# 1 Title: Facilitating feedbacks between field 2 measurements and ecosystem models

## 3 Running Title: Feedbacks between measurements and models

1. David S. LeBauer1*,*∗ Dan Wang1 Katherine T. Richter2 Carl C. Davidson2 Michael C.
2. Dietze1*,*2
3. 1 Energy Biosciences Institute, 2 Department of Plant Biology, University of Illinois,
4. Urbana, Illinois, USA

 Corresponding author

1. Telephone: (949) 433-7410
2. Fax: (217) 244-3637
3. E-mail: dlebauer@illinois.edu

# 12 Abstract

13 Ecological models help us understand how ecosystems function, predict responses to 14 global change, and identify future research needs. However, widespread use of models is

1. limited by the technical challenges of model-data synthesis and information management.
2. To address these challenges, we present a ecoinformatic workflow, the Predictive
3. Ecosystem Analyzer (PEcAn), that facilitates model analysis. Herein we describe the
4. PEcAn modules that synthesize plant trait data to estimate model parameters, propagate
5. parameter uncertainties through to model output, and evaluate the contribution of each
6. parameter to model uncertainty. We illustrate a comprehensive approach to the
7. estimation of parameter values, starting with a statement of prior knowledge that is
8. refined by species level data using Bayesian meta-analysis; this is the first use of a
9. rigorous meta-analysis to inform the parameters of a mechanistic ecosystem model.
10. Parameter uncertainty is propagated using ensemble methods to estimate model
11. uncertainty. Variance decomposition allows us to quantify the contribution of each
12. parameter to model uncertainty; this information can be used to prioritize subsequent
13. data collection. By streamlining the use of models and focusing efforts to identify and
14. constrain the dominant sources of uncertainty in model output, the approach used by 29 PEcAn can speed scientific progress.
15. We demonstrate PEcAn’s ability to incorporate data to reduce uncertainty in
16. productivity of a perennial grass monoculture (*Panicum virgatum* L.) modeled by the
17. Ecosystem Demography model. Prior estimates were specified for fifteen model
18. parameters, and species-level data were available for seven of these. Meta-analysis of
19. species-level data substantially reduced the contribution of three parameters (specific leaf
20. area [SLA], maximum carboxylation rate, and stomatal slope) to overall model
21. uncertainty. By contrast, root turnover rate, root respiration rate, and leaf width had
22. little effect on model output, therefore trait data had little impact on model uncertainty.
23. For fine root allocation the decrease in parameter uncertainty was offset by an increase in
24. model sensitivity. Remaining model uncertainty is driven by growth respiration, fine root
25. allocation, leaf turnover rater, and SLA. By establishing robust channels of feedback 41 between data collection and ecosystem modeling, PEcAn provides a framework for more 42 efficient and integrative science.
26. **keywords:** traits, ecoinformatics, ecophysiology, Ecosystem Demography, sensitivity
27. analysis, variance decomposition, ecological forecast, Bayesian, meta-analysis, ecosystem
28. model

# 46 Introduction

1. In the face of unprecedented global change there is growing demand for predictions of
2. ecosystem responses that provide actionable information for policy and management 49 (Clark et al., 2001). Currently, the response of the terrestrial biosphere remains one of
3. the largest sources of uncertainty in projections of climate change (Denman et al., 2007).
4. This uncertainty comes from a combination of the uncertainties about our conceptual
5. understanding of ecological systems, as captured by the structure and assumptions of the
6. models used to make ecological forecasts, the uncertainties in the parameters of these
7. models, and the uncertainties associated with the underlying data itself (McMahon et al.,
8. 2009). Reducing these uncertainties requires that we be able to synthesize existing 56 information, efficiently identify the dominant sources of model uncertainty and target 57 them with further field research.
9. Despite the acknowledged importance of these activities, there is often a disconnect
10. between model simulation and data collection. Both model-data synthesis and the
11. investigation of uncertainty remain challenging, while the use of models to quantitatively
12. inform data collection is extremely rare. Most modeling uses a single point estimate for
13. each parameter, effectively treating each parameter value as completely certain. However,
14. such point estimates do not account for the degree to which we understand a parameter
15. based on observations. Furthermore, the rationale for a particular estimate is often
16. unclear, as is the degree to which the estimate represents the process being observed or
17. its representation in a model. In many cases, parameter values are chosen iteratively to
18. “tune”, or “calibrate” the model output to observations. A first step toward constraining 68 model uncertainty is to account for uncertainty in model parameters instead of relying on 69 point estimates.

70 More rigorous approaches to estimating parameter values include model optimization and 71 data assimilation (Medvigy et al., 2009; Reichstein et al., 2003), as well as Bayesian

1. model-data fusion (Luo et al., 2011). However, these approaches have generally started
2. with uninformative or vague prior estimates of model parameters. These vague priors
3. ignore available data that could directly inform parameter values; the most commonly
4. used vague prior distribution is a uniform. A uniform prior assigns equal probability to
5. parameter values over its entire range, in many cases over many orders of magnitude. The
6. use of such vague priors often exacerbates problems with equifinality (Richardson and
7. Hollinger, 2005; Williams et al., 2009; Luo et al., 2009) which can produce unidentifiable
8. parameters, as well as biologically unrealistic parameter sets that generate the right model
9. output for the wrong reasons (Beven and Freer, 2001; Beven, 2006; Williams et al., 2009).
10. Another reason to use informed priors is to take advantage of one of the key strengths of
11. the Bayesian paradigm: the ability to synthesize multiple sources of information in a
12. rigorous and consistent framework. For example, plant traits related to leaf stoichiometry 84 and photosynthetic capacity are often well constrained by previous research (Skillman,
13. 2008; Reich and Oleksyn, 2004; Wright et al., 2004; Wullschleger, 1993), while other traits,
14. such as root respiration rate, are more difficult to measure and data are sparse. Informed
15. priors allow existing information to be formally integrated into model parameterization,
16. even if there is no data for the particular species or plant functional type (PFT) being 89 measured; the level of confidence in a parameter value is reflected in its variance.
17. Models have rarely been used to quantify the value of data with respect to reducing
18. uncertainty. Instead, data collection is often focused on answering specific questions in
19. specific spatial, temporal, and taxonomic contexts. In these contexts, the value of a
20. particular data set is based on the ability to answer a particular question. However, the
21. same data set may have a very different value in the context of reducing model
22. uncertainty. For example, a single data point used to inform a poorly understood but
23. influential model parameter can reduce model uncertainty more than a large collection of 97 data on a trait that is relatively well studied. In a modeling context, the value of an
24. additional data point depends both on how much it constrains parameter uncertainty and
25. the sensitivity of model output to the parameter. Thus, the ability to comprehensively
26. utilize available data in model parametrization can help to identify gaps in existing
27. knowledge, improve the ability of models to account for current understanding, and 102 inform data collection efforts by identifying the knowledge gaps most responsible for
28. uncertainty.
29. While the increasing sophistication of model-data fusion and uncertainty accounting is a
30. critical step in the right direction, the complexity of such analyses can make models even
31. less accessible. One of the reasons to make models more accessible, and to make them
32. better at synthesizing existing data, is that they are fundamentally a formal, quantitative
33. distillation of our current understanding of how a system works. As such, models can be
34. used to identify gaps in our understanding and target further research. This feedback
35. between models and data could be improved if models were routinely evaluated in a way
36. that quantifies the value of data with respect to reducing uncertainty. We fundamentally
37. believe that streamlining the informatics of modeling – the need to track, process, and
38. synthesize data and model output – will make the development and application of 114 ecological data and models more accessible, transparent, and relevant.
39. In this paper we present the Predictive Ecosystem Analyzer (PEcAn) as a step toward
40. meeting these objectives. PEcAn is a scientific workflow that manages the flows of data
41. used and produced by ecological models, and that assists with model parametrization,
42. error propagation, and error analysis. PEcAn accomplishes two goals: first, it synthesizes
43. data and propagates uncertainty through an ecosystem model; second, it places an
44. information value on subsequent data collection that enables data collection that
45. efficiently reduces uncertainty. In addition to quantifying the information content of any
46. prediction or assessment, these techniques also help identify the gaps in our knowledge of 123 ecological and biogeochemical processes (Saltelli et al., 2008).
47. PEcAn addresses the challenge of synthesizing plant trait data from the literature in a
48. way that accounts for the different scales and sources of uncertainty. Available data is 126 synthesized using a Bayesian meta-analysis, and the meta-analysis posterior estimates of 127 plant traits are used as parameters in an ecosystem model.
49. A model ensemble is a set of model runs with parameter values drawn from the
50. meta-analysis posteriors estimate of plant traits. Output from a model ensemble
51. represents the posterior predictive distributions of ecosystem responses that account for
52. trait parameter uncertainty (hereafter “model posterior” refers to the “model ensemble
53. output”). Sensitivity analysis and variance decomposition help to determine which traits
54. (model parameters) drive uncertainty in ecosystem response (model posterior) (Saltelli
55. et al., 2008; Larocque et al., 2008). These analyses help target parameters for further
56. constraint with trait data, forming a critical feedback loop that drives further field
57. research and provides an informative starting point for data assimilation. Here we 137 illustrate an application of PEcAn to the assessment of aboveground yield in a perennial 138 grass monoculture.
58. In the sections below, we provide an overview of the components of PEcAn’s integrated
59. framework for data synthesis and ecological prediction. We start with a description of the
60. methods implemented in the workflow (Implementation). This includes descriptions of
61. the database, Bayesian meta-analysis, ensemble analysis, sensitivity analysis, and
62. variance decomposition. Finally we present an example of the application of the system
63. (Application) to analyze the aboveground biomass of switchgrass (*Panicum virgatum* L.),
64. by the Ecosystem Demography model, version 2.1 (ED2) (Medvigy et al., 2009;
65. Moorcroft et al., 2001).

# 147 Implementation

## 148 PEcAn workflow

1. The Predictive Ecosystem Analyzer (PEcAn) manages the flow of information into and
2. out of ecosystem models.PEcAn is not a model itself, it is a scientific workflow consisting
3. of discrete steps, or modules. Individual modules are building blocks of the workflow,
4. represented by the rectangles in Figure 1, while flows of information are represented by
5. arrows. This makes PEcAn an encapsulated, semi-automated system for model 154 parametrization, error propagation, and analysis.
6. A central objective of the PEcAn workflow is to make the entire modeling process
7. transparent, reproducible, and adaptable to new questions (*sensu* Stodden et al., 2010;
8. Ellison, 2010). To achieve this objective, PEcAn’s adheres to “best practice” guidelines
9. for ecological data management and provenance tracking (Jones et al., 2006; Michener 159 and Jones, 2012).
10. PEcAn uses a database to track data provenance and a settings file to control workflow
11. analyses and model runs. The database records the site, date, management, species, and
12. treatment information for each trait observation used in the meta-analysis. Settings 163 related to the experimental design and computation are set and recorded in a separate file 164 for each analysis.

165 The PEcAn source code, as well as the inputs and output used in the analysis described 166 below (see Application) are provided as an appendix. However, new users are encouraged 167 to utilize the latest release available on the project web site (www.pecanproject.org).

1. This site also provides a virtual machine and a web-interface that minimize the effort
2. required to run PEcAn and begin using an ecosystem model. The PEcAn “virtual
3. machine” provides all of the required software dependencies in a pre-configured desktop
4. environment that can be run on any standard operating system using a freely available
5. virtualization software such as VirtualBox (www.virtualbox.org/) or VmWare Player
6. (www.vmware.com). The virtual machine minimizes the installation time and
7. pre-requisite knowledge required to perform analyses, and can be used to support 175 investigation, development, and education. The web interface is even easier to use, but 176 does not provide access to all of PEcAn’s functionality.
8. The PEcAn software is primarily written in R and developed in a Linux environment. It
9. also relies on a MySQL database, bash, JAGS, and specialized R packages. PEcAn has a
10. family of model-specific functions that manage the details of launching of model runs and 180 reading model output.
11. Although PEcAn does not depend on any specific model, it was developed to support
12. ecosystem models that run in high-performance computing environments; for this reason,
13. it is capable of running models locally, remotely, or through queuing systems regardless of
14. whether PEcAn is compiled locally or run as a virtual machine. The PEcAn 1.1 release
15. described herein runs with the Ecosystem Demography model, in addition, the current
16. release also supports SIPNET (Moore et al., 2008) and near term support for IBIS 187 (Kucharik et al., 2000), DayCent (Parton et al., 1998), and BioCro (Miguez et al., 2009) 188 is under development.

## 189 Trait Database

1. Model parameters are associated with corresponding prior distributions, and in many
2. cases, with species-level data. Both prior distributions and data are stored in a relational
3. database (Appendix B). PEcAn directly accesses the database, which contains additional
4. meta-data for each data set, including site descriptions, measurement conditions,
5. experimental details, and citations.

