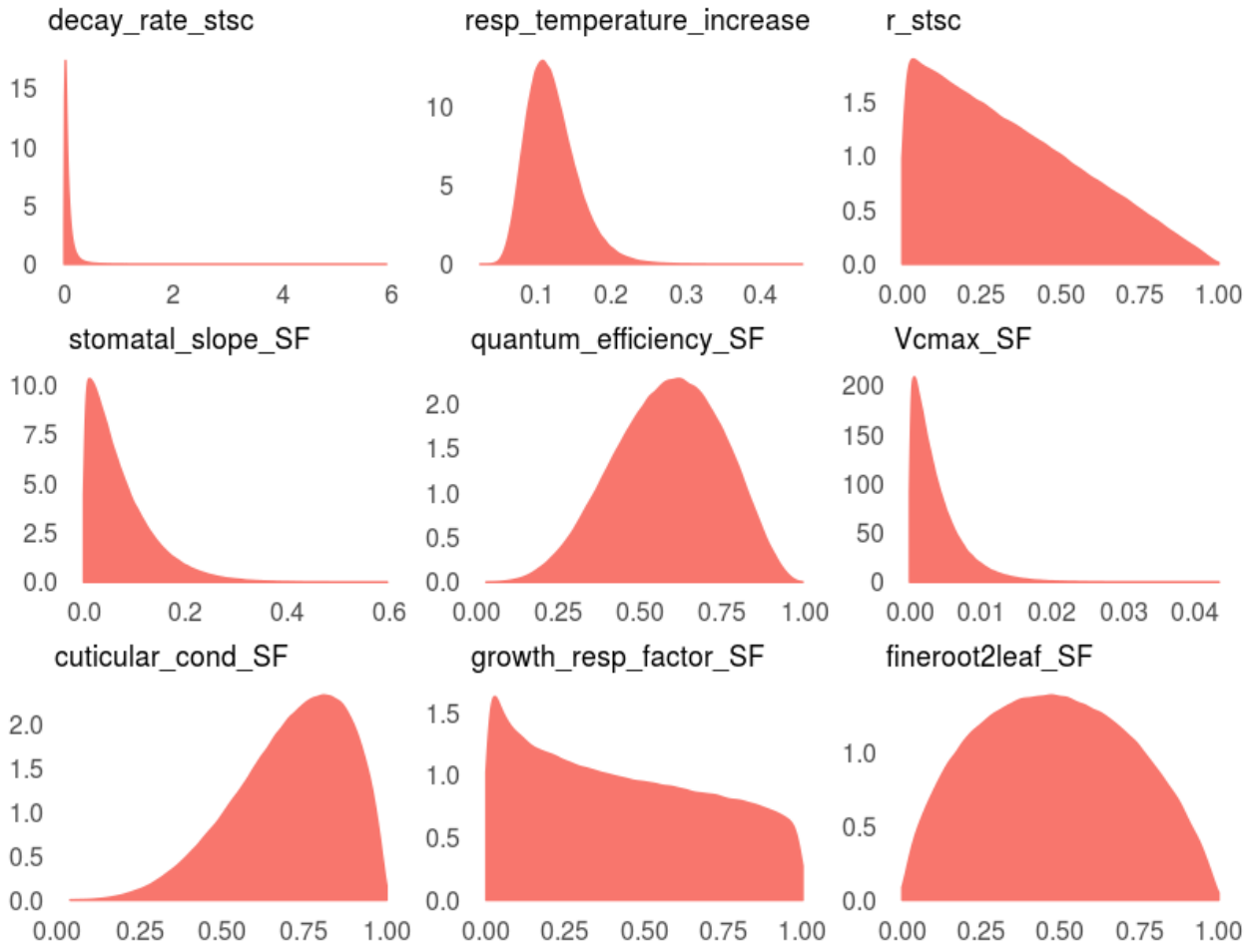


Figure S1. ED decomposition and scaling factor posteriors density distributions. Parameters common to all ED2 PFTs ending with suffix "SF" were targeted through the scaling factor.

