Biogeosciences Discuss., doi:10.5194/bg-2017-46, 2017
Manuscript under review for journal Biogeosciences
Discussion started: 16 February 2017


Biogeosciences Discussions
Open Access
EGU

# Leveraging 35 years of forest research in the southeastern U.S. to constrain carbon cycle predictions: regional data assimilation using ecosystem experiments

R. Quinn Thomas[1*], Evan Brooks[1], Annika Jersild[1], Eric Ward[2], Randolph Wynne[1], Timothy J. Albaugh[1], Heather Dinon Aldridge[3], Harold E. Burkhart[1], Jean-Christophe Domec[4,5], Thomas R. Fox[1], Carlos A. Gonzalez-Benecke[6], Timothy A. Martin[7], Asko Noormets[8], David A. Sampson[9], Robert O. Teskey[10]

[1]Department of Forest Resources and Environmental Conservation, Virginia Tech, USA,
[2]Climate Change Science Institute and Environmental Sciences Division, Oak Ridge National Laboratory, USA,
[3]State Climate Office of North Carolina, North Carolina State University, USA,
[4]Bordeaux Sciences Agro, UMR 1391 INRA-ISPA, 33175 Gradignan Cedex, France,
[5]Nicholas School of the Environment, Box 90328, Duke University, Durham, NC 27708, USA,
[6]Department of Forest Engineering, Resources and Management, Oregon State University, USA,
[7]School of Forest Resources and Conservation, University of Florida, USA,
[8]Department of Forestry and Environmental Resources, North Carolina State University, USA,
[9]Decision Center for a Desert City, Arizona State University, USA,
[10]Warnell School of Forestry and Natural Resources, University of Georgia, USA,

*Corresponding author: R. Quinn Thomas (rqthomas@vt.edu)

Biogeosciences Discuss., doi:10.5194/bg-2017-46, 2017
Manuscript under review for journal Biogeosciences
Discussion started: 16 February 2017



## Abstract

Predicting how forest carbon cycling will change in response to climate change and management depends on the collective knowledge from measurements across environmental gradients, ecosystem manipulations of global change factors, and mathematical models. Formally integrating these sources of knowledge through data assimilation, or model-data fusion, allows the use of past observations to constrain model parameters and estimate prediction uncertainty. However, the influence of different experimental treatments on those predictions depends on the exact methods and techniques used for data assimilation. Here, we introduce a hierarchical Bayesian DA approach (Data Assimilation of Pine Plantation Ecosystem Research, DAPPER) that uses observations of carbon stocks, carbon fluxes, water fluxes, and vegetation dynamics from loblolly pine plantation ecosystems across the Southeastern U.S. to constrain parameters in a modified version of the 3-PG forest growth model. The observations included major experiments that manipulated atmospheric carbon dioxide ($CO_2$) concentration, water, and nutrients, along with non-experimental studies that spanned environmental gradients across an $8.6 \times 10^5$ $km^2$ region. We optimized regionally representative posterior distributions for the most sensitive model parameters, which dependably predicted data from plots withheld from the data assimilation. The posterior distributions of parameters associated with ecosystem responses to $CO_2$, precipitation, and nutrient addition, along with the corresponding regional changes in production associated with nutrient fertilization and drought, depended on how the experimental data were assimilated. In particular, assimilating nutrient addition experiments reduced the predicted sensitivity to nutrient fertilization while assimilated water manipulation experiments increased the sensitivity to drought. Further, it was necessary to assimilate data from the $CO_2$ experimental enrichment site before other studies to constrain the parameters associated with the

Biogeosciences Discuss., doi:10.5194/bg-2017-46, 2017
Manuscript under review for journal Biogeosciences
Discussion started: 16 February 2017




48 influence of $CO_2$ on canopy photosynthesis. The ambient $CO_2$ plots were numerous and had a
49 large contribution to the cost function compared to the low number of elevated $CO_2$ plots (289
50 ambient vs. 5 elevated plots). Overall, we demonstrated how three decades of research in
51 southeastern U.S. planted pine forests can be used to develop data assimilation techniques that
52 use multiple locations, multiple data streams, and multiple ecosystem experiment types to
53 optimize parameters. This approach allows for future predictions to be consistent with a rich
54 history of ecosystem research across a region.
55
56

Biogeosciences Discuss., doi:10.5194/bg-2017-46, 2017
Manuscript under review for journal Biogeosciences
Discussion started: 16 February 2017




## 1 Introduction

Forest ecosystems absorb and store a large fraction of anthropogenic carbon dioxide ($CO_2$) emissions (Le Quere et al., 2015; Pan et al., 2011) and supply wood products to a growing human population (Shvidenko et al., 2005). Therefore, predicting future carbon sequestration and timber supply is critical for adapting forest management practices to future environmental conditions and for using forests to assist with reduction of atmospheric $CO_2$ concentrations. The key sources of information for developing these predictions are results from global change ecosystem manipulation experiments, observations of forest dynamics across environmental gradients, and process-based ecosystem models. The challenge is integrating these three tools into a common framework for creating probabilistic predictions, or forecasts (based on (Luo et al., 2011a)), that provide information on both the expected future state of the forest and the probability distribution of those future states.

Data assimilation (DA), or data-model fusion, is an increasingly used framework for integrating ecosystem observations into ecosystem models (Luo et al., 2011a; Niu et al., 2014; Williams et al., 2005). DA integrates observations with ecosystem models through statistical, often Bayesian, methods that generate probability distributions for ecosystem model parameters and initial states. DA allows for the explicit accounting of observational uncertainty (Keenan et al., 2011), the incorporation of multiple types of observations with different time scales of collection (Richardson et al., 2010), and the representation of prior knowledge through informed parameter prior distributions or specific relationships among parameters (Bloom and Williams, 2015). Using DA to parameterize ecosystem models with observations from multiple locations that leverage environmental gradients and from ecosystem manipulation experiments will allow for

Biogeosciences Discuss., doi:10.5194/bg-2017-46, 2017
Manuscript under review for journal Biogeosciences
Discussion started: 16 February 2017




forecasts to be consistent with the rich history of global change research in forest ecosystems.

Ecosystem manipulation experiments provide a controlled environment in which data collected can be used to describe how forests acclimate and operate under altered environmental conditions (Medlyn et al., 2015). These data may be used to constrain model parameters that are associated with specific physiological functions associated with, for example, carbon allocation and turnover as related to the controlled manipulation. Furthermore, the assimilation of experiments may increase parameter identifiability (reducing equifinality (Luo et al., 2009)), where two parameters have compensating controls on the same processes, by isolating the response to a manipulated driver. For example, carbon assimilation and primary productivity can be modeled as a light and temperature controlled process that is adjusted by nutrients, water, and atmospheric $CO_2$ concentration. In this case, the productivity may mathematically be equal between a parameterization that has high potential conversion of light to photosynthesis (high quantum yield) but low relative nutrient availability and a parameterization with low quantum yield but high relative nutrient availability. Therefore, the challenge is that the same rate of production can emerge from different contributions of environmental controls.

For future predictions with changing environmental conditions, the relative contribution of each environmental control should be separated in order to correctly parameterize the sensitivity to changes in the environment. Key examples of existing and past ecosystem experiments that have the potential to isolate specific parameters in DA include $CO_2$ enrichment, water manipulation, nutrient addition, and elevated soil temperature experiments. Many of these experiments are common, particularly when including nutrient addition experiments in managed forests. Other



types of experiments are less common, but the few sites with the experiments, such as whole-ecosystem $CO_2$ enrichment, include intensive measurements of numerous carbon pools and fluxes required for model optimization.

Developing optimized parameters that apply to a region requires assimilating observations that span environmental gradients to support the application of model predictions to a range of climatic conditions, nutrient availabilities, and soil water dynamics. Therefore, the DA of multiple research sites across a region is an important extension of prior DA research focused on DA at a single site with multiple types of observations (Keenan et al., 2012; Richardson et al., 2010; Weng and Luo, 2011). Incorporating multiple locations that include global change experiments in DA is associated with numerous challenges. First, prior research has demonstrated that high frequency observations (i.e., daily, or more frequent, net ecosystem exchange observations) can overwhelm the contribution of low frequency observations (i.e., annual tree diameter measurements) to the cost-function used for optimization (Richardson et al., 2010), resulting in a parameter set that predominately represents the high-frequency dynamics. DA of ecosystem experiments and regional observations can present similar issues because key contrasts isolated in an ecosystem experiment with relatively few plots may be overwhelmed by the contribution of more numerous regional observations from non-manipulated plots. For example, whole ecosystem $CO_2$ enrichment experiments are uncommon but are the only observations representing ecosystem dynamics in an environment with over 550 ppm atmospheric $CO_2$ (McCarthy et al., 2010). Therefore, DA techniques may be required that assign additional weight to unique, but rare, experiments in the DA approach. As an example, a multi-

Biogeosciences Discuss., doi:10.5194/bg-2017-46, 2017
Manuscript under review for journal Biogeosciences
Discussion started: 16 February 2017




126 stage Bayesian approach could be used where the observations from the unique experiment are
127 assimilated first and the posteriors from that assimilation are used as priors for the assimilation
128 of the remaining observations. Second, DA requires using highly simplified ecosystem models
129 because many DA methods use millions of iterations to explore parameter distributions and these
130 iterations have to be applied to both control and manipulated treatments. However, in tension
131 with the need for simple models in DA, more complex models that simulate carbon, water, and
132 nutrient dynamics are also needed to fully leverage the diversity of ecosystem manipulation
133 experiments. Monthly time-scale models of ecosystem processes may be well suited to overcome
134 these challenges for application to predicting changes in biomass over decades in response to
135 global change. First, the contribution of monthly flux and annual biomass measurements to the
136 optimized cost function is more similar in monthly than daily models (12:1 vs. 365:1). Second,
137 they are computationally more efficient than daily models commonly used in DA, allowing data
138 spanning hundreds of plots and multiple decades to be assimilated. Finally, DA is able to
139 calibrate parameters associated with carbon, nitrogen, and water cycles so that they are
140 appropriate for an aggregated monthly time step, helping prevent potential issues associated
141 when applying daily parameterizations to coarser temporal time-steps.
142
143 Southeastern U.S. planted pine forests are ideal ecosystems for exploring the application of DA
144 to carbon cycle and forest production predictions. These ecosystems are dominated by loblolly
145 pine (*Pinus taeda* L.), thus allowing for a single parameter set to be applicable to a large region
146 containing many soil types and climatic gradients. Loblolly pine represents more than one half of
147 the standing pine volume in the southern United States (11.7 million ha) and is by far the single
148 most commercially important forest tree species for the region, with more than 1 billion

Biogeosciences Discuss., doi:10.5194/bg-2017-46, 2017
Manuscript under review for journal Biogeosciences
Discussion started: 16 February 2017




149 seedlings planted annually (Fox et al., 2007; McKeand et al., 2003). There is also a rich history
150 of experimental research focused on global change factors including region-wide nutrient
151 addition (Albaugh et al., 2016; Carlson et al., 2014; Raymond et al., 2016), water exclusion
152 (Bartkowiak et al., 2015; Tang et al., 2004; Ward et al., 2015; Will et al., 2015), and water
153 addition experiments (Albaugh et al., 2004; Allen et al., 2005; Samuelson et al., 2008). The
154 region also includes a long-term ecosystem $CO_2$ enrichment study (McCarthy et al., 2010).
155 Furthermore, many of these experiments are multi-factor with water exclusion-by-nutrients (Will
156 et al., 2015), water addition-by-nutrients (Albaugh et al., 2004; Allen et al., 2005; Samuelson et
157 al., 2008), and $CO_2$-by-nutrients treatments (McCarthy et al., 2010; Oren et al., 2001). Beyond
158 experimental treatments, Southeastern U.S. loblolly pine ecosystems include at least two eddy-
159 covariance sites with high frequency measurements of carbon and water fluxes along with
160 biometric observations over many years (Noormets et al., 2010; Novick et al., 2015), and sites
161 with multi-year sap flow data (Ewers et al., 2001; Gonzalez-Benecke and Martin, 2010; Phillips
162 and Oren, 2001). Finally, there are available studies that include plots that span the regional
163 environmental gradients and extend back to the 1980s (Burkhart et al., 1985). Overall, the high
164 availability of observations of biomass stocks, leaf area index (LAI), carbon fluxes, water fluxes,
165 and vegetation dynamics that span the past 35 years in loblolly pine ecosystems, including plots
166 with experimental manipulation and plots across environmental gradients, is well suited to
167 potentially constrain model parameters and predictions of how carbon cycling responds to
168 environmental change.
169
170 Our objective was to develop a DA approach that integrated diverse data from multiple locations,
171 including ecosystem experiments, for predicting how forest productivity may respond to global

Biogeosciences Discuss., doi:10.5194/bg-2017-46, 2017
Manuscript under review for journal Biogeosciences
Discussion started: 16 February 2017




change. We applied DA techniques to optimize a monthly-time step, simple forest productivity model using southeastern U.S.-wide experimental (nutrient addition, $CO_2$ enrichment, and water manipulations) and non-experimental data from 35 years of loblolly pine plantation research in the region. Our DA approach, DAPPER (Data Assimilation of Pine Plantation Ecosystem Research), is unique in its focus on simultaneously assimilating observations from multiple locations, experimental types, and data streams into a simple ecosystem model that includes carbon, water, and (implicitly) nutrients using a hierarchal Bayesian technique to develop parameter distributions. We used the DAPPER system to evaluate the sensitivity of biomass predictions and parameter distributions to the inclusion of ecosystem experiments in DA and to predict the regional sensitivity of forest production to nutrient fertilization and drought.

## 2 Methods

### 2.1 Ecosystem Model

We used a modified version of the Physiological Principles Predicting Growth (3-PG) Model to simulate vegetation dynamics in loblolly pine stands (Bryars et al., 2013; Gonzalez-Benecke et al., 2016; Landsberg and Waring, 1997). 3-PG is a stand-level vegetation model that runs at the monthly time-step and includes vegetation carbon dynamics and a simple soil water bucket model (Figure 1). While a complete description of the 3-PG model and our modifications can be found in the Supplemental Material, the key concept for interpreting the results is that gross primary productivity (GPP) was simulated using a light-use efficiency approach where the absorbed photosynthetically active radiation (APAR) was converted to carbon based on a quantum yield. Quantum yield was simulated using a parameterized maximum quantum yield

Biogeosciences Discuss., doi:10.5194/bg-2017-46, 2017
Manuscript under review for journal Biogeosciences
Discussion started: 16 February 2017




195 ($\alpha$) that was modified by environmental conditions including air temperature, atmospheric $CO_2$,
196 available soil water and soil fertility. The available soil water and soil fertility modifiers were
197 values between 0 and 1, while the atmospheric $CO_2$ modifier had a value of 1 at 350 ppm and
198 values greater than 1 at higher $CO_2$ concentrations.
199
200 Elevated $CO_2$ modified tree physiology by increasing quantum yield, based on an increasing but
201 saturating relationship with atmospheric $CO_2$. We also added a function where the allocation to
202 foliage relative to stem biomass decreased as atmospheric $CO_2$ increased. Available soil water
203 and quantum yield were positively related through a logistic relationship between relative
204 available soil water and the quantum yield modifier, where relative available soil water was the
205 ratio of simulated available soil water to a plot-level maximum available soil water. Soil fertility
206 and quantum yield were proportionally related, where quantum yield was scaled by an estimate
207 of relative stand-level fertility where a value of 1 was the maximum fertility. The fertility
208 modifier (FR) was constant throughout a simulation of a plot and was either based on site
209 characteristics or directly optimized as a stand-level parameter. Here we used site-index, a
210 measure of the height of a stand at a specified age (25 years), and the 35-year mean annual
211 temperature as site characteristics to predict FR. For a given climate, site index captures
212 differences in soil fertility, where a lower site index corresponded to a site with lower fertility.
213 However, regional variation in site index also included the influence of climate on growth rates
214 that were already accounted for in the other environmental modifiers in the 3-PG model. To
215 account for the climatic influence on site index, a long-term climate variable (35-year mean
216 annual temperature) was included in the empirical relationship that predicted FR as an
217 increasing, but saturating, function of site index. For plots with nutrient fertilization, FR was a

Biogeosciences Discuss., doi:10.5194/bg-2017-46, 2017
Manuscript under review for journal Biogeosciences
Discussion started: 16 February 2017




directly optimized parameter. For our application of the 3-PG model using DA, we removed the previously simulated dependence of total root allocation on FR(Bryars et al., 2013; Gonzalez-Benecke et al., 2016). Therefore, plots with lower FR could be interpreted to have lower quantum yield. Other environmental conditions influenced GPP, including temperature, frosts days, and vapor pressure deficit with a description of these modifiers found in the Supplemental Material.