## 195 Trait Priors

1. A fundamental component of the Bayesian approach to parameter estimation is the use of
2. priors. Priors formally incorporate knowledge of a parameter based on previous studies
3. into a new analysis. In the current study, we leverage previously collected data from
4. non-target species to place biologically informed constraint on the distribution of a plant
5. trait parameter. When additional data for a specific species or plant functional type is
6. available, priors are further constrained before being used as model parameters. When no 202 additional data are available, these priors are used directly to parameterize the model.
7. For the *P. virgatum* example described below, priors were set using data from all plant
8. species, from only grass species, or from just C4 grass species depending on availabile 205 data. Sources of this prior information included data from previous and ad-hoc syntheses, 206 expert knowledge, and biophysical constraints (Table 1).
9. Prior distributions used in the meta-analysis were fit to one of four types of information:
10. 1) data from multiple species, 2) the posterior predictive distribution for a new species
11. from a meta-analysis of data (when error estimates were available), 3) a central tendency
12. informed by data with expert constraint on the confidence interval, or 4) expert
13. constraints on both the central tendency and confidence intervals. In case number 2, the
14. across-species meta-analysis “posterior” informs the prior for the species-level
15. meta-analysis. In all cases, maximum likelihood estimation was used to fit a prior
16. distribution. When more than one candidate distribution was considered, Aikake’s
17. Information Criterion (AIC) was used to select the best fit distribution. The choice of 216 prior was confirmed by visually inspecting the prior density functions overlain by data or

217 expert constraints (Figure 2).

## 218 Meta-analysis

219 A Hierarchical Bayes meta-analytical model (Figure 3) formally synthesizes available trait 220 data from multiple studies while accounting for various sources of uncertainty. This 221 Hierarchical Bayes approach integrates prior information and provides a flexible approach 222 to variance partitioning and parameter estimation.

1. The meta-analytical framework is useful for summarizing data sets that include summary
2. statistics. The trait data queried by PEcAn consist of a trait name, sample mean, sample 225 size, and a sample error statistic. PEcAn transforms error statistics to exact or

2

1. conservative (i.e., erring toward inflating the variance) estimates of precision (*τ* = 1*/*SE )
2. (Appendix C).
3. The sample mean is drawn from a normal distribution:

*Yk* ∼ *N*(Θ*k,τk*) (1)

1. Where *Yk* is the sample mean of the *kth* unique site by treatment combination (sample 230 unit), Θ*k* is the unobserved ’true’ value of the trait for the *kth* sample unit.
2. The meta-analysis partitions trait variability into among site, among treatment, and
3. within-unit variance. The unobserved ’true’ trait mean Θ*k* is a linear function of the 233 global trait mean, *β*0 plus random effects for study site (*β*site*j*) and treatment (*β*tr|site*ij*) 234 and a fixed effect for greenhouse (*β*gh):

Θ*k* = *β*0 + *β*site*i* + *β*tr|site*ij* + *β*gh*I*(*i*) (2)

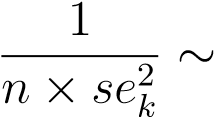
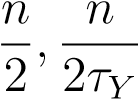
1. Where *i* indexes study site, *j* indexes each treatment within a study, and *I*(*i*) is an
2. indicator variable set to 0 for field studies and 1 for studies conducted in a greenhouse,
3. growth chamber, or pot experiment. The parameter used in the ecosystem model is the
4. posterior estimate of the global mean trait value, *β*0. *β*0, has an informed prior functional 239 form and parameter specification that varies by trait and species or PFT. Methods used
5. to elicit priors for the present study are provided in the Application section under Priors.
6. The “site” random effects (*β*site), accounts for the spatial (among-site) heterogeneity of a
7. parameter. The “treatment” random effect (*β*tr|site) accommodates differences among
8. experimental treatments. These random effects of treatment and site are assumed to be
9. Normally distributed with zero mean and they have diffuse Gamma priors on precision
10. *τ*site and *τ*tr. Control treatments and observational studies have *β*tr|site = 0. PEcAn
11. dynamically adjusts the meta-analysis model specification to include terms for each level
12. of site and treatment, or greenhouse studies as required by available data. To ensure that
13. the prior on precision remains sufficiently diffuse when the magnitude of a parameter is
14. small, the scale parameters in the gamma priors on random effect precision terms (*τ*site

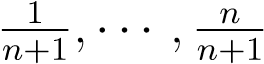
√

1. and *τ*tr|site) are scaled to (*β*¯02*/*1000) when the prior on *β*0 has a mean *β*¯0 *<* 10.
2. A “greenhouse” fixed effect *β*gh accounts for potential biases associated with plants grown 252 in a greenhouse, growth chamber, pot, or other controlled environment. This“greenhouse” 253 effect, *β*gh, has a diffuse Normal prior with a mean of zero and a precision of 0*.*01.
3. The observation precision (precision = 1/variance) for the *kth* sample mean, *τk*, is
4. determined based on the within-unit precision, *τY* , and the sample size, *n*, as *τk* = *n* × *τY*

√

1. (since *SE* = *SD/ n*). A common within sample unit precision, *τY* , is assumed in order
2. to accommodate literature values with missing sample sizes or variance estimates. The
3. sample standard error, *sek*, is drawn from a Gamma distribution with parameters 259 informed by the sample size, *n*, and within-site precision, *τY* :

 Gamma() (3)

1. *τY* has a diffuse gamma prior. Unlike the mean and variance parameters, missing values
2. of *n* cannot be estimated and are conservatively set either to 2 (when existence of a
3. variance estimate indicates *n* ≥ 2) or to 1 (if no variance estimate is given). 263 The random and fixed effects and the among study, among treatment, and within-unit 264 precisions are used to evaluate the importance of different sources of uncertainty.
4. The meta-analysis module in PEcAn is fit using JAGS software (version 2.2.0, (Plummer,
5. 2010)) called from within R code that handles data manipulations and meta-analysis
6. model specification in JAGS. JAGS uses standard Markov Chain Monte Carlo (MCMC) 268 methods (Gelman and Rubin, 1992) to approximate the posterior distribution of the
7. terms in the meta-analysis. To overdisperse the *n* MCMC chains, initial values of *β*0 are
8. set to the  quantiles of the prior on *β*0; for the default *n* = 4 chains, this
9. would be the {0*.*2*,*0*.*4*,*0*.*6*,*0*.*8} quantiles. Following Gelman and Shirley (2011), PEcAn
10. discards the first half of each chain, thins each chain to 5000 samples and then combines
11. the chains into a single vector of samples for each term in the meta-analysis model. Trace
12. plots and the Gelman-Rubin convergence diagnostic (Gelman and Rubin, 1992) are used
13. to assess chain convergence. Density plots (Figure 4) are used to visually compare the *β*0
14. chain to data and priors. The significance of the greenhouse effect is evaluated by
15. calculating a two-sided probability that *β*gh 6= 0.
16. When species-level data are unavailable, the posterior distributions are equivalent to the
17. priors.
18. Each term in the meta-analysis is represented as a vector of MCMC samples from the
19. posterior distribution. Statistical summaries of the parameters can easily be calculated
20. from these chains, and chains can also be directly sampled for use in ecosystem model 283 parameterization. When the *β*0 chains are sampled for the ecosystem model ensemble,

284 the meta-analysis posteriors become the model ensemble priors.

## 285 Model Analysis

### 286 Ensemble Analysis

287 Typically, ecosystem models are run for a single model parameterization. For example, 288 the model could be evaluated at the median value of each parameter. However, this

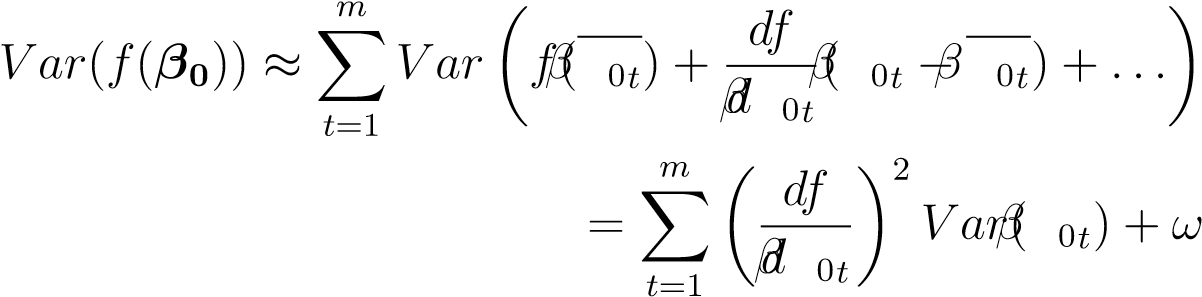
1. approach only provides a point estimate with no accounting for parameter uncertainty.
2. To propagate parameter uncertainty through the ecosystem model, PEcAn uses standard
3. ensemble-based Monte Carlo approaches. An ensemble of model runs is a set (e.g. 500 or
4. 1000) of model runs that are parameterized by sampling from the trait parameter
5. distributions. For each ensemble member, parameter sets are sampled from the full joint
6. parameter distribution of *β***0**, the vector of all model parameters. As a result, the model
7. ensemble approximates the posterior distribution of the ecosystem model output. The
8. model ensemble produces a posterior distribution of the ecosystem model output that can
9. be summarized with standard statistics (e.g. mean, standard error, and credible interval).

### 298 Sensitivity Analysis

1. Sensitivity analyses are used to understand how much a change in a model parameter
2. affects model output; sensitivity is the derivative, *df/dβ*0*t*, of the model (*f*) with respect
3. to the estimate of *β*0 for trait *t*. PEcAn approximates the sensitivities based on univariate
4. perturbations of model parameters. These approximations are necessary because
5. analytical solutions for sensitivity are not tractable for most ecosystem models, and
6. PEcAn is designed to be flexible and applicable to any such model. One disadvantage of
7. traditional perturbation-based sensitivity analyses is that the perturbations are usually
8. arbitrary, for example varying each parameter by a fixed percentage of its value
9. (Larocque et al., 2008) rather than over a meaningful range of the parameter. These
10. traditional approaches make interpretation of sensitivities difficult because they fail to
11. acknowledge the distribution or uncertainty of each parameter. In this regard, PEcAn 310 offers a distinct advantage over traditional sensitivity analyses because parameters are 311 varied based on the meta-analysis posterior parameter distributions.
12. Based on initial exploratory analyses, we found a local perturbation to be inadequate for
13. capturing the responses in most parameters so we instead estimate sensitivities using a
14. global univariate sensitivity analysis. By default, PEcAn evaluates each parameter at the
15. posterior median and at the six posterior quantiles equivalent to ±[1*,*2*,*3]*σ* in the
16. standard normal while holding all other variables constant at their posterior median. The
17. relationship between model output and each model parameter *β*0*t* is then approximated
18. by a natural cubic spline *gt*(*β*0*t*) that interpolates through the evaluation points. The
19. model sensitivity to each parameter is approximated by the derivative of the spline
20. (*dgt/dβ*0*t*) at the parameter mean. In addition to the sensitivity analysis, this set of 321 spline approximations is used in the variance decomposition, in partitioning residual
21. variance, and in evaluating the effect of ensemble size on the estimate of model variance.
22. To facilitate comparisons among the trait sensitivities, despite differences in the units on
23. different traits, we tabulate the coefficient of variation (normalized parameter variance)
24. and the elasticity (sensitivity with terms *df* and *dβ*0*t* standardized by the mean model 326 output and parameter mean respectively).

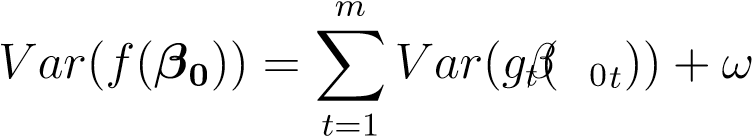
### 327 Variance Decomposition

1. Variance decomposition aims to explain how much each input parameter contributes to
2. uncertainty in model output (Cariboni et al., 2007). Although the present analysis 330 focuses on model parameters, these methods can be extended to address uncertainty in 331 initial conditions or model drivers.
3. The Delta Method uses Taylor series expansion to approximate the probability
4. distribution of a continuous function of random variables (Oehlert, 1992; pp. 240–245 in 334 Casella and Berger, 2001). In this study, the model output *f*(*β***0**) is a function of a vector
5. of the full set of parameters. After approximating the distribution of *f*(*β***0**), it is possible
6. to estimate the variance of the model output. The first step is to derive the Taylor series
7. approximation of the variance of a function (Casella and Berger, 2001, equation 5.5.9 in):

(4)

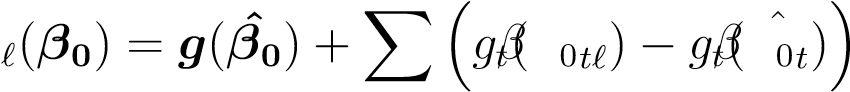
(5)

1. where *m* is the number of parameters in the model, and the error term *ω* accounts for
2. higher order terms in the Taylor series, and *β*0*t* is the estimate of *β*0 from the 340 meta-analysis (equation 2) for each trait, *t*.
3. With this approximation, it is straightforward to estimate the variance contributed by
4. each parameter. The terms in this form of the variance decomposition can be estimated
5. directly from the preceding analyses: *V ar*(*f*(*β***0**)) is the variance of the model ensemble;
6. *V ar*(*β*0*t*) is the posterior variance of trait *β*0*t* from the meta-analysis (equation 2); and
7. *df/dβ*0*t* is the model sensitivity at the parameter mean *β*0*t*. The resulting assertion is
8. that the variance of model output is equal to the sum over the variance of each trait 347 times its sensitivity squared plus a closure term, *ω*.
9. We found that the traditional Taylor polynomial approach to variance decomposition
10. produced a poor closure of the total variance of the model output: for more sensitive
11. parameters, a linear approximation of *f*(*β***0**) provided unrealistic estimates of the
12. sensitivity function that overestimated variance. Increasing the order of the Taylor series
13. expansion actually exacerbated this problem (results not shown). One problem with the
14. polynomial approximation is that, unlike polynomials, most response variables in
15. ecosystems and ecosystem models tend to be asymptotic at both high and low values of a
16. trait. For example, when assessing aboveground biomass there is a lower bound of zero
17. biomass and most parameters become progressively less sensitive, if not genuinely
18. asymptotic, at their upper bound. This asymptotic behavior is poorly approximated by a
19. polynomial because polynomials are unbounded at extreme parameter values. Therefore, 359 we sought a better approximation for the variance decomposition.
20. First, we formulated a more generalized form of the variance decomposition (equation 4):