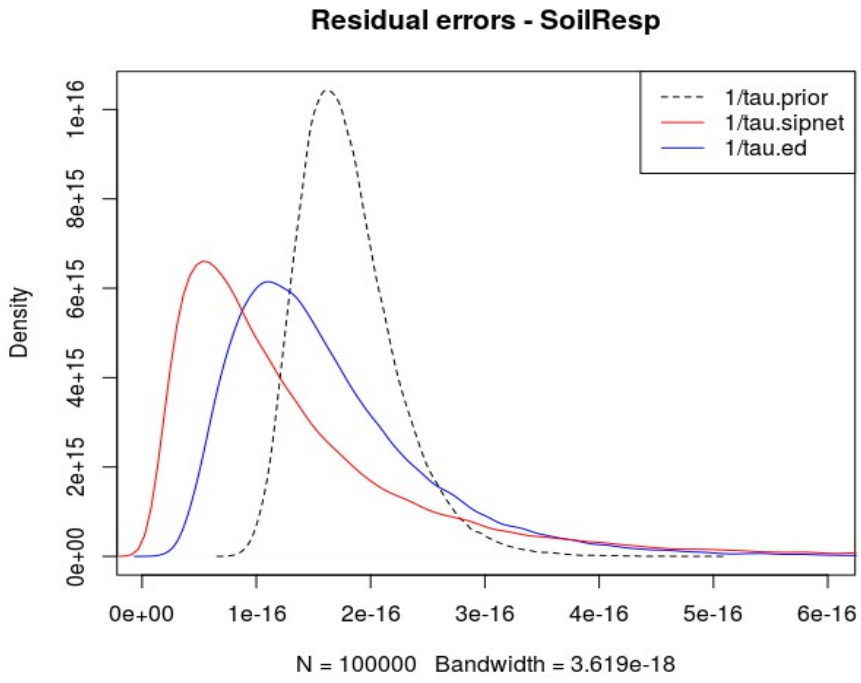


Figure S2. Posterior probability density distribution of variance (reciprocal of the precision, 1/τ) parameter of the Soil Respiration likelihood after emulator-PDA.