Each month, net primary production (a parameterized and constant proportion of GPP) was allocated to foliage, stem (stemwood, stembark, and branches), coarse roots, and fine roots. Differing from previous applications of 3-PG to loblolly pine ecosystems, we modified the model to simulate fine roots and coarse roots separately. 3-PG also simulated simple population dynamics by including stem density as a state variable. Stem density and stem biomass pools were reduced by both density-dependent and density-independent mortality (a new modification), with the former based on the concept of self-thinning. Finally, we added a simple model of hardwood understory vegetation to enable the use of estimates of gross primary productivity and evapotranspiration from eddy-covariance tower studies with significant understories. Details of the model can be found in the Supplemental Material.

The water cycle was a simple bucket model with transpiration predicted using a Penman-Monteith approach (Bryars et al., 2013; Gonzalez-Benecke et al., 2016; Landsberg and Waring, 1997). The canopy conductance used in the Penman-Monteith subroutine was modified by environmental conditions. The modifiers include the same available soil water and vapor pressure deficit modifier as used in the GPP calculation. Maximum canopy conductance

Biogeosciences Discuss., doi:10.5194/bg-2017-46, 2017
Manuscript under review for journal Biogeosciences
Discussion started: 16 February 2017




241 occurred when simulated LAI exceeded a parameterized value of leaf area index (LAI).
242 Evaporation was equal to the precipitation intercepted by the canopy. Runoff occurred when the
243 available soil water exceeded a plot-specific maximum available soil water. As in prior
244 applications of 3-PG, available soil water was not allowed take a value below a minimum
245 available soil water, resulting in an implicit irrigation in very dry conditions.
246
247 The 3-PG model used in this study simulated the monthly change in eleven state variables per
248 plot: four stocks for loblolly pines, five stocks for understory hardwoods, loblolly pine stem
249 density (stems $ha^{-1}$), and available soil water. The key fluxes that were used for DA included
250 monthly GPP, monthly evapotranspiration (ET), annual root turnover, and annual foliage
251 turnover. In total, 46 parameters were required by 3-PG with 31 of the parameters optimized
252 using DA (Table 1, Supplemental Table 1, SI Table 2). The model required mean daily
253 maximum temperature, mean daily minimum temperature, daily PAR, total frost days, total rain
254 at the monthly time scale, monthly atmospheric $CO_2$, and latitude. Each plot also required
255 maximum available soil water, site index, mean annual temperature, and the initial condition of
256 the eleven state variables as model inputs (Figure 2).
257
258 **2.2 Observations**
259 We used thirteen different data streams from 294 plots at 187 unique locations spread across the
260 region to constrain model parameters (Table 2; Figure 3). The data streams covered the period
261 between 1981 to 2015. All data streams were not available in all plots (Table 2; Table 3). The
262 most common set of data streams were annual or less frequent observations of stand stem
263 biomass (defined as the sum of stemwood, stembark and branches), winter foliage biomass, and
264 living tree counts. The stem and foliage biomass were optimized using regional allometric

Biogeosciences Discuss., doi:10.5194/bg-2017-46, 2017
Manuscript under review for journal Biogeosciences
Discussion started: 16 February 2017




models based on measurements of tree diameter, height, and plot level-stem size distributions (Gonzalez-Benecke et al., 2014). The most comprehensive set of data streams was from Duke Forest where annual measurements and allometric-based estimates were made of stem biomass (loblolly pine and hardwood), coarse root biomass (loblolly pine and hardwood), fine root biomass (combined loblolly pine and hardwood), stem count (loblolly pine only), leaf turnover (combined loblolly pine and hardwood), and fine root production (combined loblolly pine and hardwood). The Duke Forest dataset (DK3 combined with the Duke FACE $CO_2$ fertilization study) also included monthly observations of LAI, gross ecosystem production (GEP; modeled gross primary productivity from net ecosystem exchange measured at an eddy-covariance tower), and ET. The set of data streams associated with a particular site and experimental design is shown in Table 3. The measurement uncertainty associated with each data stream is listed in Table 2. Since the model used a monthly time-step, and plots with only biomass and stem density observations were more common than plots with monthly flux estimates, the data used in the optimization cost function were not dominated by high frequency data streams (GEP and ET).

### 2.3 Data assimilation method

We used a hierarchal Bayesian framework to approximate the posterior probability distributions of model parameters in Table 1, the model process uncertainty parameters, and the latent model states and fluxes. The latent model states represented the 'true' stock or flux before measurement uncertainty was included in the observation. Our hierarchal approach was designed to partition uncertainty that is attributable to uncertainty in parameters, model process, and measurements (Hobbs and Hooten, 2015). Previous forest ecosystem DA efforts have either focused on parameter uncertainty, by using measurement uncertainty as the variance term in a Gaussian cost

Biogeosciences Discuss., doi:10.5194/bg-2017-46, 2017
Manuscript under review for journal Biogeosciences
Discussion started: 16 February 2017




function, or on total uncertainty by directly estimating the Gaussian variance term. The latter combines measurement uncertainty and process uncertainty into the same parameter and is unable to be used for developing prediction intervals, as prediction intervals only include parameter and process errors (Dietze et al., 2013; Hobbs and Hooten, 2015). Here, our focus was on estimating the probability distribution of forest biomass before uncertainty is added through measurement.

First, we estimated the probability of a latent state or flux ($z_{i,m,p}$) for each data point (i) from each data stream (m) in a plot (p) using the 3-PG model with the plot FR . This included the optimized parameters ($\theta_F$), fixed parameters ($\theta_C$), soil characteristic inputs (S), climate inputs (C), site index (SI), fertility ($FR_p$), and initial conditions (I) required by the 3-PG to simulate each plot, $f(\theta_F, \theta_c, C, S, I, FR_p)$. The latent state ($z_{i,m,p}$) was assumed to be normally distributed with the mean from the 3-PG simulation and an optimized, data stream-specific, process variance $\sigma^2_{m,(process)}$

$$p(\text{process}|\text{process parameters}) =$$

$$P\left(z_{i,m,p}\middle|f\left(\theta_F,\theta_C,C,S,I,FR_p\right),\sigma^2_{m\ (process)}\right)$$

$$\sim \text{Normal}\left(z_{i,m,p}\middle|f\left(\theta_F,\theta_C,C,S,I,FR_p\right),\sigma^2_{m\ (process)}\right) \qquad \text{Equation 1}$$

The unobserved true state related to the observed state through a data observation model. In the sampling model, the measured state ($y_{i,m,p}$) was a random sample from a normal distribution with a mean of the true state and a data point-specific standard deviation ($\sigma^2_{i,m,p}$).

$$p(\text{data}|\text{process,data parameters}) =$$

Biogeosciences Discuss., doi:10.5194/bg-2017-46, 2017
Manuscript under review for journal Biogeosciences
Discussion started: 16 February 2017




311 $P\left(y_{i,m,p}\middle|z_{i,m,p},\sigma^2_{i,m,p}\right)\sim\text{Normal}\left(y_{i,m,p}\middle|z_{i,m,p},\sigma^2_{i,m,p}\right)$ Equation 2

312

313 This standard deviation ($\sigma^2_{i,m,p}$) represented measurement uncertainty and was similar to the

314 denominator in least-squares approach that is commonly used in DA (Bloom and Williams,

315 2015; Keenan et al., 2011).

316

317 Each parameter ($\theta_F$) that was optimized using the Bayesian method had a prior probability that is

318 specified in Table 1. The prior distribution for the standard deviation $\sigma^2_{m,(process)}$ parameters

319 were uniformly distributed:

320

321 $p(\text{process parameters}|\text{priors})\times p(\text{priors})= P(\sigma^2_m)\times P(\theta_F)$ Equation 3

322

323 where

324

325 $P(\sigma^2_m)\sim\text{unif}(0.001,100)$ Equation 4

326

327 and

328

329 $P(\theta_F)\sim$See Supplemental Table 1 Equation 5

330

331 Finally, following the description of the plot specific $FR_p$ described above, the probability for

332 fertilized treatments was based on a comparison to the control treatment FR.

333

Biogeosciences Discuss., doi:10.5194/bg-2017-46, 2017
Manuscript under review for journal Biogeosciences
Discussion started: 16 February 2017




334 $$P(FR_p|\theta_F,E)=\begin{cases}1 \text{ if non-fertilized}\\ 1 \text{ if fertilized and } FR_p \geq FR \text{ of control plot}\\ 0 \text{ if fertilized and } FR_p < FR \text{ of control plot}\end{cases}$$ Equation 6

335

336 Our complete Bayesian model for estimating the posterior distributions for the parameters ($\theta_F$),

337 process uncertainty ($\sigma^2_{m,(process)}$), and unobserved true states ($z_{i,m,p}$) was:

338

339 $$P\left(\theta_F,\sigma^2_m,z_{i,m,p}\middle|y_{i,m,p},\sigma^2_{i,m,p},\theta_C,S,C,SI,I\right) \propto$$

340 $$P(z_{i,m,p}|f(\theta_F,\theta_C,FR_p,E),\sigma^2_m)P(y_{i,m,p}|z_{i,m,p},\sigma^2_{i,m,p})P(FR_p|\theta_F,E)P(\theta_F)P(\sigma^2_m)$$ Equation 7

341

342 We numerically estimated the posterior distributions using the Monte-Carlo Markov Chain –

343 Metropolis Hasting (MCMC-MH) algorithm (Zobitz et al., 2011). This approach has been widely

344 used to approximate parameter distributions in ecosystem DA research (Fox et al., 2009;

345 Trudinger et al., 2007; Williams et al., 2005; Zobitz et al., 2011). We adapted the size of the

346 jump for each parameter (i.e., how far a proposed new value can potentially be from the current

347 value) to ensure the acceptance rate of the parameter set is between 22% and 43% (Ziehn et al.,

348 2012). All MCMC-MH chains were run for 30 million iterations with the first 15 million

349 iterations discarded as the burn-in. Three chains were run and compared for convergence and we

350 sampled every 1000$^{th}$ parameter in the final 15 million iterations of the MCMC-MH chain. This

351 thinned chain was used in the analysis described below. The 3-PG model and MCMC-MH

352 algorithm were programed in FORTRAN 90 and used OpenMP to parallelize the simulation of

353 each plot within an iteration of the MCMC-MH algorithm.

354

Biogeosciences Discuss., doi:10.5194/bg-2017-46, 2017
Manuscript under review for journal Biogeosciences
Discussion started: 16 February 2017




355 **2.4 Model simulations**

356 Each plot simulated required initial conditions for each model state, climate inputs, soil
357 characteristic inputs, and site index. We used the first observation at the plot as the initial
358 conditions for the loblolly pine vegetation states (foliage biomass, stem biomass, coarse root
359 biomass, fine root biomass, and stem number). When observations of coarse biomass and fine
360 root biomass were not available, these stocks were initialized as a mean region-wide proportion
361 of the observed stem biomass. However, the value of initial root biomass in plots without
362 observations was not important because the plots without root observations did not contribute to
363 the root cost function and root biomass does not influence any other functions in the model. In
364 the two plots with flux observations (US-Dk3 and US-NC2), hardwood understory was also
365 initialized using the first set of observations. Initial fine root and coarse biomass was distributed
366 between loblolly pine and hardwoods based on their relative contribution of total initial foliage
367 biomass. The initialized available soil water was assumed to be equal to the maximum available
368 soil water because most plots were initialized in winter months when plant demand for water is
369 minimal. The maximum available soil water in each plot was extracted from the SSURGO soils
370 dataset (Staff, 2016). We assumed that the minimum available soil water was zero. The value we
371 used corresponded to the maximum available soil water for the top 1.5 m of the soil. Because we
372 focused on a region-wide optimization, we used region-wide 4-km estimates of observed
373 monthly meteorology as inputs and to collect the 35-year mean annual temperature for each plot
374 (Abatzoglou, 2013). Site index was based on height measurements at age 25 in each plot or
375 calculated by combining observations of height at younger ages with an empirical model

Biogeosciences Discuss., doi:10.5194/bg-2017-46, 2017
Manuscript under review for journal Biogeosciences
Discussion started: 16 February 2017




376 (Dieguez-Aranda et al., 2006).

377

378 We simulated the experiments by altering the environmental modifiers or by modifying the
379 environmental inputs. Nutrient addition experiments were simulated by directly estimating FR,
380 rather than calculating from Equation 2, and by requiring the optimized FR in the fertilized plot
381 to be equal to or greater than the FR in the control plots. Throughfall exclusion experiments were
382 simulated by decreasing rain inputs by 30% in the treatment plots. This assumed that the
383 fractional reduction in precipitation and throughfall were equal. The SETRES Irrigation
384 experiments were simulated by adding 650 mm to precipitation between April and October. $CO_2$
385 enrichment experiments were simulated by setting the atmospheric $CO_2$ input equal to the
386 treatment mean from the elevated $CO_2$ rings (570 ppm). While not an experiment, one plot (US-
387 NC2) included a thinning treatment during the period of observation. We simulated the thinning
388 by specifying a decrease in the stem count that matched the proportion removed at the site, with
389 the biomass of each tree equivalent to the average of trees in the plot.