 (6)

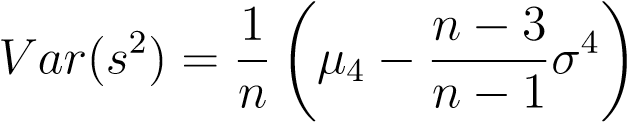
1. The spline *gt*(*β*0*t*) is a statistical emulator of the model response to trait *t* that
2. transforms *β*0*t* from the parameter domain to the model domain. The univariate
3. contribution of each parameter to variance of the model output is thus *V ar*(*gt*(*β*0*t*)).
4. Equation 6) only requires *β*0*t* from the preceding meta-analysis, *gt*(*β*0*t*) from the 365 sensitivity analysis, and *V ar*(*f*(*β***0**)) from the ensemble analysis.
5. The final term, *ω*, is the closure between the right hand side and the left hand side of the
6. variance decomposition; *ω* represents the effects of the higher order terms in the Taylor
7. approximation and the covariance terms between parameters. This closure term is
8. intended to represent parameter interactions that are excluded from the univariate
9. variance decomposition (equation 6). Negative trade-offs among physiological traits
10. would result in *ω* less than zero. However, our estimate of *ω* also includes errors
11. associated with using finite sample sizes, the spline approximation in each *gt*(*β*0*t*), and
12. biological range restrictions on model output that are not reflected in the variance 374 decomposition (equation 6).
13. One approach to partition the error in the closure term is to use the univariate spline
14. functions from the sensitivity analysis to estimate what the model output would be for
15. each of the parameter sets used in the model ensemble; we call this estimate the “spline
16. ensemble”:

*m*

*g* (7)

*t*=1

1. In this equation, *g`*(*β***0**) is the spline estimate of the model output for the *`*th ensemble 380 member and *β*ˆ0*t* is the posterior median parameter value.
2. Although the individual splines may respect range restrictions on output variables (e.g.
3. biomass values cannot fall below zero), combinations of the splines evaluated for a set of
4. unfavorable traits can fall outside these ranges. For parameter sets that give a
5. biologically implausible estimate of negative biomass (*g`*(*β***0**) *<* 0), the estimate is set to
6. zero. The only difference between the variance of the spline ensemble (equation 7) and
7. the variance decomposition (equation 6) is that range restrictions are not corrected for in
8. the variance decomposition. Therefore, the spline ensemble allows us to estimate the
9. effect of using combinations of spline estimates that do not respect the zero bound on
10. biomass in the variance decomposition. The difference between the model ensemble and
11. the spline ensemble provides an estimate of parameter interactions in the model because 391 the spline ensemble does not include the parameter interactions that exist in the model.
12. The precision of the estimate of model ensemble variance is affected by the number of
13. runs in the ensemble. When the computational expense of the model itself limits the
14. ensemble size, there can be significant uncertainty in the estimate of ensemble variance.
15. The uncertainty in a sample variance is estimated as

 (8)

1. (Mood et al., 1974, , p 239) where *µ*4 is the fourth central moment. *V ar*(*s*2) scales
2. inversely with sample size. The effect of the limited model ensemble size on uncertainty
3. in the estimate of ensemble variance is measured in two ways. The first way is to
4. calculate *V ar*(*s*2) for the model ensemble (*n* = 500). The second way is to compare
5. *V ar*(*s*2) of the spline ensemble with 500 and 10*,*000 runs. The 95%CI for *s*2 is calculated
6. as *s*2 ± 1*.*96*ss*2 where *ss*2 = p*V ar*(*s*2).
7. The errors introduced from using a spline approximation of the model response can not
8. be estimated based on the existing output, but it is small in comparison to the other 404 effects given the range restrictions imposed by the spline interpolation.
9. The results of a model ensemble are posterior estimates of aboveground biomass.
10. However, we also distinguish between ensembles depending on the nature of model
11. parameters. First, we ran a “prior model ensemble” using an ensemble of parameter sets
12. drawn from prior distributions, and then a “posterior model ensemble” drawn from 409 meta-analysis posteriors.

# 410 Application: Switchgrass Monoculture

1. We demonstrate the application of PEcAn to estimate the aboveground yield of an
2. experimental switchgrass (*Panicum virgatum*) monoculture. The first step to applying
3. PEcAn was to construct an appropriate set of priors based on data syntheses and expert
4. knowledge. These priors were conservative estimates of the plant trait parameters based
5. on information other than species level data. Next, switchgrass trait data from both
6. previous studies and field measurements were summarized using meta-analysis to
7. constrain the prior parameter estimates. The Ecosystem Demography model version 2.1, 418 (Medvigy et al., 2009; Moorcroft et al., 2001) was used to simulate plant growth.
8. The model ensemble and sensitivity analysis were performed using both the prior and
9. posterior parameter estimates. By comparing the prior model ensemble to the posterior 421 model ensemble, we are able to evaluate the ability of species level data to reduce model
10. uncertainty.
11. To evaluate model performance, we compare the ensemble estimates of aboveground
12. biomass with observed yields (Heaton et al., 2008; Wang et al., 2010, Figure 5).

## 425 Site

1. Switchgrass (*Panicum virgatum*) is a perennial grass native to North America that has
2. received attention as a potential cellulosic biofuel crop (McLaughlin and Kszos, 2005;
3. Wang et al., 2010). We modeled the aboveground biomass production of a switchgrass
4. monoculture and compared model estimates to a monoculture planted in 2002 at the
5. University of Illinois Agricultural Research and Education Center in Urbana, IL
6. (40.09 N, 88.2 W). The climate at this site is characterized by hot, humid summers and
7. cold winters with a 50 year (1959-2009) mean annual temperature of 11 ◦C and mean
8. annual precipitation of 1000 mm yr−1 (Angel, 2010). Meteorological data used to drive
9. the model were downloaded from the North American Regional Reanalysis (Mesinger 435 et al., 2006). Soil is a silt loam from the Drummer-Flanagan soil series; texture data was
10. obtained through the USDA NRCS web soil survey website (websoilsurvey.nrcs.usda.gov).
11. The yield and other aspects of this ecosystem have previously been reported (Heaton 438 et al., 2008).

## 439 Ecosystem Demography Model

1. We used the Ecosystem Demography Model, version 2 to model the productivity and soil
2. carbon pools in this switchgrass agro-ecosystem. ED2 is a terrestrial biosphere model
3. that couples age- and stage-structured plant community dynamics with ecophysiological
4. and biogeochemical models. The biophysical land-surface model in ED2 allows plant
5. uptake and growth to respond dynamically to changes in weather and soil hydrology
6. (Medvigy et al., 2009). ED2 has the ability to link short-term, physiological responses to
7. environmental conditions with realistic, long-term successional changes in ecosystem
8. structure and composition (Moorcroft et al., 2001). While other models have both
9. succession and physiology, ED2 also has explicit spatial scaling, a sub-daily time-step, 449 and the ability to couple with to a land surface model (Dietze and Latimer, 2011).
10. ED2 incorporates a mechanistic description of plant growth that accounts for the fast
11. temporal responses of plants to changes in environmental conditions. In this study, we
12. vary fifteen model parameters based on observable plant traits that control carbon 453 uptake, carbon allocation, turnover, and reproduction (Table 1, Figures 2, 4).
13. ED2 calculates photosynthetic rates using the enzyme kinetic model developed for C3
14. plants (Farquhar and Sharkey, 1982; Ball et al., 1987) and the modifications for C4
15. (Collatz et al., 1992). Vc,max sets the upper bound on the rate of Rubisco-limited
16. photosynthesis, while light limited photosynthesis is constrained by the quantum
17. efficiency parameter, and a threshold parameter controls the minimum temperature at
18. which photosynthesis will occur. Stomatal conductance is calculated using the Leuning 460 variant of the Ball-Berry model (Leuning, 1995) and is controlled by the stomatal slope 461 parameter. Leaf boundary layer conductance depends on the leaf width parameter.
19. Together, stomatal conductance and leaf boundary layer conductance affect carbon and
20. moisture fluxes and the leaf energy balance. Specific leaf area (SLA) determines the 464 amount of leaf area produced per unit leaf biomass investment.
21. In addition to photosynthesis, ED2 also accounts for carbon allocation to growth,
22. respiration, and for the turnover rate of carbon pools. These parameters include: one to
23. partition between leaf and fine root growth; one for allocation to reproduction; two
24. respiration parameters associated with growth respiration and root maintenance 469 respiration; and two parameters to control the rates of leaf and root turnover. 470 Finally, three demographic parameters control seed dispersal, seedling mortality, and

471 adult mortality due to carbon limitation (Table 1).

## 472 Priors

### 473 Priors from data

1. Priors were estimated by finding the best fit distribution to raw data sets include SLA
2. and leaf turnover rate from the GLOPNET database (Wright et al. (2004), *n* = 125*,*40
3. respectively), root turnover rate (Gill and Jackson (2000), *n* = 66), and quantum yield
4. (Skillman (2008), *n* = 56). Candidate distributions for these priors were Gamma,
5. Weibull, log-Normal, and F because each of these traits is bound at zero. In all cases we
6. are interested in using the full distribution of across-species data as our prior constraint
7. on what one individual species is capable of doing, as opposed to using the estimate of 481 the mean of this distribution as our prior.
8. Quantum yield data represent a survey of published values of quantum yield in C4
9. monocots (Skillman, 2008); original data were provided by the author and restricted to 484 measurements made under photorespiratory conditions (ambient CO2 and O2) (J. 485 Skillman, personal communication). Given the narrow range of data (CV = 11%), the 486 normal distribution was also considered but was not the best fit.

### 487 Priors from meta-analysis

1. We used meta-analysis to calculate a prior from data when summary statistics and
2. sample sizes were available. The meta-analysis model used to calculate prior distributions
3. is similar to the one used by PEcAn to summarize species-level data (equation 2), with
4. three differences. First, there are no site, treatment, or greenhouse effects. Second, data
5. from multiple species were used. Third, we generated a posterior predictive distribution
6. to predict the distribution of trait values for an unobserved C4 plant species, unlike the 494 species-level meta-analysis, which estimated the global mean parameter value. Thus, the
7. model included plant functional type (PFT) as a random effect:

Θspecies = *β*0 + *β*PFT (9)

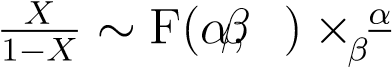
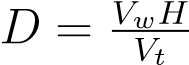
1. Stomatal slope is the empirical slope coefficient in the (Leuning, 1995) model of stomatal
2. conductance. Estimates of this parameter are sparse, so we collected new data for this
3. study. Stomatal slope was estimated using measurements of four leaves from each of five
4. field-grown energy crop species during the 2010 growing season (Appendix A). The five
5. species included two C4 grasses: Miscanthus (*Miscanthus x giganteus*) and Switchgrass
6. (*P virgatum*) planted in 2008 and three deciduous tree species: Red Maple
7. (*Acer rubrum*), Eastern Cottonwood (*Populus deltoides*, and Sherburne Willow
8. *Salix x Sherburne*) planted in 2010 as 2 year old saplings. All plants were grown at the
9. Energy Biosciences Institute Energy Farm (40*o*10’N, 88*o*03”W). We used the data from
10. the three tree species and Miscanthus to calculate the posterior predictive distribution of
11. an unobserved C4 grass species, and used this distribution as the prior estimate for 507 Switchgrass stomatal slope.
12. Maximal carboxylation rate (Vcmax) data consists of ninety-four C3 species (Wullschleger,
13. 1993) plus three C4 species (Kubien and Sage, 2004; Massad et al., 2007; Wang et al.,
14. 2011). To express Vcmax at a common temperature of 25*oC* for all species, we applied an 511 Arrhenius temperature correction (Appendix C). The Wullschleger (1993) data set

512 included a 95% CI and an asymptotic error calculated by the SAS nlin proceedure (Stan 513 Wullschleger, personal communication). We used the asymptotic error as an estimate of

SE2. Plant

1. SE, the 95% CI to estimate SD SD =, and then estimated *n* as ˆ*n* = SD
2. species were classified into five functional types: C3 grass, C4 grass, forb, woody
3. non-gymnosperm, and gymnosperm based on species records in the USDA PLANTS 517 Database (USDA and NRCS, 2011). Ambiguous species (those with both forb and woody
4. growth forms, *n* = 15) were excluded.
5. Leaf width data represent 18 grass species grown in a common garden greenhouse
6. experiment (Oyarzabal et al., 2008). *P. virgatum* was among the 18 species, and was
7. excluded from the prior estimation but used as raw data in the meta-analysis. The
8. remaining seventeen species were divided into C3 and C4 functional types. Relative to
9. the small sample of C4 grasses, switchgrass leaf width was an outlier; inflating the
10. variance four-fold reduced the prior information content to account for this descrepency.
11. Root respiration rate values were measured on thirty-six plants representing five
12. functional types, including six C4 grass species (Tjoelker et al., 2005). As before, *P.*
13. *virgatum* data was excluded from the prior estimation and used as raw data in the 528 species-level meta-analysis.