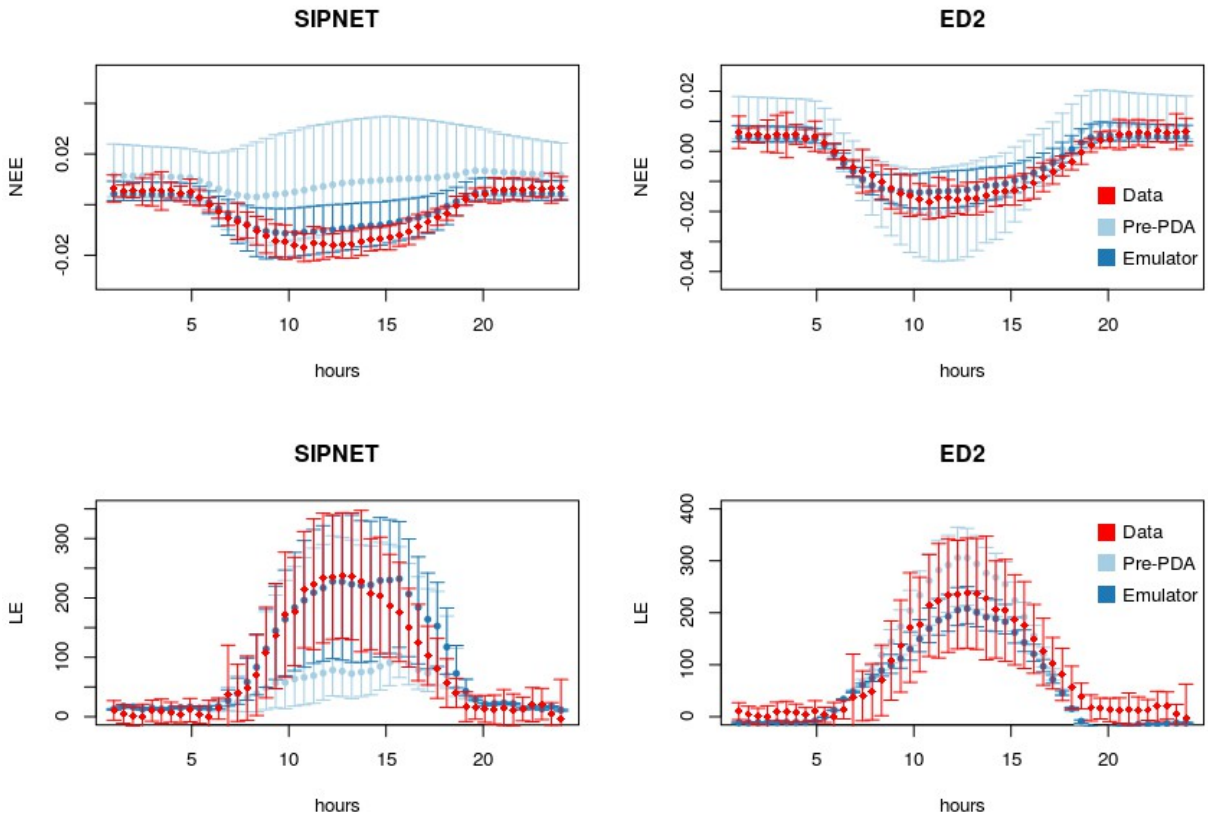


Figure S3. Diurnal cycles of NEE and LE fluxes for June-July-August months over the simulation period (2005-2006) before and after the calibration. Error bars represent the variation over the JJA period.

Table S6. Links to the Workflow IDs. The input/output files associated with each workflow can be accessed via the history table on the following link "pecan2.bu.edu/pecan/history.php". Alternatively, each workflowID can be replaced with the workflowID of the following link to directly access the workflow page:

"pecan2.bu.edu/pecan/08-finished.php?workflowid=1000008379"

The left menu on the page can be used to navigate through PEcAn settings, input and output files.

PDA: Parameter Data Assimilation, EA: Ensemble Analysis, UA: Uncertainty Analysis

Model	Experiment	Workflow ID
SIPNET	Pre-PDA EA / UA	1000008379
SIPNET	Emulator PDA - Synthetic	1000008974
SIPNET	Emulator PDA - Real Data	1000008503
SIPNET	Emulator Post-PDA EA	1000008503
SIPNET	Bruteforce PDA - Real Data	1000008530
		1000008531
		1000008532
SIPNET	Bruteforce Post-PDA EA	1000008923
ED2	Pre-PDA EA / UA	1000009051
ED2	Emulator PDA - Real Data	1000009052
ED2	Emulator Post-PDA EA	1000009052

Table S7. Links to the Workflow IDs of scaling experiments. Parameters targeted are in this order cumulatively: som_respiration_rate, soil_respiration_Q10, soilWHC, psnTOpt (4), leafGrowth, leaf_turnover_rate (6), half_saturation_PAR, dVPDSlope (8), AmaxFrac, dVpdExp (10)

Model	# of params	# of knots	Workflow ID
SIPNET	4	960	1000008978
		480	1000008979
		240	1000008980
		120	1000008981
SIPNET	6	960	1000008942
		480	1000008975
		240	1000008976
		120	1000008977
SIPNET	8	960	1000008938
		480	1000008939
		240	1000008940
		120	1000008941
SIPNET	10	960	1000008993
		480	1000008991
		240	1000008994
		120	1000008995

Table S8. Scaling experiment results showing the trade-off between wall-clock time vs. the approximation error with increasing emulator knots.