390

391 **2.4 Model experiments and analysis**

392 Our analysis focused on comparing parameter distributions and predictions among simulations
393 that used different experimental treatments to estimate the posterior distributions (Table 4). To
394 examine the influence of the Duke FACE $CO_2$ fertilization, we compared a one stage vs. a two-
395 stage data assimilation process. The one stage process assimilated all observations in all plots
396 and experiments simultaneously. In this approach, the elevated $CO_2$ plots only represented 5 of
397 the 294 plots across the region and thus a relatively minor contribution to the likelihood (cost-
398 function) calculation. The two-stage process used the observations from Duke FACE, US-Dk3

Biogeosciences Discuss., doi:10.5194/bg-2017-46, 2017
Manuscript under review for journal Biogeosciences
Discussion started: 16 February 2017




399 flux site, the other flux site in North Carolina (US-NC2) to estimate parameter posteriors using
400 the priors in Table 1 and SI Table 1. These sites were grouped together because they were the
401 most data rich, had the high frequency data streams (monthly GEP, ET, and LAI), and were
402 relatively close in geography. FR was directly estimated for all plots in the first stage, with the
403 FR of a fertilized plot required to be equal to or higher than its control plot. The FR of the $CO_2$
404 experiment was equal to the corresponding control plot estimated FR. The FR of the control plot
405 was required to be greater than 0 and, if associated with a nutrient fertilization plot, less than the
406 FR of the fertilized plot.
407
408 For the second DA stage, the posterior distributions from the first stage were used as priors for
409 the assimilation of the region-wide observations from the PINEMAP, FPC RW 18, FMRC
410 Thinning, SETRES, and Waycross studies (Table 4). We compared the $CO_2$ quantum yield
411 enhancement parameter (Calpha700) between the one and two stage approaches to evaluate how
412 the estimation of $CO_2$ fertilization of plant growth depended on how the Duke FACE data are
413 used in data assimilation. We also estimated the distribution of the percentage increase in net
414 primary productivity (NPP) associated with the elevated $CO_2$ treatment using the one and two
415 stage data assimilation approaches. The distribution of the percentage increase in NPP was
416 calculated by randomly selecting 1000 parameter sets, with replacement, from the 1-stage
417 converged MCMC chains. This calculation was repeated using the 2-stage approach.
418
419 Based on the results from comparing the one and two stage approaches (see results below), we
420 proceeded using the two-stage approach to examine the influence of the water manipulation and
421 nutrient fertilization experiments on posterior distributions and predictions. To evaluate the

Biogeosciences Discuss., doi:10.5194/bg-2017-46, 2017
Manuscript under review for journal Biogeosciences
Discussion started: 16 February 2017




influence of water manipulation experiments, we repeated the second stage of the data assimilation without the plots where water was added or subtracted. To evaluate the influence of the nutrient manipulation experiments, we first repeated the first stage of data assimilation without the nutrient addition plots in the Duke FACE experiment and used those posteriors as priors to the second stage. This ensured that the priors to the second stage of data assimilation did not include information from nutrient addition experiments. The second stage then excluded the other nutrient manipulation experiments in the region.

To examine how the exclusion of the water manipulation experiments influenced parameter inference and predictions, we first examined how the parameter distributions changed from initial priors through the two assimilation stages. With respect to the water manipulation experiments, we focused on the shape of the relationship between available soil water and the quantum yield and stomatal conductance modifier (governed by parameters SW1 and SW2) with and without assimilating the water manipulation experiments. To illustrate the capacity to estimate the probability distribution of predictions using the posterior uncertainty in parameters, we analyzed a focal site in Georgia, near the center of the loblolly pine range (circle in Figure 2). At the focal site, we predicted the sensitivity of stem biomass at age 25 (hereby referred to as $STEM_{25}$) to a 30% increase and a 30% decrease in annual precipitation with and without assimilating the water experiments. A 30% percent decrease in precipitation mirrors the magnitude of reduction in the experimental throughfall reduction studies used in DA (Table 3 and Figure 3). Our prediction distributions were calculated by integrating across the parameter uncertainty by repeating simulations using 1000 random draws from the converged chain of the posteriors. Finally, we predicted the regional response to a reduction in precipitation from



445 historical using the median posterior parameter values from the data assimilation with and
446 without the water experiments included. Our regional corresponded to the native range of
447 loblolly and used the HUC12 (USGS 12-digit Hydrological Unit Code) watershed as the scale of
448 simulation. For each HUC12 in the region we used the mean site index, 30-year mean annual
449 temperature, available soil water aggregated to the HUC12 level, and monthly meteorology as
450 inputs (Figure 2). We simulated forest development from 1989 to 2014 using actual precipitation
451 and again with a 30% reduction in precipitation. We focused our analysis on the percent change
452 in $STEM_{25}$ between the two simulations.
453
454 To examine how the exclusion of the nutrient addition experiment influenced parameter
455 inference and prediction, we focused on the difference in maximum quantum yield parameter ($\alpha$)
456 and the relationship between site index and soil fertility modifier (FR) with and without
457 assimilating the nutrient experiments. Additionally, we simulated how stem biomass at age 25
458 ($STEM_{25}$) responded to a complete removal of nutrient limitation (FR = 1) for the focal site in
459 Georgia. As in the precipitation sensitivity described above, we represented the percentage
460 change in $STEM_{25}$ between simulations with estimated FR and FR =1 as a distribution by
461 integrating across parameter uncertainty. We predicted the regional response to nutrient
462 fertilization by setting the FR at all HUC12 units (see previous paragraph) equal to 1 using the
463 median posterior parameter values from data assimilation where nutrient addition experiments
464 were either included or not. We focused on the regional pattern in the percentage change in stem
465 biomass with the predicted FR (current level fertility) and FR = 1 (nutrient limitation removed).
466
467 Finally, we assessed overall model performance of the 2-stage approach for data assimilation

Biogeosciences Discuss., doi:10.5194/bg-2017-46, 2017
Manuscript under review for journal Biogeosciences
Discussion started: 16 February 2017




468 with all experimental types included in DA, excluding the nutrient addition experiments, and
469 excluding the nutrient addition experiments using an out-of-sample approach. The approach held
470 40 random FMRC thinning study plots (Table 3) out from the assimilation, predicted the 40 plots
471 using the median parameter values, and compared the predicted stem biomass to the observed
472 stem biomass. These were plots without any manipulations of nutrients or water, were located
473 throughout the region, and had measurement ages up to 30 years old. For each plot, we only used
474 the most recent observed values to increase the time length between initialization and validation.
475 We repeated the validation for four unique sets of 40 FMRC thinning study plots.
476

477 **3 Results**

478 Our multi-site, multi-experiment, multi-data stream DA approach was able to constrain most
479 parameters in the 3-PG model (31 of 46 parameters were optimized; Table 6; Supplemental
480 Table 3; Supplemental Figure 1-3). The 31 optimized parameters were the most sensitive
481 parameters in the 3-PG model, defined by the change in total biomass at age 25 for the focal site
482 in Georgia to a 10% change in the parameter (Table 1; Supplemental Table 1). One exception
483 was the light extinction coefficient (k), which showed high sensitivity but was assumed to be
484 fixed because it strongly co-varied with the quantum yield parameter ($\alpha$). Parameters associated
485 with biomass allocation had priors with large variance but DA was able to provide posteriors
486 with relatively low variance (pFS2, pFS20, pR, and pCRS; Supplemental Figure 1; Supplemental
487 Table 3). The DA process also produced posterior distributions that had less variability than the
488 prior distribution for the important parameters associated with light-use efficiency ($\alpha$, y, FR1,
489 and FR2; Table 5). DA did not change the parameter distributions, i.e., the posterior and prior
490 distributions were similar, for the parameters that governed the temperature sensitivity of

Biogeosciences Discuss., doi:10.5194/bg-2017-46, 2017
Manuscript under review for journal Biogeosciences
Discussion started: 16 February 2017




quantum yield, the VPD sensitivity of quantum yield, and the maximum canopy conductance (Supplemental Figure 1-2; Supplemental Table 3). These parameters had strong priors supported by previous research on loblolly pine physiology. Finally, the DA approach was able to estimate the distributions of the process uncertainty parameters (Supplemental Figure 3; Supplement Table 4).

The addition of the second stage of assimilation that used region-wide observations and posteriors from the DK+NC2 assimilation modified the distributions of the parameters that related to allocation and mortality but did not provide additional constraint on the physiological parameters (Table 5). In particular, the parameters associated with the self-thinning curve and allocation of coarse roots had non-overlapping 95% credible intervals between the DK+NC2 and RW assimilation. The larger estimate for Wsx1000 and lower value for thinPower in the DK+NC2 indicated self-thinning was lower at the sites in the DK+NC2 assimilation than the average of the other sites in the region. The lower value for the pCRS parameter indicated that less NPP was allocated to coarse roots in the DK+NC2 assimilation than the RW assimilation.

The two-stage assimilation was critical for constraining the $CO_2$ quantum yield enhancement parameter (Calpha700). Both the mean of the posterior distribution and the range of the 95% credible interval were smaller for fCalpha700 when all observations were assimilated simultaneously (1-stage approach) than the distribution estimated using the 2-stage approach (Duke and NC2 assimilated before the region-wide assimilation) (Figure 5a; Table 5). Despite the same data used in both approaches, the differences in fCalpha700 led to a predicted lower enhancement of NPP associated with elevated $CO_2$ in the experiment. The 1-stage assimilation



514 approach had a median increase in NPP between the control and elevated $CO_2$ treatments of 15%
515 compared to a 27% in the two-stage approach (Figure 5b).
516
517 The RW assimilation constrained the soil fertility parameters that were necessary to enable
518 regional simulations. Our regional model using the 2-stage approach performed well compared
519 to stem biomass data not used in the assimilation. The mean bias in stem biomass of the four out-
520 of-sample validation sets was -6.7 % and the RMSE was 21.2 Mg $ha^{-1}$ (Figure 4).
521
522 Excluding the nutrient addition experiments from the DA increased the simulated level of
523 nutrient limitation but did not change the predictive capacity of the independent non-manipulated
524 validation set. DA without nutrient fertilization experiments had a greater and more uncertain
525 value for the maximum quantum yield parameter ($\alpha$; Figure 6a; Table 5). This parameter was
526 shared across all plots and modified by the environmental conditions at each plot. To compensate
527 for the higher $\alpha$ parameter when nutrient fertilization experiments were excluded from DA, the
528 two soil fertility parameters (FR1 and FR2) combined to predict a 10% lower FR values for a
529 given site index and mean annual temperature (Figure 6b). Subsequently, the prediction for the
530 percentage change in $STEM_{25}$ associated with maximum fertilization (i.e., setting FR = 1) at the
531 focal site in Georgia was 7% higher and had greater uncertainty when nutrient fertilization
532 experiments were excluded from the DA (Figure 6c). The RMSE and mean bias of the non-
533 manipulated validation set was 20.4 Mg $ha^{-1}$ and -4.8 %, respectively (SI Figure 1a)
534
535 Excluding the water manipulation experiments from the DA reduced the sensitivity to available
536 soil water but, similar to the inclusion of the nutrient addition experiments, did not change the

Biogeosciences Discuss., doi:10.5194/bg-2017-46, 2017
Manuscript under review for journal Biogeosciences
Discussion started: 16 February 2017




537 predictive capacity of the independent non-manipulated validation set. The combined differences
538 in the SW1 and SW2 parameters between the DA with and without the water manipulation
539 experiments decreased the sensitivity of quantum yield and canopy conductance to a reduction in
540 available soil water (Figure 7a). For example, at an available soil water to maximum available
541 soil water ratio of 0.50, the quantum yield and canopy conductance modifier decreased from 0.95
542 without water experiments to 0.8 with water experiments (Figure 7a). At the focal site in
543 Georgia, the sensitivity of $STEM_{25}$ to a reduction in annual precipitation (Figure 7b) was larger
544 when the water experiments were included in the DA (-8.5% median change in $STEM_{25}$ for a
545 30% reduction in precipitation) than when the experiments were excluded (-4.1% median change
546 in $STEM_{25}$ for a 30% reduction in precipitation). Similarly, the predictions of $STEM_{25}$ change
547 associated with a 30% increase in precipitation (median: 3.8%) were higher when water
548 experiments were included than when not included (median: 1.1%). The magnitude of
549 uncertainty in the predictions did not differ substantially between forecasts with and without
550 water experiments (Figure 7b). The RMSE and mean bias of the non-manipulated validation set
551 was 19.3 Mg $ha^{-1}$ and -5.8 %, respectively (SI Figure 1b)
552
553 Regionally (i.e., the native range of loblolly pines), using the two-stage approach (RW), the most
554 productive areas were the coastal plains and the interior of Mississippi and Alabama (Figure 8).
555 These patterns were largely driven by patterns in the soil fertility factor (FR; Figure 9), reflecting
556 the sensitivity of the 3-PG model to the FR parameters (Table 1). The area weighted mean
557 $STEM_{25}$ response to fertilization (represented by setting FR = 1) across the region was 28% with
558 the highest response occurring in the far west of the region, the Piedmont of Georgia, the interior
559 of the gulf coast, and the northern reach of the region (Figure 10a). These were all areas with the

Biogeosciences Discuss., doi:10.5194/bg-2017-46, 2017
Manuscript under review for journal Biogeosciences
Discussion started: 16 February 2017




560 lowest soil fertility parameter. The least responsive region to nutrient addition was in Florida
561 (Figure 10a). Excluding the nutrient addition experiments from the DA increased the sensitivity
562 to nutrient addition (Figure 10b), as shown for the focal Georgia site (Figure 6b), but did not
563 change the spatial patterns of the response.
564
565 The sensitivity of forest production to a 30% reduction in precipitation varied across the region.
566 The most sensitive areas, the Piedmont of Georgia and the western edge of the region, predicted
567 up to a 13.1% decline in $STEM_{25}$ (Figure 11a). These were warm areas with relatively low
568 precipitation before the 30% reduction (Figure 2c). The least sensitive area was the interior of the
569 gulf coast (<1% decline; Figure 11a), the area with the highest precipitation in the region (Figure
570 2c). The regional mean reduction in $STEM_{25}$ associated with a 30% decrease in precipitation was
571 5.7% (Figure 11a). Excluding the water manipulation experiments from DA reduced the regional
572 mean sensitivity to 1.7% (Figure 11b).
573
574 **4 Discussion**
575
576 Using DA to parameterize models applied to forecasting ecosystem change requires detangling
577 the vegetation responses to temperature, precipitation, nutrients, and elevated $CO_2$. To address
578 this challenge, we introduced a regional-scale hierarchical Bayesian approach (DAPPER) that
579 assimilated data across environmental gradients and ecosystem manipulation experiments into a
580 modified version of the 3-PG model to estimate parameters and generate uncertainty estimates
581 on predictions of carbon and water cycling across the whole native range of loblolly pine.
582 Furthermore, we organized observations of carbon stocks, carbon fluxes, water fluxes, vegetation



583 structure, and vegetation dynamics that spanned 35 years of forest research (Figure 3; Table 3) in
584 a region with large and dynamic carbon fluxes (Lu et al., 2015). By combining the DAPPER
585 system with the regional set of observations, we were able to estimate parameters in a model
586 with high predictive capacity (Figure 4) and with quantified uncertainty on parameters (Table 5).
587 We also found that the predictions of forest productivity response to rising $CO_2$, altered
588 precipitation, and altered nutrient availability were highly sensitive to the types of experiments
589 used in DA as well as the methodological approach applied.
590
591 We found that including nutrient and water manipulation experiments aided in distinguishing the
592 mechanisms driving patterns in biomass across the region. Including these experiments in the
593 data-assimilation did not improve the predictive capacity of the independent validation set of
594 non-manipulation plots. However, including nutrient and water manipulation did change the
595 underlying mechanisms explaining the patterns in stem biomass. Without the nutrient and water
596 manipulation experiments, the same biomass predictions were attributable to a higher level of
597 nutrient limitation and a lower level of water limitation. This resulted in differing sensitivities to
598 changes in nutrient or water availability. Overall, this finding highlights a key challenge when
599 parameterizing ecosystem models that will be used for global change predictions, that different
600 combinations of environmental drivers can produce similar predictions of current observations.
601 Ecosystem manipulation experiments are an important tool for addressing this challenge.
602
603 Parameter and process identifiability, or equifinality, presents a challenge when parameterizing
604 ecosystem models using DA (Luo et al., 2009). One important source of equifinality is the
605 tradeoff between parameters governing the potential productivity of the vegetation and the



606 downregulation of productivity due to nutrient limitation. When using observational data at a
607 single site, a single parameter is often optimized to set a photosynthetic rate per absorbed light,
608 i.e., a quantum yield. This single parameter combines the potential photosynthesis set by climate
609 and the influence of nutrient limitation on photosynthesis into a single parameter. However,
610 separating these two processes into two or more parameters is challenging because a high
611 potential quantum yield parameter ($\alpha$) and high nutrient limitation (FR) can mathematically yield
612 the same photosynthetic rate as low potential quantum yield and low nutrient limitation. The
613 former implies a larger potential response to nutrient addition than the latter. We found that
614 including nutrient addition experiments in DA helped overcome this challenge. In the case of the
615 3-PG model used in this study, the maximum quantum parameter ($\alpha$) and soil fertility parameters
616 (FR1 and FR2) were more constrained and inferred lower levels of nutrient limitation across the
617 region when nutrient fertilization experiments were included in the DA. This finding likely
618 extends to other models that include the concept of potential productivity and productivity
619 downregulated by nutrient limitation. For example, the applications of the Data Assimilation
620 Link Ecosystem Carbon (DALEC) model (Williams et al., 2005) to DA often assumed nine of
621 the ten parameters associated with photosynthesis were fixed, thus using a single parameter to
622 represent both the quantum yield (defined as nitrogen use efficiency in DALEC) and the
623 magnitude of nitrogen limitation of a site (Fox et al., 2009). The use of a single parameter, rather
624 than using nutrient addition experiments to separate into multiple parameters, is appropriate
625 when assuming nutrient availability is static. Applications of DA to predictions of ecosystems
626 with changing nutrient availability, either through management, elevated $CO_2$, or nitrogen
627 addition, would benefit from using nutrient addition studies to quantify the magnitude of nutrient
628 limitation. Studies of known nutrient gradients could be used in lieu of nutrient addition studies,



but effort must be made to account for confounding abiotic factors, such as available soil water or climatic conditions, that may co-vary with nutrient availability.