### 529 Priors from limited data and expert knowledge

1. When available data were limited to a few observations, these were used to identify a
2. central tendency such as the mean, median, or mode, while expert knowledge was used to
3. estimate the range of a confidence interval. An optimization approach was used to fit a 533 probability distribution to this combination of data and expert constraint.
4. In order to estimate the fine root to leaf ratio for grasses, we assume fine roots account
5. for all belowground biomass (Jackson et al., 1997) and that leaves account for all above
6. ground biomass. Roots account for approximately 2*/*3 of total biomass across temperate
7. grassland biomes (Saugier et al., 2001, Table 23.1), so we constrained a beta prior on the
8. root fraction to have a mean of 2*/*3 by setting *α* = *β/*2 since the mean of a beta is
9. defined as . To convert from proportion to ratio, we used the identity: if
10. *X* ∼ Beta() then . We constrained the 95% ],
11. equivalent to a fine root to leaf ratio with a mean fixed at two and a 95%CI = [1*/*2*,*10].
12. We simulated the distribution of the fine root to leaf ratio by drawing 10000 samples
13. from a F(2*α,α*) distribution and multiplying these samples by two.
14. Seed dispersal in ED2 represents the proportion of seed dispersed outside of a 7.5m radius
15. plot, which we approximate as a beta distribution. Although no direct measurements of
16. seed dispersal were available, it was straightforward to parametrize a ballistic model of
17. seed dispersal (Ernst et al. (1992), from Creemer 1977): . This model relates
18. dispersal distance *D* to terminal velocity *Vt*, wind speed *Vw*, and seed height *H*. Although
19. more sophisticated treatments of dispersal exist and are important for estimating low
20. probability long distance dispersal events (Clark et al., 1999; Thompson and Katul, 551 2008), the Ernst et al. (1992) model is sufficient for relatively short dispersal distances 552 required in the present context.
21. Values of terminal velocity, *Vt*, of grass seeds were taken from two references, (Ernst
22. et al., 1992; Jongejans and Schippers, 1999) and these data were best described as 555 *Vt* ∼ Gamma(2*.*93*,*1*.*61).

556 Next the heights from which the seeds are dropped was estimated from calibrated 557 photographs of reproductively mature switchgrass at a field site in Urbana, IL:

1. *H* ∼ N(2*,*0*.*5). Finally, wind speed observed at the site were fit to a Weibull distribution
2. (Justus et al., 1978). *Vw* ∼ Weibull(2*.*4*,*0*.*712) (Marcelo Zeri, unpublished wind and
3. height data). Given these distributions of *Vw*, *H*, and *Vt*, sets of 100 dispersal distances
4. were simulated 10000 times to calculate the fraction of seeds in each simulation dispersed
5. beyond 7.5m,

### 563 Priors informed by expert knowledge

564 When no data were available, expert knowledge was used to estimate the central 565 tendency and confidence interval for a trait parameter. Again, optimization was used to 566 fit a probability distribution to these constraints.

1. The minimum temperature of photosynthesis for C4 grasses was given a prior based on
2. expert knowledge with a mean of 10*o*C and a 95%CI = [8*,*12]*o*C that fits a normal
3. (*µ* = 10, *σ* = 1*.*02) distribution (Don Ort, UIUC, personal communication, 2010).
4. The growth respiration factor is the proportion of daily carbon gain lost to growth
5. respiration. Because it is a proportion, the beta distribution was fit with a mean set 572 equal to the ED2 default parameter value, 0*.*33 and a 95%CI = [0*.*05*,*0*.*60], conservatively 573 based on the range of construction costs reviewed by Amthor (2000).

574 Seedling mortality factor represents the proportion of carbon allocated to reproduction 575 that goes directly to the litter pool. Given the default ED2 parameter is 0.95, we stated a 576 beta prior with a median at 0*.*95, and a 95%CI = [2*/*3*,*1].

1. The mortality factor in ED2 is the rate parameter in the negative exponential
2. relationship between carbon balance and mortality (Medvigy et al., 2009). The default 579 parameter for all plant functional types (PFT’s) in ED2 is 20, and our weakly informed 580 gamma prior sets this as the median and gives a 95%CI = [5*,*80].

581 Reproductive allocation represents the proportion of carbon in the storage pool allocated 582 to reproduction. This parameter is a proportion and has a default value of 0*.*33 in ED.

1. The Beta(2*,*4) distribution has a mean of 1*/*3 and a 95%CI = [0*.*05*,*0*.*72] representing
2. relatively high uncertainty. This distribution implies that the plant may allocate any 585 fraction of the carbon pool to reproduction between but not equal to 0 and 1 and has an 586 80% probability that less than half of the carbon pool will be allocated to reproduction.

## 587 Switchgrass Trait Meta-analysis

1. Switchgrass trait data used to constrain model parameters are stored in the Biofuel
2. Ecophysiological Trait and Yield database (BETYdb, www.betydb.org), a database
3. designed to support research on biofuel crops. BETYdb includes both previously 591 published and primary data (Appendix A). Prior to entry in the database, data were 592 converted to standard units chosen for each variable (Table 1).
4. Trait data available for *Panicum virgatum* include Vcmax, SLA, leaf width, fine root to
5. leaf ratio, root respiration, stomatal slope, and root turnover rate (Figure 4, Table 2).
6. Methods used to collect these data and site descriptions are available in the source 596 references (Appendix A). Each meta-analysis was run with four 50,000 step MCMC 597 chains.

## 598 Model Analysis

1. We executed a ten-year, 500 run ensemble of ED2 using parameter values drawn from the
2. prior or posterior parameter distributions. The model was run for the years 1995-2006 to
3. simulate the field trials conducted by Heaton et al. (2008). The model output of interest
4. was the December mean aboveground biomass (AGB) during the years 2004–2006,
5. simulating the yields of the mature stand (Heaton et al., 2008). The ensemble estimate 604 was also compared to the larger set of all reported switchgrass yield data (Wang et al.,
6. 2010).
7. Runs resulting in yields less than 2 Mg/ha were considered non-viable parameter
8. combinations. To test if prior and posterior parameter sets resulted in different fractions
9. of non-viable runs, we estimated the posterior probability of a non-viable run as a
10. binomial posterior from a beta-binomial model with a flat (Beta(1*,*1)) prior. Then, we 610 calculated the two-tailed probability that the difference between these binomial posteriors 611 was 6= 0.

# 612 Results

## 613 Trait Meta-analysis

1. Switchgrass data were collected from the literature and field for seven of the model
2. parameters: specific leaf area (SLA) (*n* = 24), leaf width (39), Vcmax (4), fine root to leaf
3. allocation ratio (4), stomatal slope (4), root respiration rate (1), and root turnover rate
4. (1). Table 2 summarizes the meta-analysis for each of these parameters, including the 618 posterior mean and 95% CI of the global mean, the fixed greenhouse effect, and each of 619 the variance components (reported as standard deviations).
5. SLA and leaf width data were from from multiple sites, but the meta-analysis provided
6. no evidence for among site variability in excess of within site variability (*σY* and *σ*site,
7. respectively, in Table 2). For the remaining traits, there was insufficient spatial sampling
8. to assess site to site variability. Greenhouse growing conditions had a positive effect on 624 both SLA (*P* = 0*.*027), and leaf width (*P* = 0*.*052).
9. Figure 4 compares parameters before and after incorporating data in the meta-analysis.
10. A reduction in parameter uncertainty is seen as the reduction in the spread of the
11. posterior (black) compared to the prior (grey) parameter distributions. The influence of
12. the prior information on the posterior distribution increased when the prior was more
13. constrained or when less data were available for use in the meta-analysis. For example,
14. data substantially constrained the uncertainty in the Vcmax and SLA posteriors relative to
15. the priors. By contrast, there was little effect of additional data on the parameter 632 estimates for fine root to leaf allocation and root respiration rate; these parameters had 633 relatively well constrained priors and limited species-specific data.

## 634 Model Analysis

### 635 Ensemble

1. Within the model ensemble analysis (Figure 5), both the prior and posterior
2. parameterizations produced yield estimates that were consistent with yields observed at
3. the Urbana site for which the model was run (Heaton et al., 2008) and with 1902
4. previously reported yields of switchgrass (Wang et al., 2010). In both the prior and
5. posterior ensembles, the predicted aboveground biomass was clearly bimodal. These two 641 modes had little overlap and a distinct break at two Mg/ha. We infered that the first
6. peak represents non-viable plants generated by unrealistic parameter sets so plants with
7. aboveground biomass less than two Mg/ha were considered “non-viable”. When
8. summarizing the model output, we consider viable and non-viable ensemble members
9. separately; all runs are considered in the senstivity analysis and variance decomposition. 646 A greater percentage of runs in the prior ensemble fell below this threshold (53*.*4 vs 36*.*6,
10. *P* ' 0).
11. Compared to the prior ensemble prediction, the data-constrained posterior runs had lower
12. median yields and a more constrained 95% credible interval (16*.*5[7*.*2*,*37] Mg/ha vs
13. 25[7*.*7*,*56] Mg/ha). This reflects the substantial shrinkage of the posterior relative to the
14. prior SD estimates of model output uncertainty (from *σ* = 19*.*7 to *σ* = 11*.*9). In
15. particular, the upper tail of the modeled yield was reduced toward the observed yields.
16. Despite the reduction in ensemble uncertainty, the ensemble posterior yield was still
17. relatively imprecise and had much greater uncertainty than the field trial (Heaton et al., 655 2008, *σ* = 4*.*1) or the meta-analysis of all observations (Wang et al., 2010, (*σ* = 5*.*4)).

656 The spline ensemble viable plants had a median 18*.*8[2*.*9*,*48] and *σ* = 13.

### 657 Sensitivity Analysis

1. Sensitivity analysis demonstrated that traits varied in their effect on on aboveground
2. biomass (Figure 6), and many of these relationships are clearly non-polynomial. For
3. example, parameters associated with photosynthesis and carbon allocation - including
4. Vcmax, SLA, growth respiration, and root allocation - were particularly sensitive. For
5. particularly sensitive parameters, the sensitivity functions had coverage of unrealistic
6. yields greater than 30 Mg/ha. Constraining SLA and Vcmax parameters with data 664 resulted in a more realistic range of yields. On the other hand, aboveground biomass was

665 largely insensitive to leaf width, seed dispersal, and mortality rate.

### 666 Variance Decomposition

1. The variance decomposition showed that data-constrained parameters substantially
2. reduced their contribution to overall model variance (Figure 7). Prior to including
3. species-specific trait data, SLA contributed the most to model uncertainty, followed by
4. growth respiration, fine root allocation, Vcmax, seedling mortality, and stomatal slope
5. (right panel, grey bars Figure 7). Incorporating species level data substantially reduced
6. the contributions of SLA, Vcmax, seedling mortality, and stomatal slope to model
7. uncertainty. For example, SLA fell from first to fourth and stomatal slope fell from sixth
8. to fourteenth in rank contribution to ensemble variance. Although the addition of data
9. reduced parameter uncertainty for fine root to leaf allocation, aboveground biomass was
10. more sensitive to this parameter at the posterior median. These changes cancelled each
11. other out, and as a result the contribution of the fine root allocation parameter to 678 ensemble variance remained constant.

679 The variance of the ensemble was less than the variance calculated in the variance 680 decomposition, and this difference is the closure term, *ω*. The closure term accounted for 681 approximately 28*.*455981108897% of the variance decomposition estimate (Table 3b).

682 There was no effect of increasing the sample size from 500 to 10000 on the variance 683 estimates (Table 3a).

# 684 Discussion

## 685 Switchgrass Trait Meta-analysis

1. When species-level data were available, the meta-analysis constrained estimates of the
2. trait mean parameter (Figure 4) and provided insight into the sources of parameter
3. uncertainty (Table 2). In the context of constraining model parameters, we were
4. interested in accounting for but not directly investigating the specific effects of site,
5. treatment, or greenhouse effects. However, we can use the meta analysis results to
6. identify sources and scales of parameter variability. This insight into parameter 692 variability helps inform future sampling designs, development of the ecosystem model, 693 and improvement of methods used to parametrize the ecosystem model.
7. Where data from multiple sites were available, we could evaluate the relative importance
8. of within versus among-site variance for the range of sites represented in the data 696 (Table 2). Recent studies demonstrate important effects of intraspecific trait variability 697 on ecosystem functioning (Breza et al., 2012; Albert et al., 2011; Violle et al., 2012).
9. Therefore, for traits that do exhibit greater variability across than within sites, explicit
10. inclusion of spatial, environmental, and even genetic information into the meta-analytical
11. model would be justified. This approach would enable the estimation of site-specific 701 parameters for use in the ecosystem model and will be investigated in future development 702 of the meta-analysis module.
12. For the other parameters (Vcmax, fine root allocation, root respiration rate, and root
13. turnover rate) data came from one site, so it is not possible to estimate the across-site
14. variability. For these traits, obtaining data from additional sites would better constrain
15. both the global mean and the across-site variance. This additional data collection is
16. particularly justified for traits that contribute most to the uncertainty in the model 708 ensemble.