m parameters (*m = {4, 6, 8, 10}*)

k knots (*k = {120, 240, 480, 960}*)

m	k	Model run time (sec)			GP fitting (sec)			100K MCMC (sec)			Deviance
		1st	2nd	3rd	1st	2nd	3rd	1st	2nd	3rd	
4	120	182	188	184	2	4	12	772	948	1144	9489
	240	366	364	359	5	27	92	941	1340	1764	9255
	480	733	748	744	28	228	707	1592	2502	3614	9230
	960	1453	1511	1505	204	1736	6615	2523	4862	7815	9308
6	120	182	180	185	2	6	14	795	1017	1221	8371
	240	365	368	366	5	27	85	1039	1519	1962	8284
	480	735	777	737	28	215	731	1544	2488	3675	8310
	960	1521	1471	1514	209	1785	6858	2360	4503	7799	8150
8	120	197	199	198	2	5	12	905	1116	1323	9825
	240	410	392	392	7	32	109	1152	1611	2107	8643
	480	745	749	754	30	236	747	1625	2596	3766	8100
	960	1517	1532	1502	217	1949	6678	2532	4827	7498	8062
10	120	187	187	187	2	7	15	988	1254	1277	9573
	240	376	368	418	5	29	92	1235	1610	2075	8682
	480	752	769	766	26	204	787	1681	2732	3489	8559
	960	1491	1507	1490	208	2015	6643	2721	5010	7831	8106