Another challenge in DA is deciding how to weigh different types of data used in model fitting (Gao et al., 2011; Wutzler and Carvalhais, 2014). Here we demonstrate that DA efforts should also consider how to weigh different types of ecosystem experiments. In our analysis, we included three types of experiments: nutrient addition, water manipulation, and $CO_2$ fertilization. The nutrient addition and water manipulation experiments were represented by multiple sites across the region while the $CO_2$ fertilization only occurred at a single location (Figure 3). We found that the parameter that represents the increase in maximum quantum yield under elevated $CO_2$ was substantially lower when all observations, sites, and experiments were assimilated simultaneously than when the $CO_2$ fertilization experiment was given greater weight. The greater weight was applied by first assimilating the $CO_2$ fertilization experiment and using the posteriors as priors for assimilating the remaining observations. Providing additional weight on the single site with unique environmental conditions (i.e., atmospheric $CO_2$ at 570 ppm) using a two-stage data-assimilation, we were able to more accurately represent the observed differences in NPP between the ambient and elevated $CO_2$ treatments at the Duke site (McCarthy et al., 2010). Given than only a few of the parameters were significantly different between the Duke site and the other studies across the region, it may be possible to optimize one parameter for the Duke site and another parameter for the other studies in a 1-stage approach that combines all the plots into a single assimilation. However, the 2-stage approach was required to identify which parameters were different between the Duke site and the other studies. Overall, we suggest that DA efforts using multiple studies and multiple experiment types identify whether particular experiments at



limited number of sites have the potential to uniquely constrain specific parameters. In this case, additional weight may be needed to avoid having the signal of the unique experiment overwhelmed by the large amount of data from the other sites and experiments.

Our analysis highlights that nutrient limitation of productivity was widespread across the region. The largest potential gains in productivity from nutrient addition were predicted in central Georgia, an area with warm annual temperatures but poor soils, as expressed in the low site index. The baseline fertility used in our regional analysis was derived from an empirical model of site index that was developed using field plots with minimal management (Sabatia and Burkhart, 2014). Subsequently our estimate of baseline fertility is likely on the low end of forest stands currently in production. Further, we recognize that the site index model had uncertainty that could be formally incorporated into the hierarchal Bayesian approach in future applications.

The soil fertility modifier has commonly been used to calibrate the 3-PG for applications to a single site, with recent work focused on developing an approach to predicting the soil fertility modifier from environmental conditions (Gonzalez-Benecke et al., 2016; Subedi et al., 2015). We have extended prior efforts to develop a simple predictive model of FR in two ways. First, we simultaneously calibrated the parameters in the empirical FR model alongside the other parameters in the 3-PG model. Prior studies have assumed fixed values for the 3-PG model parameters, fitted FR for plots with observations, and developed a relationship between FR and site index. Our Bayesian approach to simultaneously calibrating the 3-PG parameters and the FR model allowed for the estimation of uncertainty and covariation among parameters in the 3-PG and FR models. Second, we included a climate term (mean annual temperature) in the



675 relationship between site index and FR. This resulted in a lower FR for a given site index in
676 warmer locations. By including the climate term, FR can be interpreted as relative to the climate
677 at a given location and the potential productivity of a plot can be optimized by setting FR equal
678 to 1. When a climate term is not used in the empirical FR model, FR is relative to the greatest
679 site index in the region, which does not occur in the northern extent of the region even in
680 fertilized plots due to climatic constraints.
681
682 Our simulations show that loblolly pine productivity was not strongly sensitive to changes in
683 precipitation at present day temperatures and atmospheric $CO_2$. We simulated a 30% reduction in
684 annual precipitation and found a maximum of a 13.1% reduction in productivity. A 30%
685 reduction in precipitation is plausible but is more extreme than most Multivariate Adaptive
686 Constructed Analogs (MACA) downscaled climate model projections for the Representative
687 Concentration Pathway (RCP) 8.5 scenario from the CMIP5 Project (comparing the 1971-2000
688 period to the 2070-2099) (Abatzoglou and Brown, 2012; Taylor et al., 2012). Central Georgia
689 was the most responsive to precipitation reduction, paralleling the spatial patterns in the response
690 to nutrient addition, suggesting that the region is able to support high productivity but is sensitive
691 to nutrient and precipitation levels. The simulated sensitivity was likely due to poor soils (low
692 site index) and low baseline precipitation relative to the warm climate. Our predictions of low
693 sensitivity to precipitation reduction or addition were derived from assimilating observations
694 from throughfall exclusion and irrigation experiments across the region. Prior publications from
695 the studies used in DA also reported low sensitivities to water manipulations, indicating that our
696 predictions are likely not biased (Albaugh et al., 2004; Samuelson et al., 2014; Ward et al., 2015;
697 Wightman et al., 2016). For example, the throughfall exclusion experiment at the focal site in

Biogeosciences Discuss., doi:10.5194/bg-2017-46, 2017
Manuscript under review for journal Biogeosciences
Discussion started: 16 February 2017




698 Georgia, reported a 13% reduction in stem production during a dry year but a 0% reduced during
699 a wet year, resulting in a 7% reduction of productivity over a 2-year period in response to a 30%
700 reduction in throughfall (Samuelson et al., 2014). Our predicted 8.5% reduction to a 30%
701 reduction in precipitation compares well to the observed change, noting that our sensitivity
702 integrated over a 25-year rotation and included a mix of relatively wet and dry years.
703
704 The 3-PG model included a highly simplified representation of interactions between the water
705 and carbon cycles that resulted in parameterizations that, while consistent with observations, may
706 contain assumptions that require additional investigation. For example, transpiration is modeled
707 as a potential canopy transpiration that occurred if leaf area was not limiting transpiration. The
708 LAI at which leaf area was no longer limiting was a parameter that was optimized (LAIgcx in SI
709 Table 3), resulting in a value of 2.3. Interestingly, this optimized value is consistent with the
710 scant literature on this topic. In their analysis of multi-year measurements of transpiration in
711 loblolly pine, Phillips and Oren (2001) observed that transpiration per unit leaf area was
712 relatively insensitive to increases in leaf area above LAI of approximately 2.5. Iritz and Lindroth
713 (1996) reviewed transpiration data from a range of crop species and found only small increases
714 in transpiration above LAI of 3-4. These authors suggest that the threshold-type responses
715 observed were related to the range of LAI at which self-shading increases most rapidly, therefore
716 limiting increases in transpiration. The resulting model behavior of "flat" transpiration above 2.3
717 LAI, with gradually decreasing photosynthesis above that value, results in increasing water use
718 efficiency at higher LAI values. The parameterization of the relationships between transpiration
719 and photosynthesis in 3-PG would likely benefit fromadditional data beyond the two eddy-
720 covariance studies with ET observations used here. For example, canopy conductance estimates,

Biogeosciences Discuss., doi:10.5194/bg-2017-46, 2017
Manuscript under review for journal Biogeosciences
Discussion started: 16 February 2017




721 and their associated uncertainty have been derived from assimilating observations from sap-flow
722 measurements into a model that scales from the sensor measurements to canopy transpiration
723 using LAI observations (Bell et al., 2015). This sap-flow to canopy conductance scaling
724 approach (the State Space Canopy Conductance (StaCC) model (Bell et al., 2015)) produces a
725 probability distribution of monthly canopy conductance that could be integrated into the
726 DAPPER system by treating the posterior estimates of StaCC as the distribution of the data in
727 equation 2. Second, the optimized parameters that described the relationship between relative
728 available soil water and the modifier of photosynthesis and transpiration predicted a modifier
729 value greater than zero when the relative available soil water was zero. This resulted in positive
730 values from photosynthesis and transpiration when the average available soil water during the
731 month was zero. In practice, the monthly available soil water was rarely zero during simulations,
732 which presents a challenge constraining the shape of the available soil water modifier. The priors
733 for the two available soil moisture modifiers (SW1 and SW2) had ranges that permitted the
734 modifier to be zero. Therefore, additional data is likely needed during very dry conditions to
735 develop a more physically based parameterization. Alternatively, the parameterization of a non-
736 zero soil moisture modifier at zero available soil water may be due to trees having access to
737 water at soil depths deeper than the top 1.5 m of soil represented by the bucket in 3-PG. Overall,
738 it is important to view the parameterization presented here as a phenomenological relationship
739 that is consistent with observations from throughfall exclusion and irrigation experiments as well
740 as observations across regional gradients in precipitation.
741
742 Beyond the specifics of the 3-PG modeling efforts, the DA of regional observations into a
743 monthly, computationally tractable ecosystem model can potentially inform Earth system



744 modeling efforts. While the details of physiology differ between 3-PG and global land-surface
745 models, the concepts governing NPP allocation are similar. Therefore, DA using the 3-PG model
746 can be used to parameterize the allocation patterns of similar plant types in a global model. One
747 land-surface model, the Community Land Model (CLM), includes parameters that govern the
748 ratio of stem to leaf allocation, ratio of coarse root to stem allocation, and the ratio of leaf to fine
749 root allocation, parameters that are also optimized in DAPPER. As an example, the ratio of fine
750 root to leaf allocation in CLM 4.0 and 4.5 for temperate pine plant function type is set to 1,
751 resulting in equal annual allocation of carbon to foliage and fine roots (Oleson et al., 2013). In
752 contrast, we found that the median ratio of fine root to foliage allocation was substantially lower
753 at 0.13 (Table 6). Therefore, simulations in the CLM with the lower value of root allocation
754 would have higher allocation to aboveground tissues if the loblolly pine parameters from our
755 analysis were used. This would increase carbon accumulation in woody tissues and could alter
756 predictions of nutrient limitation because stems have higher C:N ratios. Other parameters,
757 including the stem to coarse root ratio, are closer to the values used in the CLM.
758
759 **5 Conclusions**
760 DA is increasingly used for ecological forecasting due to its ability represent prior knowledge,
761 integrate observations into the parameterization, and estimate multiple components of
762 uncertainty, including observation, parameter, and process representation uncertainty (Dietze et
763 al., 2013; Luo et al., 2011b; Niu et al., 2014). Our application of DA to loblolly pine plantations
764 of the southeastern U.S demonstrated that these ecosystems are well suited as a test-bed for the
765 development of DA techniques, particularly techniques for assimilating ecosystem experiments.
766 Further, we found that assimilating ecosystem manipulative experiments into a simple ecosystem



767 model changed predictions quantifying how forest productivity responds to environmental
768 change, highlighting the importance of networks of ecosystem manipulation experiments for
769 helping to parameterize and evaluate ecosystems models (Medlyn et al., 2015).
770

771 **6 Data availability**

772 Observations used in the DA can be found in the following: Duke FACE study can be found in
773 McCarthy et al. (McCarthy et al., 2010), the PINEMAP studies are available through the TerraC
774 database (http://terrac.ifas.ufl.edu), the DK3 eddy-flux tower data are available through the
775 Ameriflux database (http://ameriflux-data.lbl.gov) , the Waycross data can be found in Bryars et
776 al. (2003), the NC2 data are available upon request with Asko Noormets, the FMRC and FPC are
777 available through membership with the cooperatives. The parameter chains and 3-PG are
778 available upon request from R. Quinn Thomas.
779

780 **Acknowledgments**

781 Funding support came from USDA-NIFA Project 2015-67003-23485 and the Pine Integrated
782 Network: Education, Mitigation, and Adaptation project (PINEMAP), a Coordinated
783 Agricultural Project funded by the USDA National Institute of Food and Agriculture, Award
784 #2011-68002-30185. Additional funding support came from USDA-NIFA McIntire-Stennis
785 Program. The Virginia Space Grant Consortium Graduate STEM Research Fellowship Program
786 provided partial support for A. Jersild. Computational support was provided by Virginia Tech
787 Advanced Research Computing. This research was also supported by grants from the French
788 Research Agency (MACACC ANR-13-AGRO-0005 and MARIS ANR-14-CE03-0007). We
789 thank Luke Smallman and Mat Williams for helpful discussions about data assimilation. We



790 thank the corporate and government agency members of the FPC and FMRC research
791 cooperatives for supporting the extensive long-term experimental and observational plots in
792 those datasets.

793

794 **References**

795

796 Abatzoglou, J. T.: Development of gridded surface meteorological data for ecological
797 applications and modelling, International Journal of Climatology, 33(1), 121–131,
798 doi:10.1002/joc.3413, 2013.

799 Abatzoglou, J. T. and Brown, T. J.: A comparison of statistical downscaling methods suited for
800 wildfire applications, International Journal of Climatology, 32(5), 772–780,
801 doi:10.1002/joc.2312, 2012.

802 Albaugh, T. J., Albaugh, J. M., Fox, T. R., Allen, H. L., Rubilar, R. A., Trichet, P., Loustau, D.
803 and Linder, S.: Tamm Review: Light use efficiency and carbon storage in nutrient and water
804 experiments on major forest plantation species, Forest Ecol Manag, 376, 333–342,
805 doi:10.1016/j.foreco.2016.05.031, 2016.

806 Albaugh, T. J., Allen, H. L. and Kress, L. W.: Root and stem partitioning of *Pinus taeda*, Trees,
807 20(2), 176–185, doi:10.1007/s00468-005-0024-4, 2005.