## 709 Model Ensemble

1. Despite the large reduction in model uncertainty from the prior to the posterior model
2. ensemble, the uncertainty in projected yield is substantial (Figure 5) and further
3. constraint would increase the utility of this model output. However, the explicit
4. accounting of parameter uncertainty is an important first step toward producing more
5. informative model output. If model parameters had been treated as fixed constants, we
6. would have no estimate of model uncertainty; without an estimate of uncertainty, the
7. similarity between the modeled (16*.*5 Mg/ha) and observed (12*.*0 Mg/ha) median yields
8. would be difficult to interpret; a naive interpretation could create false confidence in the
9. model. Including the non-viable plants would have made the model mean more accurate
10. (Figure 5), but the 90%CI would have been less accurate, containing the possibility that
11. switchgrass would not grow in Champaign County, Illinois, even though extensive 721 research (Heaton et al., 2008; VanLoocke et al., 2012, personal observation) demonstrates 722 that it does grow very well in this area.
12. The reduction in median modeled yield in the posterior relative to the prior model
13. ensemble 5 is consistent with the reduced probability of high SLA and Vcmax values in the
14. posterior relative to the prior distributions. As expected, the use of switchgrass trait data
15. to inform model parameters succeeded in both reducing total uncertainty and bringing
16. modeled yield in line with observations of switchgrass yields both at this site (Heaton
17. et al., 2008) and globally (Wang et al., 2010). Reducing uncertainty in model outputs, in 729 this case yield, is key to increasing the value of ecological forecasts (Clark et al., 2001).
18. While reducing uncertainty does not necessarily increase model accuracy, an estimate of
19. model uncertainty is a first step toward generating meaningful statistical inference from
20. the model itself. Without an estimate of model uncertainty, it is not possible to make
21. such a basic inference as the probability that the model predictions overlap with observed
22. data; this limits the capacity of models to inform research and applied problems (Clark
23. et al., 2001). Instead, comparisons of ecosystem models with observations have focused
24. on differences and correlations between model output and data (Bellocchi et al., 2010;
25. Schwalm et al., 2010; Dietze et al., 2011) without providing a confidence interval around
26. the model output itself. The ability to identify, with confidence, the conditions under
27. which a model produces valid output helps determine appropriate applications of the
28. model and it helps to identify targets for further model development (Williams et al.,
29. 2009). While parameter uncertainty is clearly just one of many sources of uncertainty in
30. models (McMahon et al., 2009), and constraining model parameters by no means
31. guarantees that a model will match reality, is difficult to assess the accuracy of a model if
32. it has low precision. In deterministic models, such as most ecosystem models, parameter 745 uncertainty is a major driver of the precision of a model.
33. In this study, we can state with 90% Confidence that the mean Switchgrass yield during
34. the Heaton et al. (2008) study should have been between 7.2 and 37, and if we had made
35. this prediction in advance, we could have said that we were correct because the mean did
36. fall within this range. We can also see that the model uncertainty contains the 90% CI
37. for observed switchgrass yields globally (Wang et al., 2010), even though we know that
38. important drivers of variability in the global meta-analysis (e.g., climate, soil) are
39. different from the source of uncertainty in our model predictions (e.g., parameters). The
40. model ensemble left open the possibility that the yields could have been much more or
41. much less than was actually observed, and we conclude that much of this variability could
42. be constrained with additional trait level data or data assimilation. Wang et al 201x (in
43. review, Ecological Applications #12-0854) provides an example of combining the PEcAn
44. meta-analysis and variance decomposition with data assimilation of biomass to constrain
45. uncertainty in parameter estimates and improve the accuracy and precision of model
46. output. Once the model can make more precise predictions, it will be possible to begin 760 investigation of other sources of uncertainty, such as model structure and state variables
47. (e.g. climate, soil).
48. Although the present analysis focuses on modeled aboveground biomass, PEcAn can 763 analyze any model output, including pools and fluxes of water, energy, and carbon.
49. Indeed, PEcAn’s approach to the synthesis of data and mechanistic models is
50. independent of the system being modeled, and thus has potential applications far beyond 766 the scope of its current development to support ecosystem modeling.

## 767 Variance Decomposition

1. Variance decomposition quantified the contribution of each parameter to model
2. uncertainty, helping to identify a subset of parameters for further constraint. SLA, Vcmax,
3. fine root to leaf ratio, and leaf turnover rate dominated uncertainty in yield prior to
4. incorporating species level data. Therefore, SLA, which can be measured quickly and at
5. low cost, would make a good first target for reducing uncertainty when a new species is
6. evaluated. SLA also correlates strongly with other important model parameters, such as
7. photosynthetic rate, leaf lifespan, and nitrogen content (Wright et al., 2004). The ranking
8. of parameters based on variance contribution depends on the response variable of choice
9. (in this case, aboveground biomass) as well as the conditions of the run (duration, soil,
10. climate), and the species or PFT being evaluated. In general, for a given species and
11. model output, overall patterns of parameter importance are consistent across a broad 779 range of climates (Wang et al., 201x, in review, Ecological Applications #12-0854).
12. Variance decomposition (equation 6) demonstrates that it is not parameter uncertainty or
13. model sensitivity alone, but the combination of the two, that determines the importance
14. of a variable. For example, despite the high uncertainty in seed dispersal, switchgrass
15. yield is insensitive to this parameter (Figures 6, 7), therefore a better understanding of
16. switchgrass seed dispersal would not reduce model uncertainty. By contrast, although
17. uncertainty in the growth respiration is not particularly large, switchgrass yield was very
18. sensitive to growth respiration, and for this reason growth respiration is the greatest
19. contributor to model uncertainty. In addition, although no seedling mortality data were
20. available, model sensitivity to this parameter was much lower in the posterior compared
21. to prior runs. Using the sensitivity analysis or parameter uncertainties alone would thus
22. lead to incorrect conclusions about what parameters are most important and an
23. inefficient approach to reducing predictive uncertainties.
24. This analysis only represents the first step toward more comprehensive accounting of
25. known sources of uncertainty. The next step in reducing uncertainty would be to use the
26. results of the variance decomposition to target the most influential model parameters for
27. further constraint through data collection. We have demonstrated how the use of 796 species-level data to constrain parameter uncertainty reduced ensemble variance, resulting 797 in a new set of targets for additional field observations and refined literature surveys.
28. Traits that are difficult to measure, such as belowground carbon cycling, can be indirectly
29. constrained with ecosystem-level observations using data assimilation (Luo et al., 2009,
30. 2011). Integrating data assimilation into PEcAn will allow ecosystem-level observations
31. to further constrain parameters for which trait level observations are difficult or
32. impossible to obtain. To date most Bayesian data assimilation approaches applied by
33. ecologists have employed flat, uninformative priors (assigning equal probability to values
34. over many orders of magnitude) , which has lead to the problems of parameter
35. identifiably and the criticism that model parameters are allowed to take on biologically
36. unrealistic values. The use of the meta-analysis posteriors as priors in the data
37. assimilation step ensures that any parameter estimates are consistent with what is known
38. about different plant traits. In this way Bayesian methods are, in effect, updating the
39. literature-derived estimates with new data and providing a coherent and rigorous 810 framework for combining multiple different types of data.
40. In addition, by effectively restricting parameter space based on observed values, the use
41. of informed priors in data assimilation reduces problems of equifinality and identifiability.
42. This is consistent with the argument by Beven and Freer (2001) that only the feasible 814 parameter range should be sampled.
43. To a first order the spline interpolations of the univariate relationships between
44. parameters and aboveground biomass (Figure 6) provide a good estimate of the total
45. model variance. The closure term (Table 3b) accounted for approximately 5*.*2Mg ha−1 or
46. 28% of the variance decomposition estimate (18*.*1 Mg ha−1, Table 3a), suggesting that
47. while parameter interactions are important, univariate parameter uncertainty drives
48. overall model variance. One key concern of parameter interactions is that the
49. combination of the variance decomposition terms would result in the prediction of
50. negative yields, which is clearly biologically impossible. By comparing the spline
51. ensemble, where this term is truncated, to the spline-based variance decomposition we 824 can conclude that this truncation effect accounts for 4*.*1 or 80% of the closure term in the
52. variance decomposition.
53. By contrast, evaluating the spline ensemble for different ensemble sizes shows that
54. ensemble size had negligible effect on the mean variance estimate although it does
55. improve the precision of this estimate (Table 3a). Finally, the difference between the
56. model and spline ensembles (Table 3b) suggests that the absence of parameter
57. interactions in the variance decomposition account for the remaining 20% of the closure
58. term (*<* 6% of the variance decomposition calculation), which could be improved by a
59. multivariate meta-analysis and sensitivity analysis, both of which are planned for future 833 development of PEcAn. Overall, the closure term is relatively well constrained even when 834 the parameter interactions are assumed to be linear.

## 835 Model-field work feedback

1. Variance decomposition can be used to inform data collection by identifying candidate
2. parameters for further refinement based on their contributions to model variance. Recall
3. that this variance contribution is a function of parameter sensitivity and the parameters’
4. probability density (equation 6, Figure 7). Sensitivity is a function of the model and so
5. there is no direct way to reduce sensitivity. However, because *V ar*(*f*) ∝ *V ar*(*β*0), it is
6. possible to reduce the model uncertainty by reducing parameter variances.
7. Through simple power analyses one can explicitly estimate the relationship between an
8. increase in sample size and the reduction in posterior variance. Not only can we calculate 844 the reduction in parameter uncertainty that would be expected for a given sample size,
9. but using equation 6 we can also express this in terms of reductions in the variance of the
10. model output. This then allows us to directly compare the value of different data types in 847 a common currency.
11. Quantitatively evaluating the relationship between data and model uncertainty provides a
12. path of communication between field research and modeling, opening the door for a new
13. framework in which modeling and field work can be mutually informative. Given the
14. current data and model uncertainties, it is possible to identify effective data acquisition
15. strategies based on this analysis. For example, data could be ranked by the ratio of
16. reduction in model uncertainty to the cost of acquiring each sample in terms of dollars
17. and/or man hours. In this way, data collection could be optimized in terms of the 855 efficiency at which new information is gained.
18. These approaches close the model-data loop by enabling models to inform data collection,
19. and data to inform models. Such a shift has the potential to put field ecologists and
20. modelers in closer connection. It also gives us the tools to answer the long standing
21. question among many field ecologists about what exactly modelers need to know. Indeed,
22. this shift highlights a need for greater accessibility to models by the general research
23. community so that field ecologists can drive this loop directly. This is exactly the 862 objective of PEcAn – to encapsulate these tasks in a way that makes the analysis of 863 models transparent, repeatable, and accessible.
24. In addition to informing sample size, the parameter meta-analysis can inform
25. experimental design by providing valuable information on the scales of variability. For
26. example, when data from multiple sites is available, the meta-analysis partitions among
27. site and within site variance. This information can be used to construct optimal sampling 868 designs which balance intensive vs extensive sampling, and may help identify 869 environmental covariates that should be measured in order to explain parameter
28. variability.
29. Based on our switchgrass example, variance partitioning also highlights broad data needs
30. and the discrepancy between the relative ease of parameterizing aboveground processes
31. compared to below ground processes. Indeed, one of the greatest challenges in ecosystem
32. ecology is the ability to constrain below ground processes such as root allocation,
33. respiration, and turnover. These parameters are uncertain precisely because measurement
34. is difficult, often indirect, and data may reflect the diverse methods used to indirectly
35. estimate a pool or flux. Many parameters in the ED2 model correspond to processes that
36. are not directly observable. For example, the root respiration parameter in ED2 is not
37. total root respiration but just maintenance respiration, while measurements typically can
38. not separate growth, maintenance, and rhizosphere respiration. Whole-plant growth
39. respiration, which is currently the most important model parameter, is also difficult to
40. estimate directly from measurements (Amthor, 2000). In this case, data assimilation is
41. likely the most efficient route to constrain this parameter; data assimilation would 884 effectively use mass balance of ecosystem carbon exchange to estimate this respiration 885 parameter once other parameters are more directly constrained by data.

## 886 Future Directions

1. The analyses presented here represent the first phase in the development of the PEcAn
2. project. In the near term we intend to expand the existing analyses to include a
3. multivariate meta-analysis and sensitivity analysis to reduce model uncertainty by
4. accounting for parameter covariances. In addition, we plan to implement the power
5. analyses discussed above to more quantitatively inform data collection. A data
6. assimilation module is in progress for PEcAn that will allow the use of ecosystem level
7. data including plot-level inventory data, eddy covariance fluxes, and remote sensing
8. imagery to enter the analysis and provide additional constraint on uncertainty in both
9. parameters and output. The basic concept of variance decomposition will also be
10. expanded to investigate other sources of variability, such as uncertainty in initial
11. conditions or in driver data. We are implementing ecosystem models other than ED2 898 within the PEcAn workflow. This will provide opportunities for multi-model ensemble 899 forecasting and assessing data requirements across models.

900 Integrating modeling into a workflow system has distinct advantages not just for model 901 analysis but also for managing the flows of information coming in and out of the model.