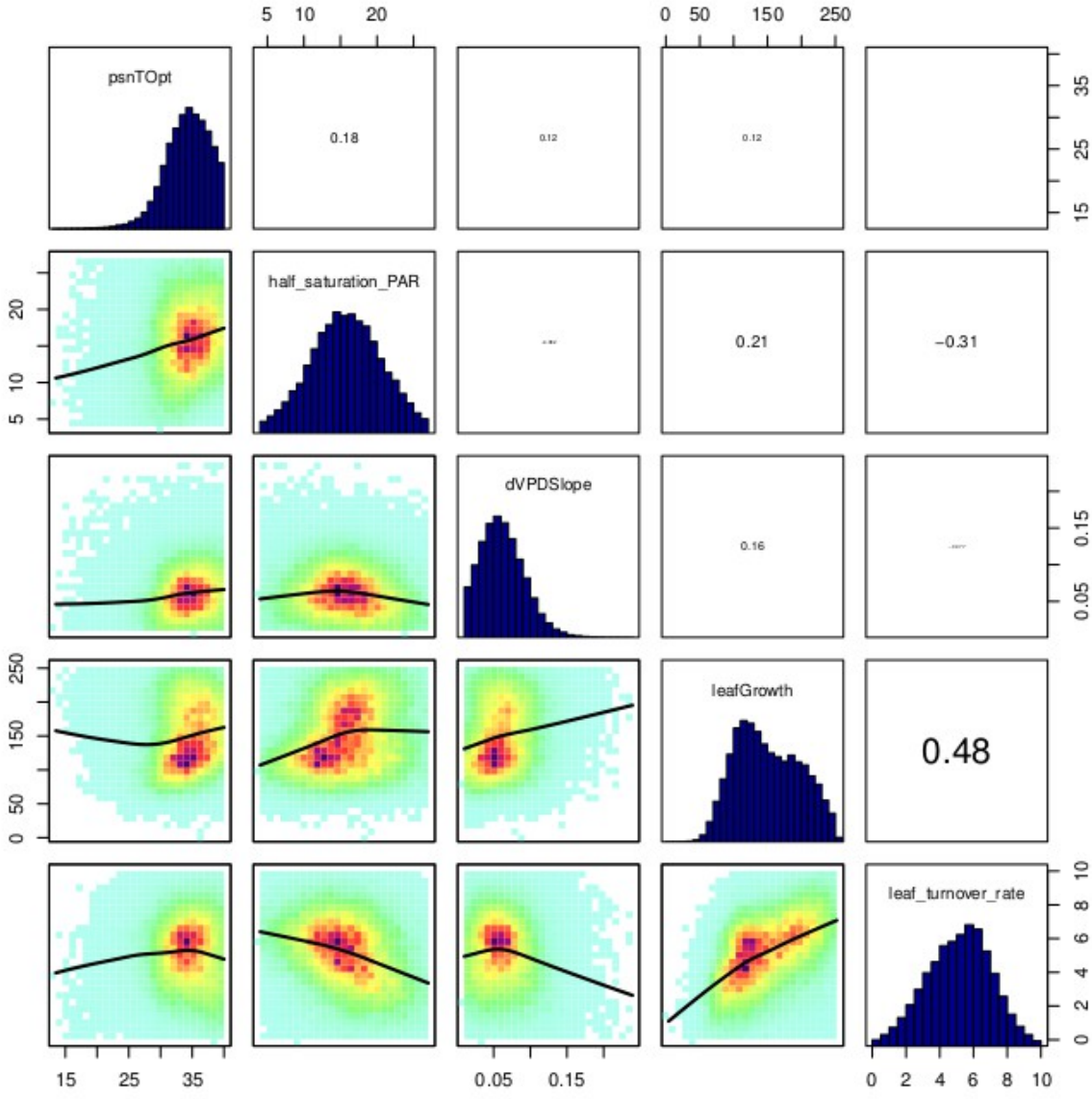