808 Albaugh, T. J., Lee Allen, H., Dougherty, P. M. and Johnsen, K. H.: Long term growth responses
809 of loblolly pine to optimal nutrient and water resource availability, Forest Ecol Manag, 192(1),
810 3–19, doi:10.1016/j.foreco.2004.01.002, 2004.

811 Albaugh, T., Fox, T., Allen, H. and Rubilar, R.: Juvenile southern pine response to fertilization is
812 influenced by soil drainage and texture, Forests, 6(8), 2799–2819, doi:10.3390/f6082799, 2015.

813 Allen, C. B., Will, R. E. and Jacobson, M. A.: Production efficiency and radiation use efficiency
814 of four tree species receiving irrigation and fertilization, Forest Science, 51(6), 556–569, 2005.

815 Bartkowiak, S. M., Samuelson, L. J., McGuire, M. A. and Teskey, R. O.: Fertilization increases
816 sensitivity of canopy stomatal conductance and transpiration to throughfall reduction in an 8-
817 year-old loblolly pine plantation, Forest Ecol Manag, 354, 87–96,
818 doi:10.1016/j.foreco.2015.06.033, 2015.

819 Bell, D. M., Ward, E. J., Oishi, A. C., Oren, R., Flikkema, P. G. and Clark, J. S.: A state-space
820 modeling approach to estimating canopy conductance and associated uncertainties from sap flux
821 density data, edited by D. Whitehead,, 35(7), 792–802, doi:10.1093/treephys/tpv041, 2015.

822 Bloom, A. A. and Williams, M.: Constraining ecosystem carbon dynamics in a data-limited
823 world: integrating ecological "common sense" in a model–data fusion framework,
824 Biogeosciences, 12(5), 1299–1315, doi:10.5194/bg-12-1299-2015, 2015.

825 Bryars, C., Maier, C., Zhao, D., Kane, M., Borders, B., Will, R. and Teskey, R.: Fixed
826 physiological parameters in the 3-PG model produced accurate estimates of loblolly pine growth
827 on sites in different geographic regions, Forest Ecol Manag, 289, 501–514,
828 doi:10.1016/j.foreco.2012.09.031, 2013.

829 Burkhart, H. E., Cloeren, D. C. and Amateis, R. L.: Yield relationships in unthinned loblolly pine
830 plantations on cutover, site-prepared lands, Southern Journal of Applied Forestry, 9(2), 84–91,
831 1985.

832 Carlson, C. A., Fox, T. R., Allen, H. L., Albaugh, T. J., Rubilar, R. A. and Stape, J. L.: Growth
833 responses of loblolly pine in the Southeast United States to midrotation applications of nitrogen,
834 phosphorus, potassium, and micronutrients, Forest Science, 60(1), 157–169,
835 doi:10.5849/forsci.12-158, 2014.

836 DeLucia, E. H., Drake, J. E., Thomas, R. B. and Gonzalez-Meler, M.: Forest carbon use
837 efficiency: is respiration a constant fraction of gross primary production? Global Change
838 Biology, 13(6), 1157–1167, doi:10.1111/j.1365-2486.2007.01365.x, 2007.

839 Dieguez-Aranda, U., Burkhart, H. E. and Amateis, R. L.: Dynamic site model for loblolly pine
840 (Pinus taeda L.) plantations in the United States, Forest Science, 52(3), 262–272, 2006.

841 Dietze, M. C., LeBauer, D. S. and Kooper, R.: On improving the communication between
842 models and data, Plant Cell Environ, 36(9), 1575–1585, doi:10.1111/pce.12043, 2013.

843 Ewers, B. E., Oren, R., Phillips, N., Stromgren, M. and Linder, S.: Mean canopy stomatal
844 conductance responses to water and nutrient availabilities in Picea abies and Pinus taeda, Tree
845 Physiology, 21(12-13), 841–850, 2001.

846 Fox, A., Williams, M., Richardson, A. D., Cameron, D., Gove, J. H., Quaife, T., Ricciuto, D.,
847 Reichstein, M., Tomelleri, E., Trudinger, C. M. and Van Wijk, M. T.: The REFLEX project:
848 Comparing different algorithms and implementations for the inversion of a terrestrial ecosystem
849 model against eddy covariance data, Agr Forest Meteorol, 149(10), 1597–1615,
850 doi:10.1016/j.agrformet.2009.05.002, 2009.

851 Fox, T. R., Jokela, E. J. and Allen, H. L.: The Development of Pine Plantation Silviculture in the
852 Southern United States, Journal of Forestry, 105(7), 337–347, 2007.

853 Gao, C., Wang, H., Weng, E., Lakshmivarahan, S., Zhang, Y. and Luo, Y.: Assimilation of
854 multiple data sets with the ensemble Kalman filter to improve forecasts of forest carbon
855 dynamics, Ecological Applications, 21(5), 1461–1473, doi:10.1890/09-1234.1, 2011.

856 Gonzalez-Benecke, C. A. and Martin, T. A.: Water availability and genetic effects on water
857 relations of loblolly pine (Pinus taeda) stands, Tree Physiology, 30(3), 376–392,
858 doi:10.1093/treephys/tpp118, 2010.

859 Gonzalez-Benecke, C. A., Gezan, S. A., Albaugh, T. J., Allen, H. L., Burkhart, H. E., Fox, T. R.,
860 Jokela, E. J., Maier, C. A., Martin, T. A., Rubilar, R. A. and Samuelson, L. J.: Local and general
861 above-stump biomass functions for loblolly pine and slash pine trees, Forest Ecol Manag, 334,

862 254–276, doi:10.1016/j.foreco.2014.09.002, 2014.

863 Gonzalez-Benecke, C. A., Teskey, R. O., Martin, T. A., Jokela, E. J., Fox, T. R., Kane, M. B.
864 and Noormets, A.: Regional validation and improved parameterization of the 3-PG model for
865 Pinus taeda stands, Forest Ecol Manag, 361, 237–256, doi:10.1016/j.foreco.2015.11.025, 2016.

866 Hobbs, N. T. and Hooten, M. B.: Bayesian Models: A Statistical Primer for Ecologists, Princeton
867 University Press, Princeton. 2015.

868 Iritz, Z. and Lindroth, A.: Energy partitioning in relation to leaf area development of short-
869 rotation willow coppice, Agr Forest Meteorol, 81(1-2), 119–130, doi:10.1016/0168-
870 1923(95)02306-2, 1996.

871 Keenan, T. F., Carbone, M. S., Reichstein, M. and Richardson, A. D.: The model–data fusion
872 pitfall: assuming certainty in an uncertain world, Oecologia, 167(3), 587–597,
873 doi:10.1007/s00442-011-2106-x, 2011.

874 Keenan, T. F., Davidson, E., Moffat, A. M., Munger, W. and Richardson, A. D.: Using model-
875 data fusion to interpret past trends, and quantify uncertainties in future projections, of terrestrial
876 ecosystem carbon cycling, Global Change Biology, 18(8), 2555–2569, doi:10.1111/j.1365-
877 2486.2012.02684.x, 2012.

878 Landsberg, J. and Waring, R.: A generalised model of forest productivity using simplified
879 concepts of radiation-use efficiency, carbon balance and partitioning, Forest Ecol Manag, 95(3),
880 209–228, doi:10.1016/S0378-1127(97)00026-1, 1997.

881 Le Quere, C., Moriarty, R., Andrew, R. M., Canadell, J. G., Sitch, S., Korsbakken, J. I.,
882 Friedlingstein, P., Peters, G. P., Andres, R. J., Boden, T. A., Houghton, R. A., House, J. I.,
883 Keeling, R. F., Tans, P., Arneth, A., Bakker, D. C. E., Barbero, L., Bopp, L., Chang, J.,
884 Chevallier, F., Chini, L. P., Ciais, P., Fader, M., Feely, R. A., Gkritzalis, T., Harris, I., Hauck, J.,
885 Ilyina, T., Jain, A. K., Kato, E., Kitidis, V., Klein Goldewijk, K., Koven, C., Landsch utzer, P.,
886 Lauvset, S. K., Lef evre, N., Lenton, A., Lima, I. D., Metzl, N., Millero, F., Munro, D. R.,
887 Murata, A., Nabel, J. E. M. S., Nakaoka, S., Nojiri, Y., O'Brien, K., Olsen, A., Ono, T., P 'erez,
888 F. F., Pfeil, B., Pierrot, D., Poulter, B., Rehder, G., R odenbeck, C., Saito, S., Schuster, U.,
889 Schwinger, J., S 'ef 'erian, R., Steinhoff, T., Stocker, B. D., Sutton, A. J., TAKAHASHI, T.,
890 Tilbrook, B., van der Laan-Luijkx, I. T., van der Werf, G. R., van Heuven, S., Vandemark, D.,
891 Viovy, N., Wiltshire, A., Zaehle, S. and Zeng, N.: Global Carbon Budget 2015, Earth Syst. Sci.
892 Data, 7(2), 349–396, doi:10.5194/essd-7-349-2015, 2015.

893 Lu, X., Lu, X., Kicklighter, D. W., Kicklighter, D., Melillo, J. M., Melillo, J. M., Reilly, J. M.,
894 Reilly, J. M., Xu, L. and Wu, L.: Land carbon sequestration within the conterminous United
895 States: Regional- and state-level analyses, J Geophys Res-Biogeo, 120(2), 379–398,
896 doi:10.1002/2014JG002818, 2015.

897 Luo, Y. Q., Luo, Y., Ogle, K., Ogle, K., Tucker, C., Tucker, C., Fei, S., Fei, S., Gao, C., Gao, C.,
898 LaDeau, S., LaDeau, S., Clark, J. S., Clark, J. S. and Schimel, D. S.: Ecological forecasting and
899 data assimilation in a data-rich era, Ecological Applications, 21(5), 1429–1442, 2011a.

900 Luo, Y., Ogle, K., Tucker, C., Fei, S., Gao, C., LaDeau, S., Clark, J. S. and Schimel, D. S.:
901 Ecological forecasting and data assimilation in a data-rich era, Ecological Applications, 21(5),
902 1429–1442, doi:10.1890/09-1275.1, 2011b.

903 Luo, Y., Weng, E., Wu, X., Gao, C., Zhou, X. and Zhang, L.: Parameter identifiability,
904 constraint, and equifinality in data assimilation with ecosystem models, Ecological Applications,
905 19(3), 571–574, doi:10.1890/08-0561.1, 2009.

906 McCarthy, H. R., Oren, R., Johnsen, K. H., Gallet-Budynek, A., Pritchard, S. G., Cook, C. W.,
907 LaDeau, S. L., Jackson, R. B. and Finzi, A. C.: Re-assessment of plant carbon dynamics at the
908 Duke free-air CO2 enrichment site: interactions of atmospheric [CO2] with nitrogen and water
909 availability over stand development, New Phytol, 185(2), 514–528, doi:10.1111/j.1469-
910 8137.2009.03078.x, 2010.

911 McKeand, S., Mullin, T., Byram, T. and White, T.: Deployment of genetically improved loblolly
912 and slash pines in the south, Journal of Forestry, 101(3), 32–37, 2003.

913 Medlyn, B. E., Zaehle, S., De Kauwe, M. G., Walker, A. P., Dietze, M. C., Hanson, P. J.,
914 Hickler, T., Jain, A. K., Luo, Y., Parton, W., Prentice, I. C., Thornton, P. E., Wang, S., Wang,
915 Y.-P., Weng, E., Iversen, C. M., McCarthy, H. R., Warren, J. M., Oren, R. and Norby, R. J.:
916 Using ecosystem experiments to improve vegetation models, Nature Climate Change, 5(6), 528–
917 534, doi:10.1038/nclimate2621, 2015.

918 Niu, S., Luo, Y., Dietze, M. C., Keenan, T. F., Shi, Z., Li, J. and III, F. S. C.: The role of data
919 assimilation in predictive ecology, Ecosphere, 5(5), art65–16, doi:10.1890/ES13-00273.1, 2014.

920 Noormets, A., Gavazzi, M. J., McNulty, S. G., Domec, J.-C., Sun, G., King, J. S. and Chen, J.:
921 Response of carbon fluxes to drought in a coastal plain loblolly pine forest, Glob Change Biol,
922 16(1), 272–287, doi:10.1111/j.1365-2486.2009.01928.x, 2010.

923 Novick, K. A., Oishi, A. C., Ward, E. J., Siqueira, M. B. S., Juang, J.-Y. and Stoy, P. C.: On the
924 difference in the net ecosystem exchange of CO2 between deciduous and evergreen forests in the
925 southeastern United States, Glob Change Biol, 21(2), 827–842, doi:10.1111/gcb.12723, 2015.

926 Oleson, K., Lawrence, D. M., Bonan, G. B., Drewniak, B., Huang, M., Koven, C. D., Levis, S.,
927 Li, F., Riley, W. J., Subin, Z. M., Swenson, S. C., Thornton, P. E., Bozbiyik, A., Fisher, R.,
928 Heald, C. L., Kluzek, E., Lamarque, J. F., Lawrence, P. J., Leung, L. R., Lipscomb, W., Muszala,
929 S., Ricciuto, D. M., Sacks, W., Sun, Y., Tang, J. and Yang, Z. L.: Technical Description of
930 version 4.5 of the Community Land Model (CLM), NCAR Technical Note NCAR, Boulder.
931 2013.

932 Oren, R., Ellsworth, D., Johnsen, K., Phillips, N., Ewers, B., Maier, C., Schafer, K., McCarthy,
933 H., Hendrey, G., McNulty, S. G. and Katul, G.: Soil fertility limits carbon sequestration by forest
934 ecosystems in a $CO_2$-enriched atmosphere, Nature, 411(6836), 469–472, doi:10.1038/35078064,
935 2001.

936 Pan, Y., Birdsey, R. A., Fang, J., Houghton, R., Kauppi, P. E., Kurz, W. A., Phillips, O. L.,
937 Shvidenko, A., Lewis, S. L., Canadell, J. G., Ciais, P., Jackson, R. B., Pacala, S. W., McGuire,

938 A. D., Piao, S. L., Rautiainen, A., Sitch, S. and Hayes, D.: A large and persistent carbon sink in
939 the world's forests, Science, 333(6045), 988–993, doi:10.1126/science.1201609, 2011.

940 Phillips, N. and Oren, R.: Intra- and inter-annual variation in transpiration of a pine forest,
941 Ecological Applications, 11(2), 385–396, 2001.

942 Raymond, J. E., Fox, T. R., Strahm, B. D. and Zerpa, J.: Differences in the recovery of four
943 different nitrogen containing fertilizers after two application seasons in pine plantations across
944 the southeastern United States, Forest Ecol Manag, 380(C), 161–171,
945 doi:10.1016/j.foreco.2016.08.044, 2016.

946 Richardson, A. D., Williams, M., Hollinger, D. Y., Moore, D. J. P., Dail, D. B., Davidson, E. A.,
947 Scott, N. A., Evans, R. S., Hughes, H., Lee, J. T., Rodrigues, C. and Savage, K.: Estimating
948 parameters of a forest ecosystem C model with measurements of stocks and fluxes as joint
949 constraints, Oecologia, 164, 25–40, doi:10.1007/s00442-010-1628-y, 2010.

950 Sabatia, C. O. and Burkhart, H. E.: Predicting site index of plantation loblolly pine from
951 biophysical variables, Forest Ecol Manag, 326, 142–156, doi:10.1016/j.foreco.2014.04.019,
952 2014.