1. In this sense we also envision PEcAn as an informatics and data management tool.
2. Finally, it is our hope that other researchers will find PEcAn useful and contribute 904 modules that extend the functionality of the system in creative and exciting ways.

# 905 Conclusion

1. In this paper, we demonstrate an approach to the parametrization of a terrestrial
2. biosphere model that synthesizes available data while providing a robust accounting of
3. parameter uncertainty. We also present a scientific workflow that enables more efficient
4. constraint of this uncertainty by identifying the key areas requiring data collection and
5. model refinement. By quantifying the effect that each parameter has on model output
6. uncertainty, this analysis identifies additional data that, once obtained, would improve 912 model precision. In addition, the analysis calculates probabilities of alternate potential

913 outcomes, resulting in more useful assessments.

# 914 Acknowledgments

1. Author contributions: DSL and MCD developed the concept, statistical analyses, and
2. writing; DSL implemented analyses and visualization; MCD provided scientific and 917 statistical guidance; DW provided feedback on design and implementation of BETY,
3. PEcAn, and this manuscript; KTR collected stomatal slope data; CDD contributed code.
4. Other contributions: Michael Sterling Burns assisted the writing process. Patrick
5. Mulrooney assisted with database design and implementation. John Skillman provided
6. raw data; John Skillman, Peter Reich, Ian Wright, and Stan Wullschleger provided
7. assistance with interpretation of their published data. DSL, DW, David Bettinardi, Andy 923 Tu, Xiaohui Feng, and others entered previously published data into the database and
8. contributed to the design of the web-based data entry workflow interface. Andrew Leakey,
9. Evan Delucia, and Don Ort assisted with prior estimation. Statistical and programming
10. advice was obtained from contributors to crossvalidated.com, stackoverflow.com, and
11. other stackexchange websites. We appreciate all individuals, including but not limited to
12. the authors of publications cited in this paper, whose time, efforts, and intellect designed
13. and produced the data used here, and whose recognition of the importance of archiving 930 and sharing large data sets contributes to the advancement of science.

931 Funding for this research was provided by the Energy Biosciences Institute and an NSF 932 Advances in Bioinformatics grant #1062547 to MCD.

# 933 References

934 Albert, C. H., F. Grassein, F. M. Schurr, G. Vieilledent, and C. Violle. 2011. When and 935 how should intraspecific variability be considered in trait-based plant ecology?

936 Perspectives in Plant Ecology, Evolution and Systematics **13**:217–225. URL 937 http://linkinghub.elsevier.com/retrieve/pii/S143383191100028X.

1. Amthor, J. 2000. The McCree-de Wit-Penning de Vries-Thornley Respiration Paradigms:
2. 30 Years Later. Annals of Botany **86**:1–20. URL
3. http://linkinghub.elsevier.com/retrieve/doi/10.1006/anbo.2000.1175.
4. Angel, J., 2010. Illinois State Climatologist Data for Station 118749 (Urbana). URL 942 http://www.isws.illinois.edu/atmos/statecli/Summary/118740.htm.
5. Ball, J., I. Woodrow, and J. Berry, 1987. A model predicting stomatal conductance and
6. its contribution to the control of photosynthesis under different environmental 945 conditions. Pages 221—-224 *in* J. Biggins, editor. Progress in Photosynthesis Research.
7. Martinus Nijhoff Publishers, Netherlands.
8. Bellocchi, G., M. Rivington, M. Donatelli, and K. Matthews. 2010. Validation of
9. biophysical models: issues and methodologies. A review. Agronomy for Sustainable
10. Development **30**:109–130. URL
11. http://www.springerlink.com/index/10.1051/agro/2009001.
12. Beven, K. 2006. A manifesto for the equifinality thesis. Journal of Hydrology **320**:18–36. 952 URL http://linkinghub.elsevier.com/retrieve/pii/S002216940500332X.
13. Beven, K., and J. Freer. 2001. Equifinality, data assimilation, and uncertainty estimation
14. in mechanistic modelling of complex environmental systems using the GLUE
15. methodology. Journal of Hydrology **249**:11–29. URL
16. http://linkinghub.elsevier.com/retrieve/pii/S0022169401004218.
17. Breza, L. C., L. Souza, N. J. Sanders, and A. T. Classen. 2012. Within and between
18. population variation in plant traits predicts ecosystem
19. functions associated with a dominant plant species. Ecology and Evolution **1**:no–no. URL
20. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3287307&tool=pmcentrez&ren
21. http://doi.wiley.com/10.1002/ece3.223.
22. Cariboni, J., D. Gatelli, R. Liska, and A. Saltelli. 2007. The role of sensitivity analysis in
23. ecological modelling. Ecological Modelling **203**:167–182. URL
24. http://linkinghub.elsevier.com/retrieve/pii/S0304380006005734.
25. Casella, G., and R. L. Berger. 2001. Statistical Inference. second edition. Duxbury Press, 966 Pacific Grove, CA.

967 Chapin III, F. S., P. A. Matson, and H. A. Mooney. 2002. Principles of Terrestrial 968 Ecosystem Ecology. Springer.

969 Clark, J. S., S. R. Carpenter, M. Barber, S. Collins, A. Dobson, J. a. Foley, D. M. Lodge, 970 M. Pascual, R. Pielke, W. Pizer, C. Pringle, W. V. Reid, K. a. Rose, O. Sala, W. H.

1. Schlesinger, D. H. Wall, and D. Wear. 2001. Ecological forecasts: an emerging
2. imperative. Science (New York, N.Y.) **293**:657–60. URL 973 http://www.ncbi.nlm.nih.gov/pubmed/11474103.
3. Clark, J. S., M. Silman, R. Kern, E. Macklin, and J. HilleRisLambers. 1999. Seed
4. dispersal near and far: patterns across temperate and tropical forests. Ecology 976 **80**:1475–1494. URL http://www.esajournals.org/doi/full/10.1890/0012977 9658(1999)080[1475:SDNAFP]2.0.CO;2.

978 Collatz, G., M. Ribas-Carbo, and J. Berry. 1992. Coupled Photosynthesis-Stomatal 979 Conductance Model for Leaves of C4 Plants. Functional Plant Biology **19**:519—-538.

1. URL http://www.publish.csiro.au/paper/PP9920519.
2. Denman, K., G. Brasseur, A. Chidthaisong, P. Ciais, P. Cox, R. Dickinson,
3. D. Hauglustaine, C. Heinze, E. Holland, D. Jacob, U. Lohmann, S. Ramachandran,
4. P. D. S. Dias, S. W. Lohmann, S Ramachandran, P.L. Da Silva Dias, X. Zhang, 984 X. Zhang, S. R. Lohmann, P. da Silva Dias, S. Wofsy, and X. Zhang, 2007. Couplings

985 between changes in the climate system and biogeochemistry. In: Climate Change 2007: 986 The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment

1. Report of the Intergovernmental Panel on Climate Change. Chapter 7, pages 499–587 .
2. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.
3. URL http://www.osti.gov/bridge/product.biblio.jsp?osti\_id=934721.
4. Dietze, M. C., and A. M. Latimer, 2011. Forest simulators. Pages 307–316 *in* A. Hastings 991 and L. Gross, editors. Sourcebook in Theoretical Ecology. Univ California Press, 992 Berkeley.
5. Dietze, M. C., R. Vargas, A. D. Richardson, P. C. Stoy, A. G. Barr, R. S. Anderson,
6. M. A. Arain, I. T. Baker, T. A. Black, J. M. Chen, P. Ciais, L. B. Flanagan, C. M. 995 Gough, R. F. Grant, D. Hollinger, R. C. Izaurralde, C. J. Kucharik, P. Lafleur, S. Liu, 996 E. Lokupitiya, Y. Luo, J. W. Munger, C. Peng, B. Poulter, D. T. Price, D. M.
7. Ricciuto, W. J. Riley, A. K. Sahoo, K. Schaefer, A. E. Suyker, H. Tian, C. Tonitto,
8. H. Verbeeck, S. B. Verma, W. Wang, and E. Weng. 2011. Characterizing the 999 performance of ecosystem models across time scales: A spectral analysis of the North 1000 American Carbon Program site-level synthesis. Journal of Geophysical Research **116**.
9. URL http://www.agu.org/pubs/crossref/2011/2011JG001661.shtml.
10. Ellison, A. M. 2010. Repeatability and transparency in ecological research. Ecology
11. **91**:2536–2539. URL http://www.esajournals.org/doi/abs/10.1890/09-0032.1.
12. Ernst, W. H. O., E. M. Veenendaal, and M. M. Kebakile. 1992. Possibilities for dispersal
13. in annual and perrennial grasses in a savanna in Botswana. Vegetatio **102**:1–11. URL 1006 http://www.springerlink.com/index/10.1007/BF00031700.
14. Farquhar, G. D., and T. D. Sharkey. 1982. Stomatal Conductance and Photosynthesis.
15. Annual Review of Plant Physiology **33**:317–345. URL
16. http://www.annualreviews.org/doi/abs/10.1146/annurev.pp.33.060182.001533. 1010 Gelman, A., and D. B. Rubin. 1992. Inference from Iterative Simulation Using Multiple
17. Sequences. Statistical Science **7**:457–472. URL
18. http://projecteuclid.org/euclid.ss/1177011136.
19. Gelman, A., and K. Shirley, 2011. Inference from simulations and monitoring 1014 convergence. Pages 163—-174 *in* S. Brooks, A. Gelman, G. Jones, and X.-L. Meng, 1015 editors. Handbook of Markov Chain Monte Carlo. CRC Press LLC.
20. Gill, R. A., and R. B. Jackson. 2000. Global Patterns of Root Turnover for Terrestrial
21. Ecosystems. New Phytologist **147**:13–31. URL 1018 http://www.jstor.org/stable/2588686.
22. Heaton, E. A., F. G. Dohleman, and S. P. Long. 2008. Meeting US biofuel goals with less
23. land: the potential of Miscanthus. Global Change Biology **14**:2000–2014. URL 1021 http://www3.interscience.wiley.com/journal/120119109/abstract.
24. Jackson, R. B., H. A. Mooney, and E. D. Schulze. 1997. A global budget for fine root
25. biomass, surface area, and nutrient contents. Proceedings of the National Academy of
26. Sciences of the United States of America **94**:7362–6. URL
27. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=23826.
28. Jones, M. B., M. P. Schildhauer, O. Reichman, and S. Bowers. 2006. The New
29. Bioinformatics: Integrating Ecological Data from the Gene to the Biosphere. Annual
30. Review of Ecology, Evolution, and Systematics **37**:519–544. URL
31. http://www.annualreviews.org/doi/abs/10.1146/annurev.ecolsys.37.091305.110031.
32. Jongejans, E., and P. Schippers. 1999. Modeling seed dispersal by wind in herbaceous 1031 species. Oikos **87**:362–372. URL http://www.jstor.org/stable/3546752.
33. Justus, C., W. Hargraves, A. Mikhail, and D. Graber. 1978. Methods for estimating wind
34. speed frequency distributions. J. Appl. Meteorol. **17**:350–353. URL
35. http://www.osti.gov/energycitations/product.biblio.jsp?osti\_id=5127748.
36. Kubien, D. S., and R. F. Sage. 2004. Low-temperature photosynthetic performance of a
37. C4 grass and a co-occurring C3 grass native to high latitudes. Plant, Cell and
38. Environment **27**:907–916. URL
39. http://doi.wiley.com/10.1111/j.1365-3040.2004.01196.x.
40. Kucharik, C. J., J. A. Foley, C. Delire, V. A. Fisher, M. T. Coe, J. D. Lenters,
41. C. Young-Molling, N. Ramankutty, J. M. Norman, and S. T. Gower. 2000. Testing the
42. performance of a dynamic global ecosystem model: Water balance, carbon balance, and
43. vegetation structure. Global Biogeochemical Cycles **14**:795. URL
44. http://www.agu.org/pubs/crossref/2000/1999GB001138.shtml.
45. Larocque, G. R., J. S. Bhatti, R. Boutin, and O. Chertov. 2008. Uncertainty analysis in
46. carbon cycle models of forest ecosystems: Research needs and development of a
47. theoretical framework to estimate error propagation. Ecological Modelling
48. **219**:400–412. URL
49. http://linkinghub.elsevier.com/retrieve/pii/S0304380008003542.
50. Leuning, R. 1995. A critical appraisal of a combined stomatal-photosynthesis model for
51. C3 plants. Plant, Cell and Environment **18**:339–355. URL
52. http://doi.wiley.com/10.1111/j.1365-3040.1995.tb00370.x.
53. Luo, Y., K. Ogle, C. Tucker, S. Fei, C. Gao, S. LaDeau, J. S. Clark, and D. S. Schimel.
54. 2011. Ecological forecasting and data assimilation in a data-rich era. Ecological 1054 applications : a publication of the Ecological Society of America **21**:1429–42. URL 1055 http://www.ncbi.nlm.nih.gov/pubmed/21830693.
55. Luo, Y., E. Weng, X. Wu, C. Gao, X. Zhou, and L. Zhang. 2009. Parameter
56. identifiability, constraint, and equifinality in data assimilation with ecosystem models.
57. Ecological applications : a publication of the Ecological Society of America **19**:571–4.
58. URL http://www.ncbi.nlm.nih.gov/pubmed/19425417.
59. Massad, R.-S., A. Tuzet, and O. Bethenod. 2007. The effect of temperature on C(4)-type
60. leaf photosynthesis parameters. Plant, cell & environment **30**:1191–204. URL 1062 http://www.ncbi.nlm.nih.gov/pubmed/17661755.
61. McLaughlin, S., and L. Kszos. 2005. Development of switchgrass (Panicum virgatum) as
62. a bioenergy feedstock in the United States. Biomass and Bioenergy **28**:515–535. URL 1065 http://linkinghub.elsevier.com/retrieve/pii/S0961953404001904.
63. McMahon, S. M., M. C. Dietze, M. H. Hersh, E. V. Moran, and J. S. Clark. 2009. A
64. predictive framework to understand forest responses to global change. Annals of the
65. New York Academy of Sciences **1162**:221–36. URL
66. http://www.ncbi.nlm.nih.gov/pubmed/19432650.
67. Medvigy, D., S. C. Wofsy, J. W. Munger, D. Y. Hollinger, and P. R. Moorcroft. 2009.
68. Mechanistic scaling of ecosystem function and dynamics in space and time: Ecosystem
69. Demography model version 2. Journal of Geophysical Research **114**:1–21. URL 1073 http://www.agu.org/pubs/crossref/2009/2008JG000812.shtml.
70. Mesinger, F., G. DiMego, E. Kalnay, K. Mitchell, P. C. Shafran, W. Ebisuzaki, D. Jovi´c,
71. J. Woollen, E. Rogers, E. H. Berbery, M. B. Ek, Y. Fan, R. Grumbine, W. Higgins,
72. H. Li, Y. Lin, G. Manikin, D. Parrish, and W. Shi. 2006. North American Regional
73. Reanalysis. Bulletin of the American Meteorological Society **87**:343. URL 1078 http://journals.ametsoc.org/doi/abs/10.1175/BAMS-87-3-343.
74. Michener, W. K., and M. B. Jones. 2012. Ecoinformatics: supporting ecology as a
75. data-intensive science. Trends in Ecology & Evolution **27**:85–93. URL
76. http://linkinghub.elsevier.com/retrieve/pii/S0169534711003399.
77. Miguez, F., X. Zhu, S. Humphries, G. Bollero, and S. Long. 2009. A semimechanistic
78. model predicting the growth and production of the bioenergy crop Miscanthus A˜˚U 1084 giganteus: description, parameterization and validation. GCB Bioenergy **1**:282–296.
79. URL http://doi.wiley.com/10.1111/j.1757-1707.2009.01019.x.
80. Mood, A. M., F. A. Graybill, and D. C. Boes. 1974. Introduction to the Theory of 1087 Statistics. 3rd edition. Mcgraw-Hill College.
81. Moorcroft, P. R., G. C. Hurtt, and S. W. Pacala. 2001. A Method for Scaling Vegetation
82. Dynamics: the Ecosystem Demography Model (ED). Ecological Monographs
83. **71**:557–586. URL http://www.esajournals.org/doi/abs/10.1890/00121091 9615%282001%29071%5B0557%3AAMFSVD%5D2.0.CO%3B2.
84. Moore, D. J., J. Hu, W. J. Sacks, D. S. Schimel, and R. K. Monson. 2008. Estimating
85. transpiration and the sensitivity of carbon uptake to water availability in a subalpine
86. forest using a simple ecosystem process model informed by measured net CO2 and
87. H2O fluxes. Agricultural and Forest Meteorology **148**:1467–1477. URL
88. http://linkinghub.elsevier.com/retrieve/pii/S0168192308001263.
89. Oehlert, G. 1992. A note on the delta method. American Statistician **46**:27–29. URL 1098 http://www.jstor.org/stable/10.2307/2684406.
90. Oyarzabal, M., J. M. Paruelo, F. Pino, M. Oesterheld, and W. K. Lauenroth. 2008. Trait
91. differences between grass species along a climatic gradient in South and North
92. America. Journal of Vegetation Science **19**:183–192. URL
93. http://blackwell-synergy.com/doi/abs/10.3170/2007-8-18349.
94. Parton, W. J., M. Hartman, D. Ojima, and D. Schimel. 1998. DAYCENT and its land 1104 surface submodel: description and testing. Global and Planetary Change **19**:35–48.
95. URL http://linkinghub.elsevier.com/retrieve/pii/S092181819800040X.
96. Plummer, M., 2010. JAGS Version 2.2.0 user manual.
97. Reich, P. B., and J. Oleksyn. 2004. Global patterns of plant leaf N and P in relation to
98. temperature and latitude. Proceedings of the National Academy of Sciences of the
99. United States of America **101**:11001–6. URL
100. http://www.pnas.org/cgi/content/abstract/101/30/11001.
101. Reichstein, M., J. Tenhunen, O. Roupsard, J.-M. Ourcival, S. Rambal, F. Miglietta,
102. A. Peressotti, M. Pecchiari, G. Tirone, and R. Valentini. 2003. Inverse modeling of
103. seasonal drought effects on canopy CO 2 /H 2 O exchange in three Mediterranean
104. ecosystems. Journal of Geophysical Research **108**. URL
105. http://www.agu.org/pubs/crossref/2003/2003JD003430.shtml.
106. Richardson, A., and D. Hollinger. 2005. Statistical modeling of ecosystem respiration 1117 using eddy covariance data: Maximum likelihood parameter estimation, and Monte
107. Carlo simulation of model and parameter uncertainty, applied to three simple models.
108. Agricultural and Forest Meteorology **131**:191–208. URL
109. http://linkinghub.elsevier.com/retrieve/pii/S0168192305001139.
110. Saltelli, A., M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, 1122 and S. Tarantola. 2008. Global Sensitivity Analysis. John Wiley & Sons, West Sussex, 1123 England.
111. Saugier, B., J. Roy, and H. Mooney, 2001. Estimations of global terrestrial productivity:
112. Converging toward a single number? Chapter 23, pages 543—-557 *in* B. Saugier and
113. H. Mooney, editors. Terrestrial Global Productivity. Academic Press, San Diego, CA. 1127 Schwalm, C. R., C. a. Williams, K. Schaefer, R. Anderson, M. A. Arain, I. Baker,