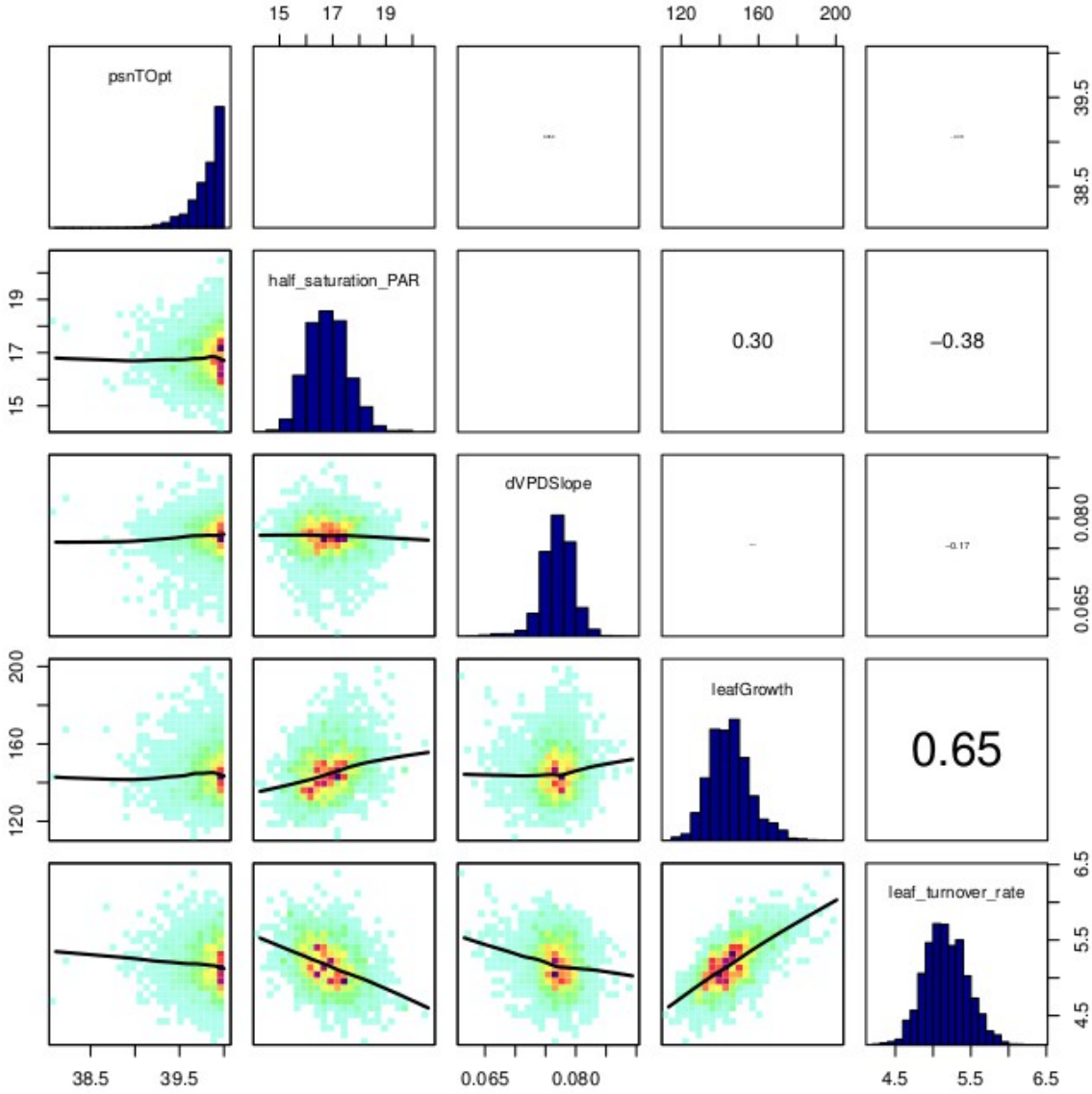