953 Samuelson, L. J., Butnor, J., Maier, C., Stokes, T. A., Johnsen, K. and Kane, M.: Growth and
954 physiology of loblolly pine in response to long-term resource management: defining growth
955 potential in the southern United States, Can. J. For. Res., 38(4), 721–732, doi:10.1139/X07-191,
956 2008.

957 Samuelson, L. J., Pell, C. J., Stokes, T. A., Bartkowiak, S. M., Akers, M. K., Kane, M.,
958 Markewitz, D., McGuire, M. A. and Teskey, R. O.: Two-year throughfall and fertilization effects
959 on leaf physiology and growth of loblolly pine in the Georgia Piedmont, Forest Ecol Manag,
960 330, 29–37, doi:10.1016/j.foreco.2014.06.030, 2014.

961 Shvidenko, A., Barber, C. V. and Persson, R.: Forest and Woodland Systems, in Ecosystems and
962 Human Well-being Current State and Trends, Volume, edited by R. Hassan, R. Scholes, and N.
963 Ash, pp. 585–621, Island Press, Washington. 2005.

964 Staff, S. S.: Soil Survey Geographic (SSURGO) Database for Southeastern U.S,, 2016.

965 Subedi, S., Fox, T. and Wynne, R.: Determination of fertility rating (FR) in the 3-PG model for
966 loblolly pine plantations in the Southeastern United States based on site index, Forests, 6(9),
967 3002–3027, doi:10.3390/f6093002, 2015.

968 Tang, Z., Sayer, M. A. S., Chambers, J. L. and Barnett, J. P.: Interactive effects of fertilization
969 and throughfall exclusion on the physiological responses and whole-tree carbon uptake of mature
970 loblolly pine, Canadian Journal of Botany, 82(6), 850–861, doi:10.1139/b04-064, 2004.

971 Taylor, K. E., Stouffer, R. J. and Meehl, G. A.: An overview of CMIP5 and the experiment
972 design, Bull. Amer. Meteor. Soc., 93(4), 485–498, doi:10.1175/BAMS-D-11-00094.1, 2012.

973 Trudinger, C. M., Raupach, M. R., Rayner, P. J., Kattge, J., Liu, Q., Pak, B., Reichstein, M.,

974 Renzullo, L., Richardson, A. D., Roxburgh, S. H., Styles, J., Wang, Y.-P., Briggs, P., Barrett, D.
975 and Nikolova, S.: OptIC project: An intercomparison of optimization techniques for parameter
976 estimation in terrestrial biogeochemical models, J. Geophys. Res., 112(G2), G02027–17,
977 doi:10.1029/2006JG000367, 2007.

978 Ward, E. J., Domec, J.-C., Laviner, M. A., Fox, T. R., Sun, G., McNulty, S., King, J. and
979 Noormets, A.: Fertilization intensifies drought stress: Water use and stomatal conductance of
980 Pinus taeda in a midrotation fertilization and throughfall reduction experiment, Forest Ecol
981 Manag, 355, 72–82, doi:10.1016/j.foreco.2015.04.009, 2015.

982 Weng, E. and Luo, Y.: Relative information contributions of model vs. data to short- and long-
983 term forecasts of forest carbon dynamics, Ecological Applications, 21(5), 1490–1505,
984 doi:10.1890/09-1394.1, 2011.

985 Wightman, M., Martin, T., González-Benecke, C., Jokela, E., Cropper, W., Jr. and WARD, E.:
986 Loblolly Pine Productivity and Water Relations in Response to Throughfall Reduction and
987 Fertilizer Application on a Poorly Drained Site in Northern Florida, Forests, 7(10), 214–19,
988 doi:10.3390/f7100214, 2016.

989 Will, R., Fox, T., Akers, M., Domec, J.-C., González-Benecke, C., Jokela, E., Kane, M., Laviner,
990 M., Lokuta, G., Markewitz, D., McGuire, M., Meek, C., Noormets, A., Samuelson, L., Seiler, J.,
991 Strahm, B., Teskey, R., Vogel, J., WARD, E., West, J., Wilson, D. and Martin, T.: A range-wide
992 experiment to investigate nutrient and soil moisture interactions in loblolly pine plantations,
993 Forests, 6(6), 2014–2028, doi:10.3390/f6062014, 2015.

994 Williams, M., Schwarz, P., Law, B. E., Irvine, J. and Kurpius, M.: An improved analysis of
995 forest carbon dynamics using data assimilation, Global Change Biology, 11(1), 89–105,
996 doi:10.1111/j.1365-2486.2004.00891.x, 2005.

997 Wutzler, T. and Carvalhais, N.: Balancing multiple constraints in model-data integration:
998 Weights and the parameter block approach, J Geophys Res-Biogeo, 119(11), 2112–2129,
999 doi:10.1002/2014JG002650, 2014.

1000 Zobitz, J. M., Desai, A. R., Moore, D. J. P. and Chadwick, M. A.: A primer for data assimilation
1001 with ecological models using Markov Chain Monte Carlo (MCMC), Oecologia, 167(3), 599–
1002 611, doi:10.1007/s00442-011-2107-9, 2011.

1003

1004

Biogeosciences Discuss., doi:10.5194/bg-2017-46, 2017
Manuscript under review for journal Biogeosciences
Discussion started: 16 February 2017




Table 1. A subset of parameters optimized using data assimilation, prior distributions, and the sensitivity of total biomass at age 25 to the parameter. These are the parameters referred to in the results and discussion, other optimized model parameter can be found in the supplemental material.

| Parameter | Parameter description | Units | Sensitivity* | Prior distribution | Prior parameters | Reference for prior |
|---|---|---|---|---|---|---|
| $\alpha$ | Canopy quantum efficiency (pines) | mol C mol $PAR^{-1}$ | 0.84 | Uniform | Min = 0.02<br>Max = 0.1 | Vague |
| y | Ratio NPP/GPP | - | 0.84 | Uniform | Max= 0.66<br>Min = 0.30 | 1 |
| fCalpha700 | Proportional increase in canopy quantum efficiency between 350 and 700 ppm CO2 | - | 0.08 | Uniform | Min = 1.05<br>Max = 2.0 | Vague |
| fCpFS700 | Proportional decrease in allocation to foliage between 350 and 700 ppm CO2 | - | 0.00[#] | Uniform | Min = 0.50<br>Max = 1.00 | Vague |
| SWconst | Moisture ratio deficit when downregulation is 0.5 | - | 0.06 | Uniform | Min = 0.6<br>Max = 1.8 | 2, Vague |
| SWpower | Power of moisture ratio deficit | - | 0.06 | Uniform | Min = 1<br>Max= 13 | 2, Vague |
| FR1 | Fertility rating parameter 1 (mean annual temperature coefficient) | - | 0.23 | Uniform | Min = 0.0<br>Max = 1.0 | Vague |
| FR2 | Fertility rating parameter 2 (site index age 25 coefficient) | - | 0.39 | Uniform | Min = 0.0<br>Max = 1.0 | Vague |
| wSx1000 | Maximum stem mass per tree at 1000 trees/ha | kg $tree^{-1}$ | 0.43 | Normal | Mean = 235<br>Sd = 25 | 3,4 |
| thinPower | Power in self thinning law | - | 0.25 | Uniform | Min = 1.1<br>Max = 1.80 | 3,4 |
| pCRS | Ratio of coarse roots to stem allocation | - | 0.08 | Uniform | Min = 0.15<br>Max = 0.35 | 5 |

1(DeLucia et al., 2007);[2](Landsberg and Waring, 1997), [3](Bryars et al., 2013),[4](Gonzalez-Benecke et al., 2016), 5(Albaugh et al., 2005)
* Sensitivity is 1 when a 10% increase in the parameter results in a 10% change in total biomass. [#]Sensitivity is 0 when a 10% increase in the parameters does not change total biomass by a value greater than 0.01%.

Biogeosciences Discuss., doi:10.5194/bg-2017-46, 2017
Manuscript under review for journal Biogeosciences
Discussion started: 16 February 2017




1010

Table 2. Regional observational data streams used in data assimilation.

| Data stream | Measurement frequency | Measurement or estimation technique | Uncertainty | Stream ID for Table 4 |
|---|---|---|---|---|
| Foliage biomass (Pine) | Annual or less | Allometric relationship | Based on propagating the allometric model uncertainty in Gonzalez-Benecke et al. 2014. Varied by observation. | 1 |
| Foliage biomass (hardwood) | Annual or less | Allometric relationship | Assumed zero | 2 |
| Stem biomass (pine) | Annual or less | Allometric relationship | Based on propagating the allometric model uncertainty in Gonzalez-Benecke et al. 2014. Varied by observation. | 3 |
| Stem biomass (hardwood) | Annual or less | Allometric relationship | Assumed zero | 4 |
| Coarse root biomass (combined) | Annual or less | Allometric relationship | Standard deviation (SD) = 10% of observation | 5 |
| Fine root biomass (combined) | Annual or less | Allometric relationship | SD = 10% of observation | 6 |
| Foliage biomass turnover (combined) | Annual | Litterfall traps | SD = 2.5% of observation | 7 |
| Fine root biomass turnover (combined) | Annual | Mini-rhizotrons | SD = 10% of observation | 8 |
| Pine stem count | Annual or less | Counting individuals | 1% (assumed small) | 9 |
| Leaf area index (pine) | Monthly to annual | Litter traps or LI 2000 | If litter trap method: SD = 2.5% of observation<br>If LI-2000 method: SD = 10% of observation | 10 |
| Leaf area index (hardwood) | Monthly to annual | Litter traps or LI 2000 | If litter trap method: SD = 2.5% of observation<br>If LI-2000 method: SD = 10% of observation | 11 |
| Leaf area index (combined) | Only used if not separated into pine and hardwood | Litter traps or LI 2000 | If litter trap method: SD = 2.5% of observation<br>If LI-2000 method: SD = 10% of observation | 12 |
| Gross Ecosystem Production | Monthly | Modeled from flux eddy-covariance net ecosystem exchange | SD = 10% of observation | 13 |
| Evapotranspiration | Monthly | Eddy-covariance | SD = 10% of observation | 14 |

1011
1012

Biogeosciences Discuss., doi:10.5194/bg-2017-46, 2017
Manuscript under review for journal Biogeosciences
Discussion started: 16 February 2017


Table 3. Descriptions of the studies used in data assimilation.

| Study name | Number of locations | Number of plots per site | Experimental treatments (plots) | Data streams (Table 2) | Measurement Years | Measurement Stand Ages (years) | Reference |
|---|---|---|---|---|---|---|---|
| FMRC[1] Thinning Study | 163 | 1 | None | 1, 3,9 | 1981 - 2003 | 8 - 30 | (Burkhart et al., 1985) |
| FPC[2] Region-wide 18 | 18 | 2 | Nutrient addition | 1, 3,9 | 2011-2014 | 12-21 | (Albaugh et al., 2015) |
| PINEMAP[3] | 4 | 16 | Nutrient addition, 30% throughfall, Nutrient x throughfall | 1, 3,9 | 2011-2015 | 3 – 13 | (Will et al., 2015) |
| Waycross | 1 | 2 | Nutrient addition | 3,9,10 | 1991-2010 | 4-23 | (Bryars et al., 2013) |
| SETRES[4] | 1 | 16 | Nutrient addition, irrigation, nutrient x irrigation | 1,3,5,6,9, 10 | 1991-2006 | 8 - 23 | (Albaugh et al., 2004) |
| Duke FACE[5] and flux | 1 | 12 | $CO_2$, nutrient addition, $CO_2$ x nutrient addition | 2,3,4,5,6, 7,8,9,10, 11,13,14 | 1996-2004 | 13-22 | (McCarthy et al., 2010; Novick et al., 2015) |
| NC2 Flux | 1 | 1 | None | 2,3,4,5,6, 7,9,10,11 ,12,13,14 | 2005-2014 | 12-22 | (Noormets et al., 2010) |
| Total | 187 | 294 | | | 1981 - 2014 | 4 - 30 | |

[1]Forest Modeling Research Cooperative; [2] Forest Productivity Cooperative; [3] Pine Integrated Network: Education, Mitigation, and Adaptation project (PINEMAP); [4] Southeast Tree Research and Education Site; [5] Free Air Carbon Enrichment

Biogeosciences Discuss., doi:10.5194/bg-2017-46, 2017
Manuscript under review for journal Biogeosciences
Discussion started: 16 February 2017


Table 4. Description of the different data assimilation approaches used.

| Simulation Name | Treatments included in assimilation | Number of plots |
|---|---|---|
| All | 1-stage data assimilation. All plots and experiments in the region were used simultaneously. | 294 |
| DK+NC2 | 1st stage of 2-stage assimilation. All plots at the Duke eddy flux (DK3), Duke Free Air $CO_2$ Enrichment Study, and NC2 eddy flux site; includes $CO_2$ enrichment and nutrient addition experiments at the Duke site | 13 |
| DK+NC2-fert | 1st stage of 2-stage assimilation. Same as DK+NC2 but without nutrient fertilization plots | 10 |
| RW | 2nd stage of 2-stage assimilation. Region-wide assimilation of FRMC, FPC, PINEMAP, Waycross, and SETRES sites. Uses the posteriors of the DK+NC2 simulation as priors. Includes nutrient addition and water manipulation experiments. This simulation is repeated four times for four different out-of-sample validation plots. | 281 |
| RW-fert | 2nd stage of 2-stage assimilation. Same as RW but without nutrient addition experiments; uses the posteriors of the DK+NC2-fert simulation as priors | 222 |
| RW-water | 2nd stage of 2-stage assimilation. Same as RW but without water manipulation experiments | 241 |

1018
1019

Biogeosciences Discuss., doi:10.5194/bg-2017-46, 2017
Manuscript under review for journal Biogeosciences
Discussion started: 16 February 2017



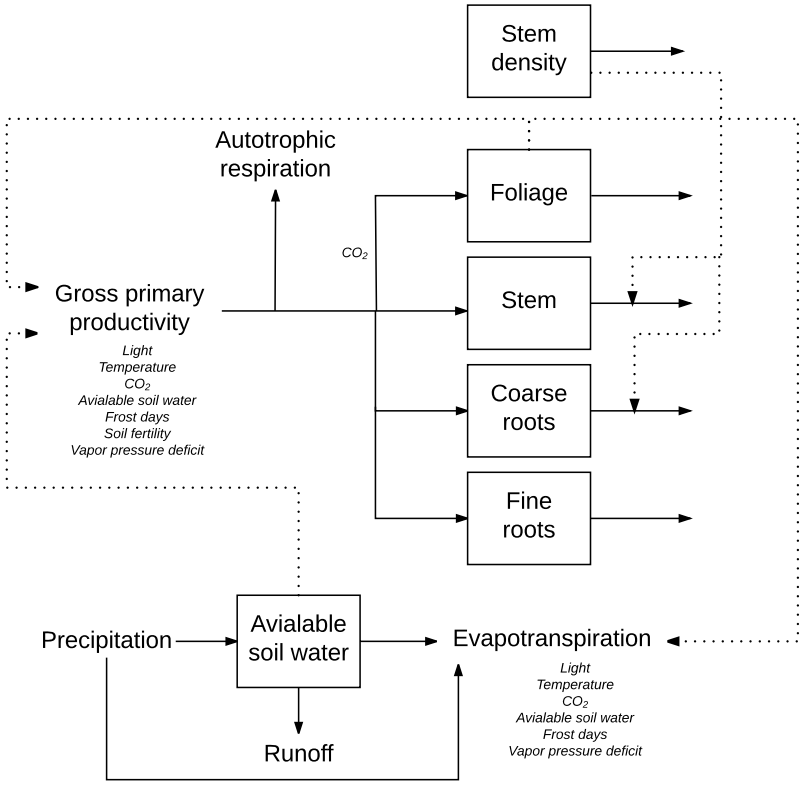

Table 5. Posterior means and 95% credible intervals for parameters listed in Table 1 using the data assimilation approaches listed in Table 4.