1128 A. Barr, T. A. Black, G. Chen, J. M. Chen, P. Ciais, K. J. Davis, A. Desai, M. Dietze, 1129 D. Dragoni, M. L. Fischer, L. B. Flanagan, R. Grant, L. Gu, D. Hollinger, R. C. 1130 Izaurralde, C. Kucharik, P. Lafleur, B. E. Law, L. Li, Z. Li, S. Liu, E. Lokupitiya, 1131 Y. Luo, S. Ma, H. Margolis, R. Matamala, H. McCaughey, R. K. Monson, W. C.

1. Oechel, C. Peng, B. Poulter, D. T. Price, D. M. Riciutto, W. Riley, A. K. Sahoo,
2. M. Sprintsin, J. Sun, H. Tian, C. Tonitto, H. Verbeeck, and S. B. Verma. 2010. A
3. model-data intercomparison of CO 2 exchange across North America: Results from the
4. North American Carbon Program site synthesis. Journal of Geophysical Research **115**.
5. URL http://www.agu.org/pubs/crossref/2010/2009JG001229.shtml.
6. Skillman, J. B. 2008. Quantum yield variation across the three pathways of
7. photosynthesis: not yet out of the dark. Journal of experimental botany **59**:1647–61.
8. URL http://www.ncbi.nlm.nih.gov/pubmed/18359752.
9. Stodden, V., D. Donoho, S. Fomel, M. Friedlander, M. Gerstein, R. LeVeque, I. Mitchell,
10. L. Larrimore, C. Wiggins, N. W. Bramble, P. O. Brown, V. J. Carey, L. DeNardis, 1142 R. Gentleman, J. D. Gezelter, A. Goodman, M. G. Knepley, J. E. Moore, F. A.

1143 Pasquale, J. Rolnick, M. Seringhaus, and R. Subramanian, 2010. Reproducible 1144 Research. URL http://doi.ieeecomputersociety.org/10.1109/MCSE.2010.113

1. http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5562471.
2. Thompson, S., and G. Katul. 2008. Plant propagation fronts and wind dispersal: an
3. analytical model to upscale from seconds to decades using superstatistics. The
4. American naturalist **171**:468–79. URL
5. http://www.ncbi.nlm.nih.gov/pubmed/18248297.
6. Tjoelker, M. G., J. M. Craine, D. Wedin, P. B. Reich, and D. Tilman. 2005. Linking leaf
7. and root trait syndromes among 39 grassland and savannah species. The New
8. phytologist **167**:493–508. URL
9. http://onlinelibrary.wiley.com/doi/10.1111/j.1469-8137.2005.01428.x/full.
10. USDA, and NRCS, 2011. The PLANTS Database. URL http://plants.usda.gov/.
11. VanLoocke, A., T. E. Twine, M. Zeri, and C. J. Bernacchi. 2012. A regional comparison
12. of water use efficiency for miscanthus, switchgrass and maize. Agricultural and Forest
13. Meteorology **164**:82–95. URL
14. http://linkinghub.elsevier.com/retrieve/pii/S0168192312001931.
15. Violle, C., B. J. Enquist, B. J. McGill, L. Jiang, C. H. Albert, C. Hulshof, V. Jung, and
16. J. Messier. 2012. The return of the variance: intraspecific variability in community
17. ecology. Trends in Ecology & Evolution **27**:244–52. URL 1162 http://www.ncbi.nlm.nih.gov/pubmed/22244797.
18. Wang, D., D. S. LeBauer, and M. C. Dietze. 2010. A quantitative review comparing the
19. yield of switchgrass in monocultures and mixtures in relation to climate and
20. management factors. GCB Bioenergy **2**:16–25. URL
21. http://blackwell-synergy.com/doi/abs/10.1111/j.1757-1707.2010.01035.x.
22. Wang, D., M. W. Maughan, J. Sun, X. Feng, F. E. Miguez, D. K. Lee, and M. C. Dietze.
23. 2011. Impacts of canopy position and nitrogen on nitrogen allocation and
24. photosynthesis of switchgrass (Panicum virgatum L.). Aspects of Applied Biology 112 1170 **112**:341—-351.
25. Williams, M., A. D. Richardson, M. Reichstein, P. C. Stoy, P. Peylin, H. Verbeeck,
26. N. Carvalhais, M. Jung, D. Y. Hollinger, J. Kattge, R. Leuning, Y. Luo, E. Tomelleri,
27. C. Trudinger, and Y.-P. Wang. 2009. Improving land surface models with FLUXNET
28. data. Biogeosciences Discussions **6**:2785–2835. URL
29. http://www.biogeosciences-discuss.net/6/2785/2009/.
30. Wright, I. J., P. B. Reich, M. Westoby, D. D. Ackerly, Z. Baruch, F. Bongers, 1177 J. Cavender-Bares, T. Chapin, J. H. C. Cornelissen, M. Diemer, J. Flexas, E. Garnier, 1178 P. K. Groom, J. Gulias, K. Hikosaka, B. B. Lamont, T. Lee, W. Lee, C. Lusk, J. J.
31. Midgley, M.-L. Navas, U. Niinemets, J. Oleksyn, N. Osada, H. Poorter, P. Poot,
32. L. Prior, V. I. Pyankov, C. Roumet, S. C. Thomas, M. G. Tjoelker, E. J. Veneklaas, 1181 and R. Villar. 2004. The worldwide leaf economics spectrum. Nature **428**:821–7. URL 1182 http://www.ncbi.nlm.nih.gov/pubmed/15103368.

1183 Wullschleger, S. D. 1993. Biochemical Limitations to Carbon Assimilation in C3 Plants 1184 Retrospective Analysis of the A/Ci Curves from 109 Species. Journal of experimental

1185 botany **44**:907–920.