Figure S4. Correlation density plot after emulator MCMC (SIPNET).



885 **Figure S5.** Correlation density plot after bruteforce MCMC (SIPNET).

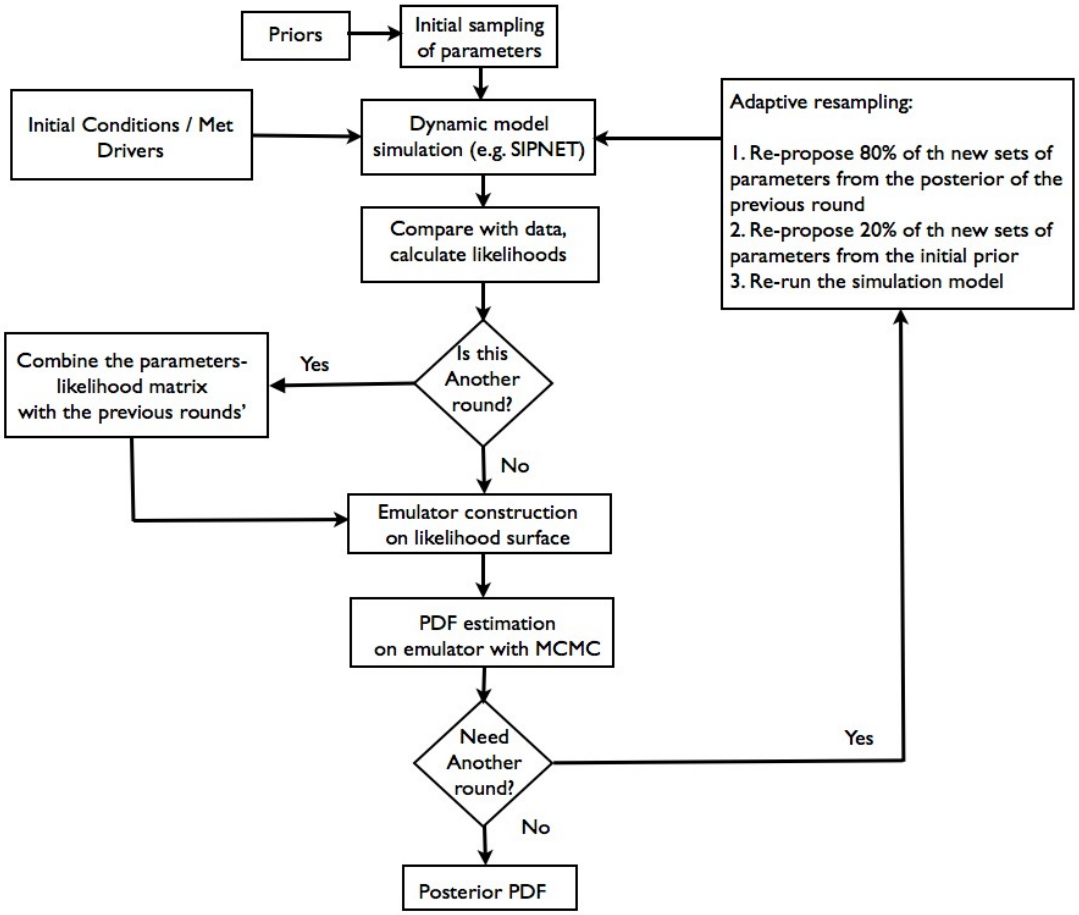


Figure S6. Schematic diagram of emulator workflow.

References

Aber, J.D., and Federer, C.A.: A generalized, lumped-parameter model of photosynthesis, evapo-transpiration and net primary production in temperate and boreal forest ecosystems. Oecologia, 92, 463–474, 1992.

Bradford, J.B., Weishampel, P., Smith, M.-L., Kolka, R., Birdsey, R.A., Ollinger, S.V., and Ryan, M.G.: Carbon pools and fluxes in small temperate forest landscapes: Variability and implications for sampling design, Forest Ecology and Management, 259(7), 1245-1254, doi: 10.1016/j.foreco.2009.04.009, 2010.

Braswell, B.H., Sacks, W.J., Linder, E., and Schimel, D.S.: Estimating diurnal to annual ecosystem parameters by synthesis of a carbon flux model with eddy covariance net ecosystem exchange observations, Global Change Biology, 11, 1-21, doi:10.1111/j.1365-2486.2005.00897.x, 2005

Dietze, M.C., and Moorcroft, P.R.: Tree mortality in the eastern and central United States: patterns and drivers, Global Change Biology, 17(11), 3312-3326, doi: 10.1111/j.1365-2486.2011.02477.x, 2011.

Dietze, M.C., Bayesian allometry models, github.com/PecanProject, Allmoetry module, in prep.

Jenkins, J.C., Chojnacky, D.C., Heath, L.S., and Birdsey, R.A.: Comprehensive database of diameter-based biomass regressions for North American tree species, General Technical Report, USDA, 2004.

Lee, M.S., Hollinger, D.Y., Keenan, T.F., Ouimette, A.P., Ollinger, S.V., and Richardson, A.D.: Model-based analysis of the impact of diffuse radiation on CO_2 exchange in a temperate deciduous forest. Agricultural and Forest Meteorology, 249, 377-389, doi:10.1016/j.agrformet.2017.11.016, 2018.

Medvigy, D., Wofsy, S.C., Munger, J.W., Hollinger, D.Y., and Moorcroft: Mechanistic scaling of ecosystem function and dynamics in space and time: Ecosystem Demography model version 2. Journal of Geophysical Research, 114, G01002, doi:.10.1029/2008JG000812, 2009.

Moorcroft, P.R., Hurtt, G.C., and Pacala, S.W.: A method for scaling vegetation dynamics: The Ecosystem Demography Model (ED), Ecological Monographs, 71(4), 557-586, doi: 10.1890/0012-9615(2001)071[0557:AMFSVD]2.0.CO;2, 2001.

Richardson, A.D., Jenkins, J.P., Braswell, B.H., Hollinger, D.Y., Ollinger, S.V., and Smith, M.-L.: Use of digital webcam images to track spring green-up in a deciduous broadleaf forest, Oecologia, 152: 323, doi: 10.1007/s00442-006-0657-z, 2007.

Sacks, W.J., Schimel, D.S., and Monson, R.K.: Coupling between carbon cycling and climate in a high-elevation, subalpine forest: a model-data fusion analysis, Oecologia, 151, 54-68, doi: 0.1007/s00442-006-0565-2, 2007.