| Parameter | RW | RW-fert | RW-water | DK+NC2 | All |
|---|---|---|---|---|---|
| $\alpha$ | 0.037 | 0.040 | 0.037 | 0.035 | 0.032 |
| | (0.034 – 0.040) | (0.036 – 0.045) | (0.035 – 0.040) | (0.030 – 0.042) | (0.030 – 0.035) |
| y | 0.48 | 0.48 | 0.48 | 0.48 | 0.52 |
| | (0.46 – 0.51) | (0.45-0.51) | (0.46 – 0.51) | (0.45 – 0.51) | (0.50 – 0.54) |
| fCalpha700 | 1.31 | 1.31 | 1.31 | 1.32 | 1.11 |
| | (1.22 – 1.40) | (1.22 – 1.40) | (1.22 – 1.40) | (1.23 – 1.41) | (1.08 – 1.15) |
| fCpFS700 | 0.84 | 0.83 | 0.84 | 0.84 | 0.99 |
| | (0.75 – 0.93) | (0.75 – 0.93) | (0.75 – 0.93) | (0.76-0.93) | (0.95 – 1.0) |
| SWconst | 1.48 | 1.31 | 1.8 | 1.30 | 1.57 |
| | (1.09 – 1.85) | (0.95 – 1.70) | (1.47 – 2.15) | (0.89 – 1.76) | (1.08 – 1.79) |
| SWpower | 1.61 | 1.29 | 2.93 | 2.20 | 1.47 |
| | (0.90 – 2.46) | (0.78 – 1.98) | (1.48 – 3.82) | (1.47 – 3.44) | (1.09 – 2.26) |
| FR1 | 0.094 | 0.096 | 0.118 | not fit | 0.094 |
| | (0.086 – 0.104) | (0.088 – 0.103) | (0.110 – 0.128) | | (0.087 – 0.102) |
| FR2 | 0.144 | 0.124 | 0.179 | not fit | 0.153 |
| | (0.133 – 0.154) | (0.108 – 0.142) | (0.156 – 0.182) | | (0.140 – 0.168) |
| wSx1000 | 176 | 180 | 180 | 258 | 181 |
| | (171 – 181) | (174 – 186) | (176 – 186) | (228 – 295) | (174 0 187) |
| thinPower | 1.67 | 1.70 | 1.71 | 1.28 | 1.61 |
| | (1.60 – 1.74) | (1.63 – 1.78) | (1.65 – 1.78) | (1.12 – 1.60) | (1.51 – 1.69) |
| pCRS | 0.26 | 0.24 | 0.25 | 0.17 | 0.28 |
| | (0.25 – 0.27) | (0.23 – 0.25) | (0.24 – 0.26) | (0.16 – 0.19) | (0.27 – 0.29) |

1020

Biogeosciences Discuss., doi:10.5194/bg-2017-46, 2017
Manuscript under review for journal Biogeosciences
Discussion started: 16 February 2017


1021 Figures
1022




1023
1024 Figure 1. A diagram of the monthly time-step 3-PG model used in this study. The stocks are represented by the
1025 boxes and the fluxes by the arrows. An influence of a stock on a flux that is not directly related to that stock is
1026 represented by the dotted lines. The environmental influences on a flux is described using italics. A description of
1027 the model can be found in the supplemental information.
1028
1029

Biogeosciences Discuss., doi:10.5194/bg-2017-46, 2017
Manuscript under review for journal Biogeosciences
Discussion started: 16 February 2017

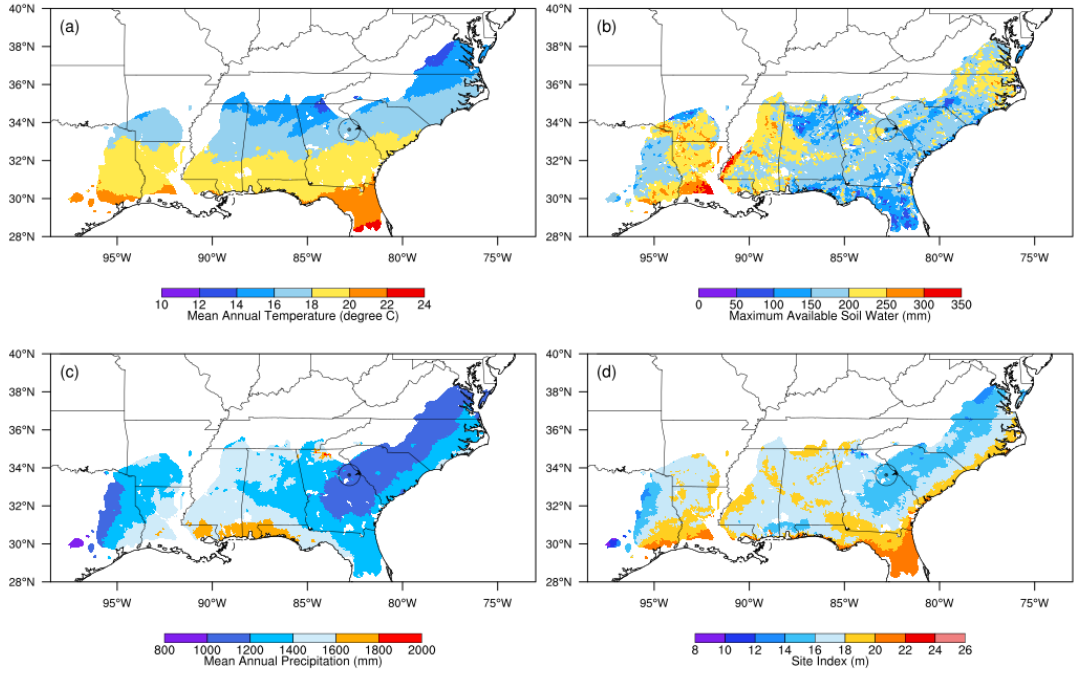




Figure 2. Key climatic and stand characteristic inputs to the regional 3-PG simulations: (a) Mean annual temperature (1979-2011) as a summary of the gradient in monthly temperature inputs used in simulations, (b) maximum available soil water for the top 1.5 meters of soil from SSURGO, (c) mean annual precipitation (1979-2011) as a summary of the gradient in monthly precipitation inputs used in simulations, and (d) site index. The focal site in Georgia highlighted in Figures 5c and 6b is represented by the circle containing the dot. The area shown is the natural range of loblolly pine (*Pinus taeda L.*).

Biogeosciences Discuss., doi:10.5194/bg-2017-46, 2017
Manuscript under review for journal Biogeosciences
Discussion started: 16 February 2017

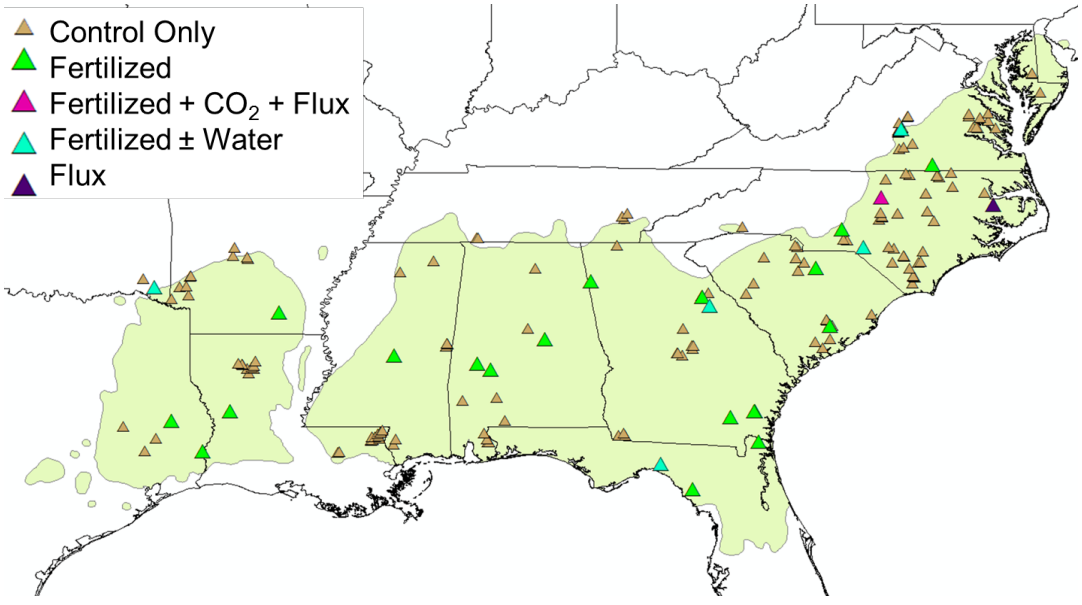




Figure 3. Map of loblolly pine distribution, plot locations used in data assimilation, and the experiment type associated with each plot. The control-only treatments were plots without any associated experimental treatment or flux measurements. Fertilized were plots with nutrient additions. $CO_2$ were plots with free-air concentration enrichment treatments. The flux treatments were plots with eddy-covariance measurements of ecosystem-scale carbon and water exchange. The water treatments included throughfall exclusion and irrigation experiments.

Biogeosciences Discuss., doi:10.5194/bg-2017-46, 2017
Manuscript under review for journal Biogeosciences
Discussion started: 16 February 2017

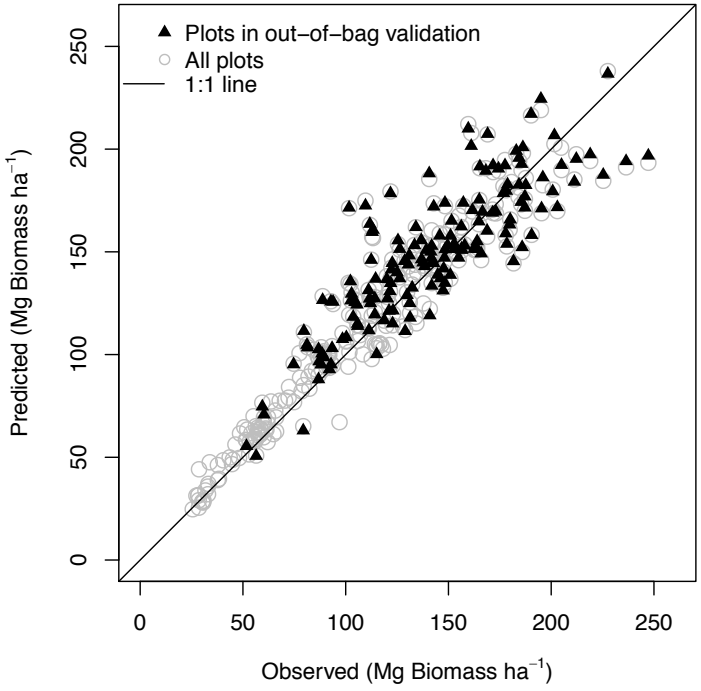




Figure 4. Model evaluation of stem biomass using the RW simulation described in Table 5. The gray circles correspond to predictions where all plots were used in data assimilation. The black triangles correspond to predictions where 120 plots were not included in data assimilation and represent an independent evaluation of model predictions (out-of-bag validation). For each plot, we used the measurement with the longest interval between initialization and measurement for evaluation.

Biogeosciences Discuss., doi:10.5194/bg-2017-46, 2017
Manuscript under review for journal Biogeosciences
Discussion started: 16 February 2017

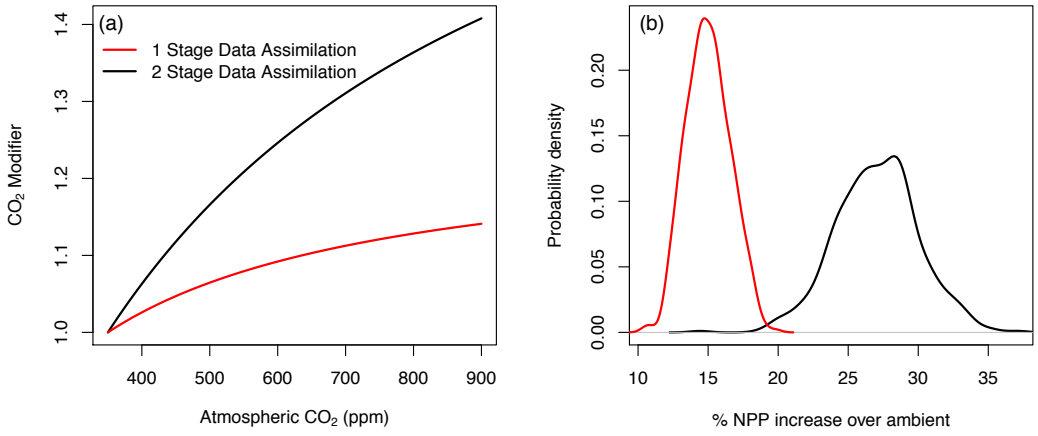




Figure 5. The influence of the data assimilation approach on predictions of how productivity responds to atmospheric $CO_2$. (a) The relationship between atmospheric $CO_2$ concentration and the modifier of light-use efficiency when all plots and experiments are assimilated simultaneously (1 stage) and when the Duke and NC2 plots are assimilated before assimilating the remaining observations across the region (2 stage). (b) The probability distribution of predicted response of NPP to the elevated $CO_2$ at the Duke FACE experiment for the two assimilation approaches. Uncertainty was estimated by integrating the parameter uncertainty estimated through data assimilation (see Methods).

Biogeosciences Discuss., doi:10.5194/bg-2017-46, 2017
Manuscript under review for journal Biogeosciences
Discussion started: 16 February 2017

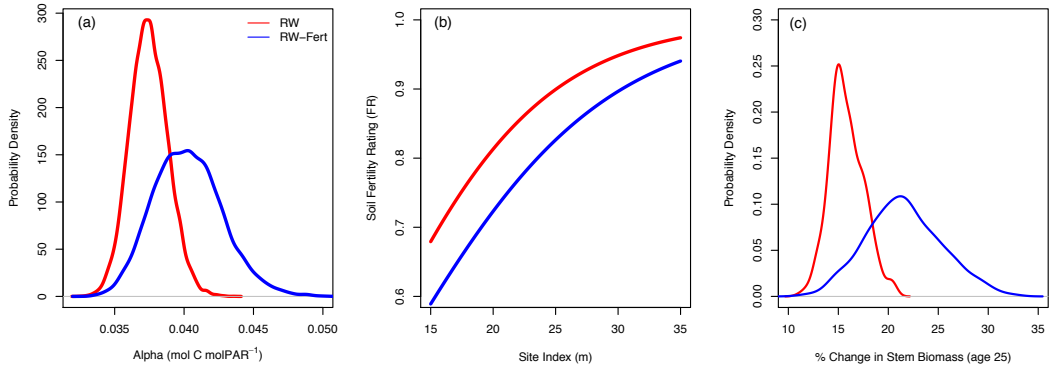
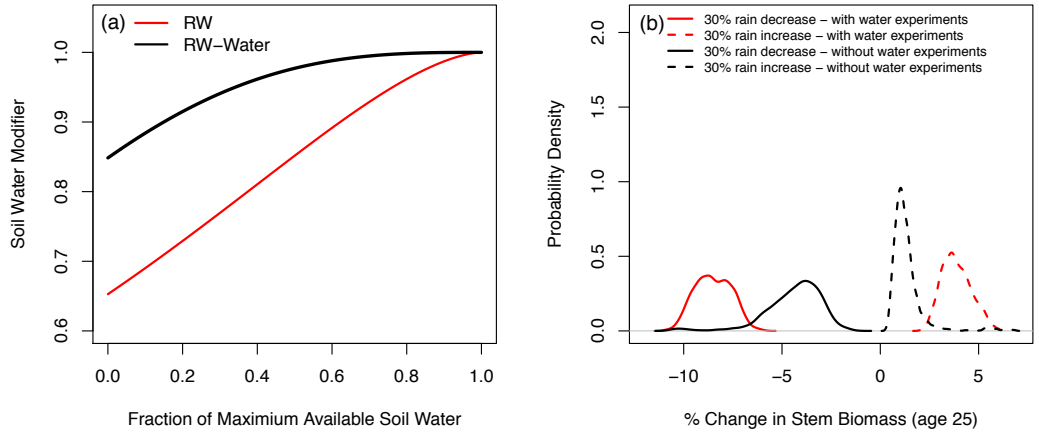




Figure 6. The influence of the data assimilation (DA) approach on predictions of light-use efficiency, nutrient limitation, and how productivity responds to nutrient addition. (a) The posterior distribution for the potential light-use efficiency parameter (alpha). (b) The relationship between site index and the nutrient limitation modifier in 3-PG (FR). (c) The predicted distribution of the response of stem biomass to nutrient fertilization (setting FR = 1) at the focal site in Georgia. The red line corresponds to DA that included nutrient addition experiments. The blue line corresponds to DA that did not include nutrient addition experiments.

Biogeosciences Discuss., doi:10.5194/bg-2017-46, 2017
Manuscript under review for journal Biogeosciences
Discussion started: 16 February 2017

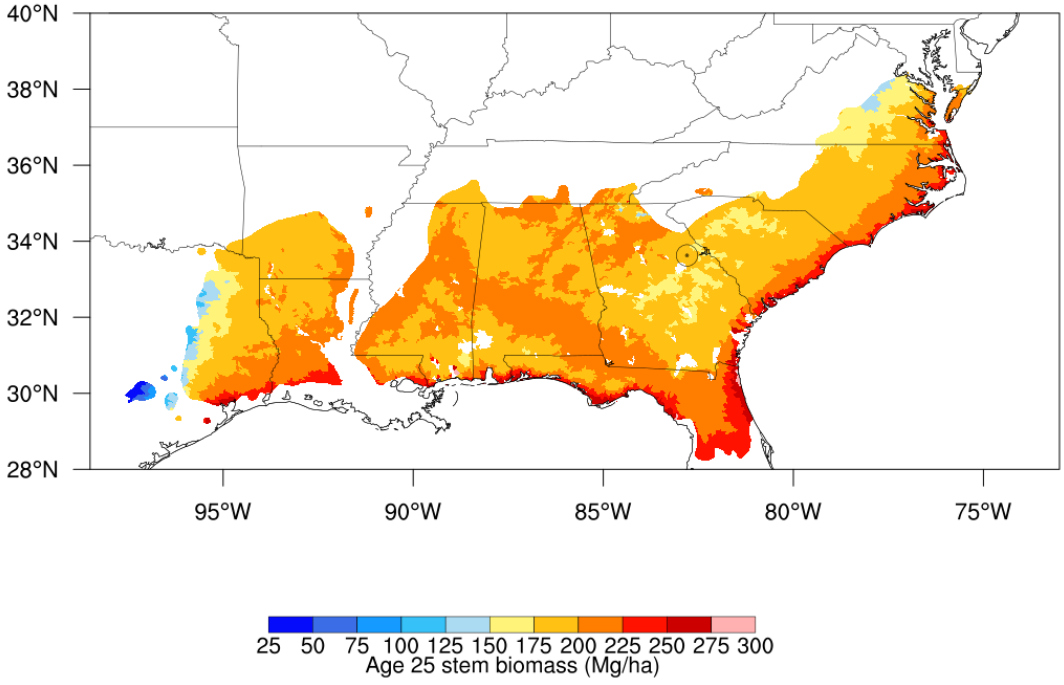

1077




1078
1079 Figure 7. The influence of the data assimilation (DA) approach on predictions of water limitation
1080 and how productivity responds to a change in precipitation. (a) The relationship between fraction
1081 of maximum available soil water, as predicted by 3-PG, and the modifier of light-use efficiency
1082 and canopy conductance. (b) The predicted distribution of the response of stem biomass to a 30%
1083 increase (dashed lines) and a 30% decrease (solid lines) in precipitation at the focal site in
1084 Georgia. The red line corresponds to DA that included water manipulation experiments. The
1085 black line corresponds to DA that did not include water manipulation experiments.
1086
1087

Biogeosciences Discuss., doi:10.5194/bg-2017-46, 2017
Manuscript under review for journal Biogeosciences
Discussion started: 16 February 2017


1088

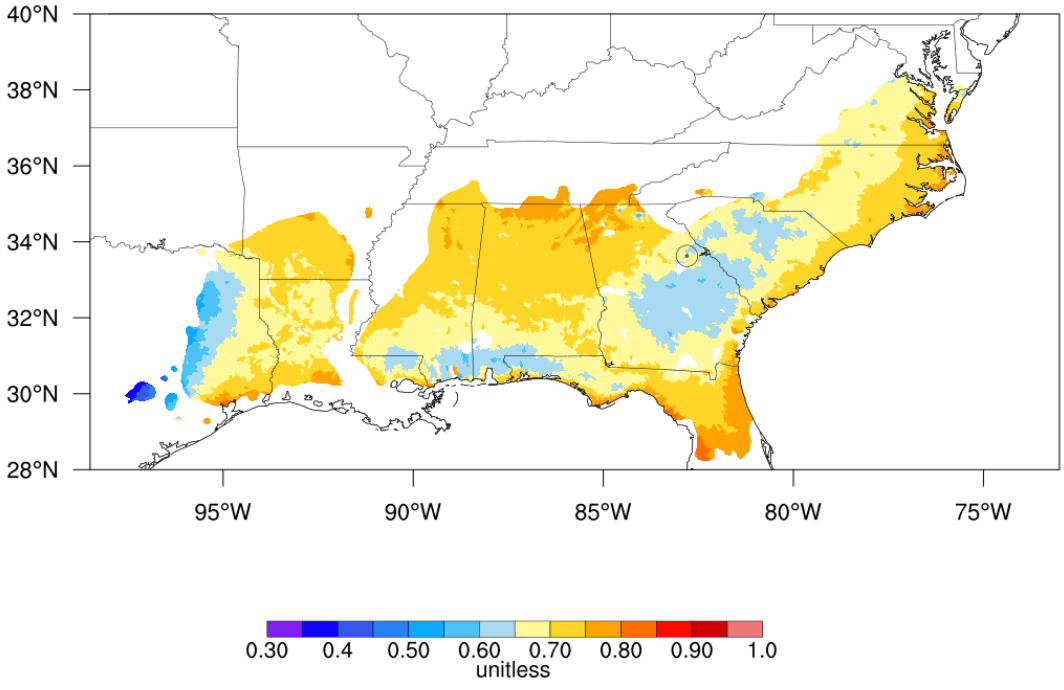



1089
1090 Figure 8. Regional predictions of stem biomass stocks for a 25-year-old stand planted in 1979.
1091 Parameters used the predictions were from the RW data assimilation approach described in Table
1092 5. The focal site in Georgia highlighted in Figures 5c and 6b is represented by the circle
1093 containing the dot.
1094

Biogeosciences Discuss., doi:10.5194/bg-2017-46, 2017
Manuscript under review for journal Biogeosciences
Discussion started: 16 February 2017



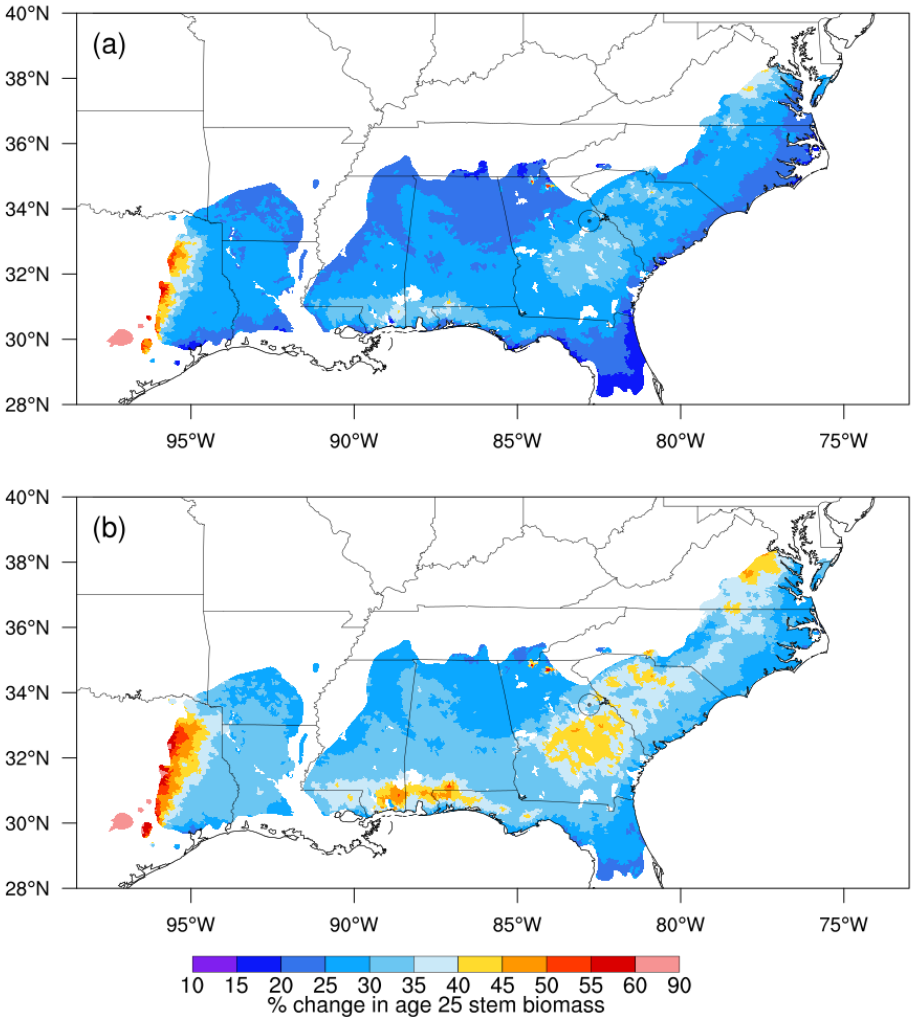

1095
1096
1097




1098
1099 Figure 9: Regional predictions of the soil fertility factor (FR) used in 3-PG to define the nutrient
1100 status of the simulated stand. Parameters used the predictions were from the RW data
1101 assimilation approach described in Table 5. The focal site in Georgia highlighted in Figures 5c
1102 and 6b is represented by the circle containing the dot.

Biogeosciences Discuss., doi:10.5194/bg-2017-46, 2017
Manuscript under review for journal Biogeosciences
Discussion started: 16 February 2017


1103




1104

1105 Figure 10: Regional predictions of the change in stem biomass of a 25-year stand when nutrient
1106 limitation is completely removed through nutrient addition (simulated by setting FR = 1).
1107 Predictions from data assimilation that included nutrient addition experiments are shown in (a)
1108 and prediction data assimilation that did not include nutrient addition experiments are shown in
1109 (b). The focal site in Georgia highlighted in Figures 5c and 6b is represented by the circle
1110 containing the dot.
1111

Biogeosciences Discuss., doi:10.5194/bg-2017-46, 2017
Manuscript under review for journal Biogeosciences
Discussion started: 16 February 2017


Biogeosciences Discussions
Open Access
EGU

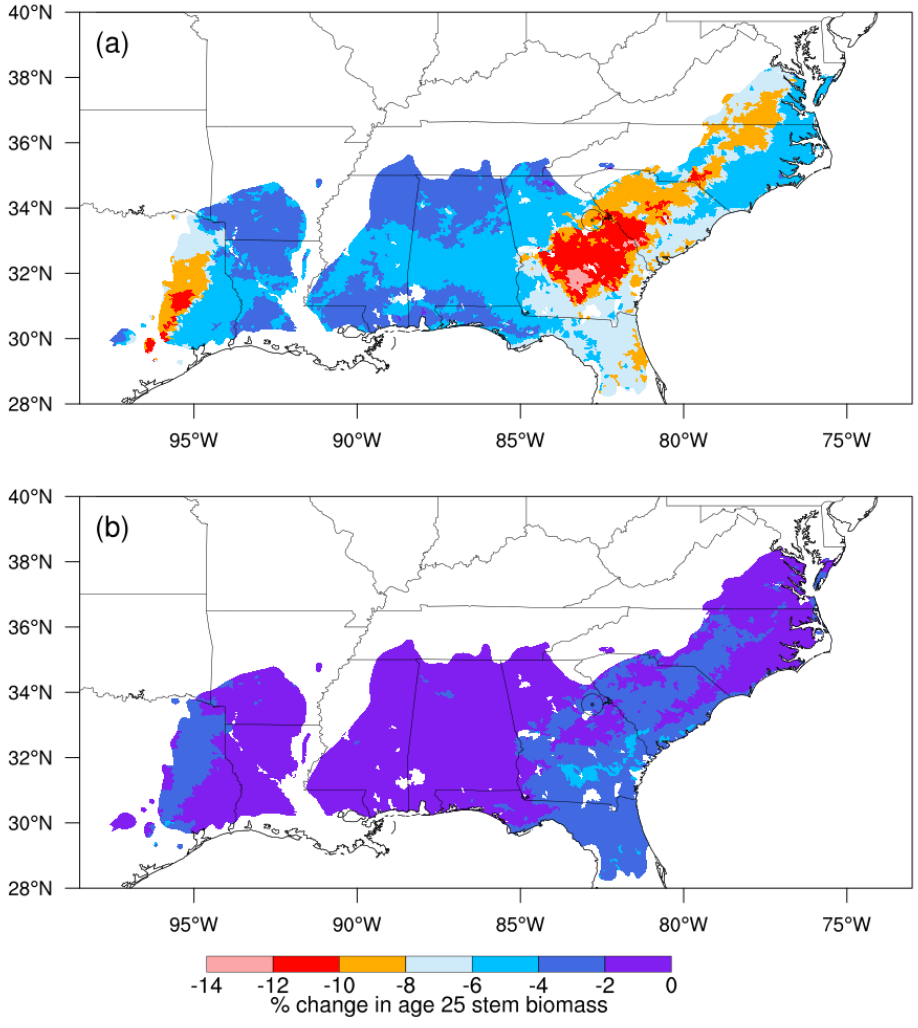



Figure 11. Regional predictions of the change in stem biomass of a 25-year stand when annual precipitation is reduced by 30%. Predictions from data assimilation that included water manipulation experiments are shown in (a) and prediction data assimilation that did not include water manipulation experiments are shown in (b). The focal site in Georgia highlighted in Figures 6c and 7b is represented by the circle containing the dot.