# 1186 List of Tables

1. 1 **Prior Distributions** Prior distributions used in meta-analysis and model
2. parameterization. Prior sources come from citations as indicated except ∗ by
3. authors or † based on default ED2 parameterizations, as described in text.
4. The ’Clade’ column indicates the group of plants for which the priors were
5. derived. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 54
6. 2 **Meta-analysis Results** Results of meta-analysis of Switchgrass data for six
7. physiological traits. The global mean parameter, *β*0, is used to parametrize
8. the Ecosystem Demography model and is described in more detail by Fig-
9. ure 4. The variance components are transformed from precision to the stan-
10. dard deviation scale for ease of interpretation. Values are reported as the
11. parameter median with the 95% CI in parentheses. Units are the same as
12. in Table 1. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 55 1199 3 **a) Variance Estimates** Comparison of sample variances (*s*, on stan-
13. dard deviation scale) for the aboveground biomass estimated from data-
14. constrained parameters calculated from model ensemble, spline-emulated
15. model ensembles, and variance decomposition. Values in parentheses are
16. estimates of uncertainty in the sample estimate of variance. Sample size, *n*,
17. refers to the size of the sample from the posterior parameter distribution.
18. **b) Components of closure term,** *ω* The closure term *ω* (equation 6) is
19. 5.2, the difference between the variance decomposition and model ensemble
20. estimates of *σ*. The closure due to parameter interactions is estimated as the
21. difference between the spline ensemble and the model ensemble; the closure
22. due to the absence of a lower bound of zero on the spline functions is esti-
23. mated as the difference between the variance decomposition and the spline
24. ensemble estimates. ∗ Analysis of the closure term is based on estimates
25. with *n* = 10000 parameter sets, except in the case of the model ensemble
26. because evaluation of the model ensemble at *n* = 10000 is computationally
27. prohibitive. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 56

ParameterUnitsCladeDistributionabNmeanLCLUCLCitation

SpecificLeafAream

2

kg

−

1

GrassGamma2.0619.00125173.236(Wrightetal.,2004)

LeafTurnoverRate1/yrGrassWeibull2.900.63404.60.9111(Wrightetal.,2004)

RootTurnoverRate1/yrGrassGamma1.670.66660.590.0731.4(GillandJackson,2000)

QuantumEfficiencypercentC4grassesWeibull90.901580.00560.0580.0460.07(Skillman,2008)

StomatalSloperatioC4GrassGamma3.633.8143.41.45.5\*

VcmaxumolCO2m

−

2

s

−

1

graminoidGamma3.4924.7097228.636(Wullschleger,1993)

LeafWidthmmC4GrassWeibull26.105.94184.42.96.2(Oyarzabaletal.,2008)

RootRespirationRateumolCO

2

kg

−

1

s

−

1

C4GrassF5.612.33355.6110(Tjoelkeretal.,2005)

FineRootAllocationratioGrassBeta0.800.8103.10.4611(ChapinIIIetal.,2002)

SeedDispersalpercentGrasslog-Normal20.1074.90300.210.140.3(JongejansandSchippers,19

PhotosynthesismintempCelsiusC4GrassF10.001.02010812\*

GrowthRespirationpercentGrasslog-Normal2.636.5200.290.0620.6\*

SeedlingMortalitypercentmonocotslog-Normal3.610.4300.890.51\*

MortalityCoefficient1/yrplantsWeibull1.470.060251.880\*

ReproductiveAllocationpercentPlantslog-Normal2.004.0000.330.0530.72\*

Table1

Variablen

*β*

0

*σ*

*Y*

*σ*

site

*σ*

treatment

|

site

*β*

greenhouse

SpecificLeafArea2416(12

*,*

20)2

*.*

8(2

*.*

5

*,*

3

*.*

2)3

*.*

2(1

*.*

6

*,*

7

*.*

3)2

*.*

4(1

*.*

1

*,*

6)6

*.*

5(1

*,*

12)

LeafWidth396(4

*.*

7

*,*

6

*.*

6)0

*.*

46(0

*.*

44

*,*

0

*.*

48)0

*.*

47(0

*.*

2

*,*

2

*.*

1)6

*.*

4(1

*.*

9

*,*

130)1

*.*

6(

−

0

*.*

033

*,*

3

*.*

5)

Vcmax424(18

*,*

30)12(8

*.*

1

*,*

17)1

*.*

2(0

*.*

098

*,*

47)

FineRootAllocation41

*.*

3(0

*.*

5

*,*

2

*.*

6)2

*.*

2(1

*.*

2

*,*

6

*.*

2)

RootRespirationRate15

*.*

1(3

*.*

7

*,*

6

*.*

6)1

*.*

2(0

*.*

39

*,*

2

*.*

3)

RootTurnoverRate10

*.*

67(0

*.*

2

*,*

1

*.*

1)0

*.*

45(0

*.*

14

*,*

0

*.*

88)

StomatalSlope44

*.*

1(3

*.*

9

*,*

4

*.*

3)0

*.*

33(0

*.*

23

*,*

0

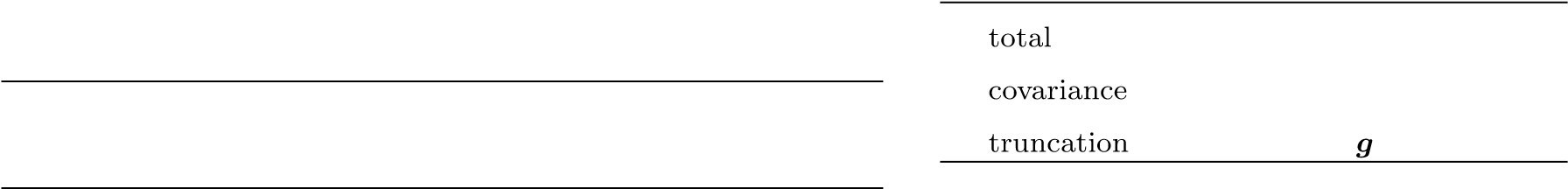
*.*

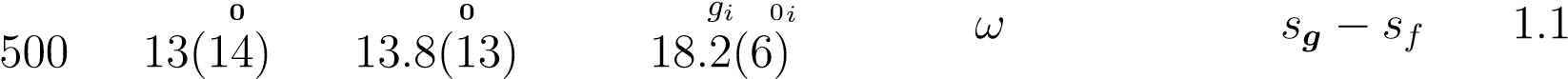
45)

Table2

model spline variance

calculation Mg/ha

 ensemble ensemble decomposition *ω* P*sgi* − *sf* 5.2

*n sf*(*β* ) *sg*(*β* ) P*s* (*β* )

10000 ∗ 14.1(2.8) 18.1(1.2) *ω* P*sgi* − *s* 4.1

(b)

(a)

Table 3

# 1215 List of Figures

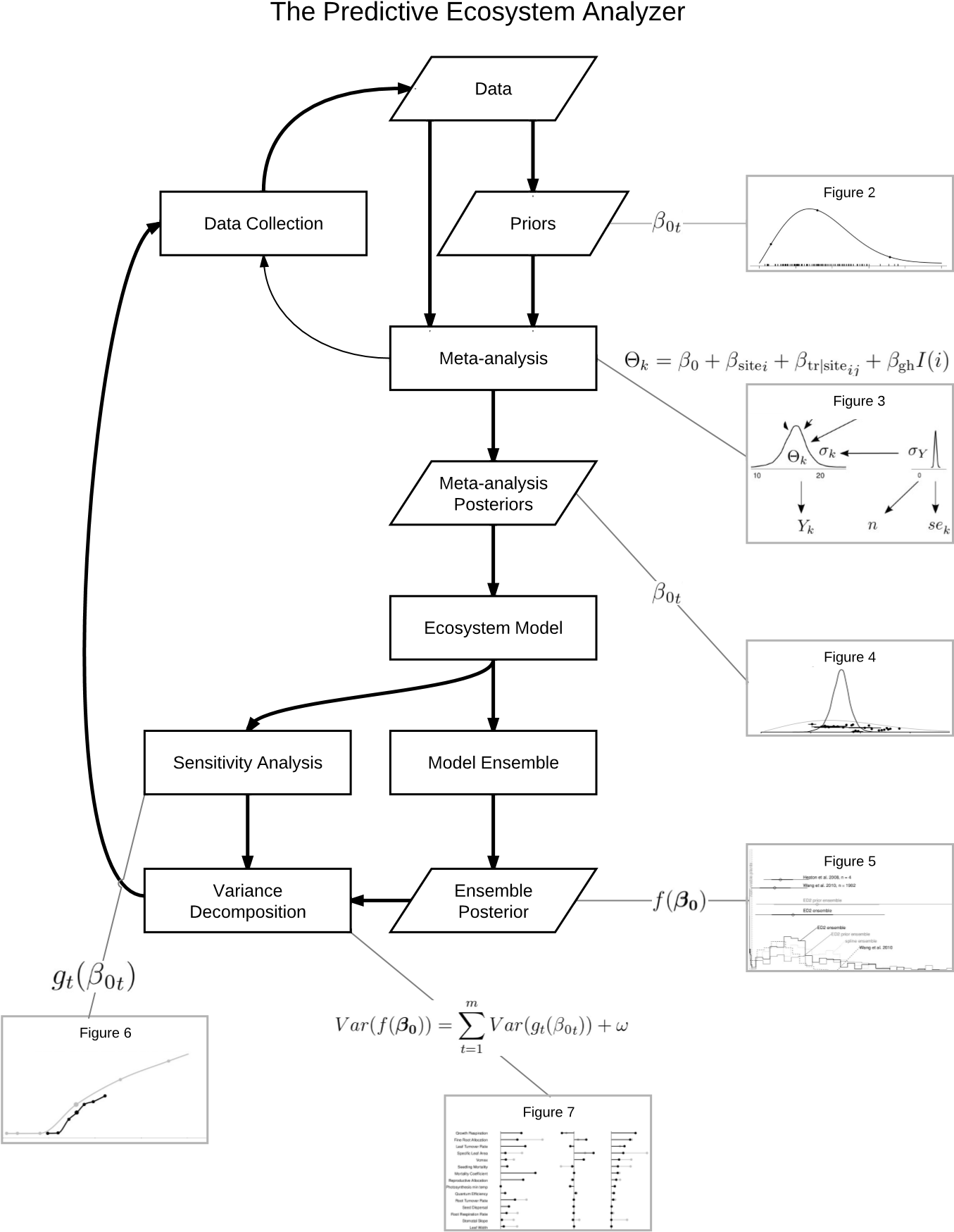
1. 1 **Overview of the PEcAn workflow**. The synthesis of plant trait data be-
2. gins by querying a database of plant trait data for data on a single species or
3. a plant functional type, and then mapping these data to the model parame-
4. ters that they inform. The database also provides probability distributions
5. that describe our prior information about the range of values that a model
6. parameter can take. Next, this information is synthesized in a Bayesian
7. meta-analysis, resulting in a posterior trait distribution that summarizes
8. the uncertainty in each parameter. The ensemble of model runs produces
9. the posterior distribution of model outputs, representing a probabilistic as-
10. sessment or forecast that accounts for input parameter uncertainty. The final
11. steps in the workflow are the sensitivity analysis and variance decomposition;
12. these steps gives insight into the relative contribution of each parameter to
13. the uncertainty in the model output, and can be used to guide additional
14. data collection that will most efficiently reduce model uncertainty. . . . . 62 1230 2 **Prior distributions** PDFs of priors with data constraints. Parameter value
15. is on the x-axis and probability density is on the y-axis, and the area under
16. each curve equals one. Three points on each line, from left to right, indicate
17. the 2*.*5th, 50th, and 97*.*5th quantiles. (From top left) Four priors fit to data
18. (data points shown as rug plot) using maximum likelihood: specific leaf
19. area and leaf turnover rate (Wright et al., 2004), root turnover rate (Gill
20. and Jackson, 2000), and quantum yield (Skillman, 2008). Four priors fit to
21. the posterior predictive distribution of an unobserved C4 grass species using
22. Bayesian meta-analysis of data from multiple plant functional types (C4
23. data shown in black, other functional types in grey): stomatal slope (present
24. study data provided in Appendix A), Vcmax of C3 plants (Wullschleger, 1993)
25. and C4 grasses (Kubien and Sage, 2004; Massad et al., 2007; Wang et al.,
26. 2011), leaf width (Oyarzabal et al., 2008), and root respiration (Tjoelker
27. et al., 2005). Priors fit to 95% CI (dashed grey line) and median (solid grey
28. line) based on ED2 defaults and expert opinion as described in the text:
29. fine root to leaf ratio (Chapin III et al., 2002), seed dispersal (Ernst et al.
30. (1992) model parameterized with site level data), minimum temperature
31. of photosynthesis (Don Ort, personal communication), growth respiration,
32. seedling mortality factor, mortality factor, and reproductive allocation. . . 63 1249 3 **Overview of the Hierarchical Bayesian meta-analysis model.** For
33. each trait, the posterior estimate of the global trait mean (*β*0) is used as
34. an input parameter in the sensitivity analysis and model ensemble (Figures
35. 6 and 5). Results from the meta-analysis of specific leaf area are as an
36. illustrative example; x-axes have units of m2kg−1 and all plots are on the
37. same scale. Each of the *k* sample means (*Yk*) are taken from published
38. articles and unpublished field measurements, and may be associated with a
39. sample standard error and sample size. When sufficient data were available,
40. site, treatment, and greenhouse effects were estimated. The within-unit
41. standard deviation, *σY* , is estimated from *se* and *n*. Site and treatment
42. random effects, represented by *β*site and *β*tr|site, are estimated for each site
43. and treatment within site with from normal distributions with mean zero
44. and standard deviations *σ*site and *σ*tr|site, respectively. Greenhouse is a fixed
45. effect. Table 2 summarizes the global mean, variance terms, and greenhouse
46. effect for the seven model parameters informed by species-level data. . . . 64 1264 4 **Prior (gray) and posterior (black) densities of trait parameters**
47. **used in the analysis.** Priors distributions are based on the traits of
48. plants within broad taxonomic or functional type groupings (e.g. all grasses).
49. When species-level data were available, they are used in a hierarchical Bayesian
50. meta-analysis, and the posterior estimate of the mean parameter value is
51. shown. Data used in the meta-analysis come from both published and direct
52. measurements of the trait on the perennial C4 grass Switchgrass (*Panicum*
53. *virgatum*). These data are represented as mean ±SE. Mismatch between
54. data and the posterior estimate of the global trait mean results from site,
55. treatment, and greenhouse effects. Data from plants grown under an experi-
56. mental treatment or in a controlled environment (e.g. in a pot or greenhouse)
57. are presented in grey; data from field-grown plants under control treatments
58. are in black. Site-level effects account for disparity between raw data and
59. parameter distribution in the SLA and leaf width plots. . . . . . . . . . . 65

## 1278 5 Ensemble average 2004-2006 post-senescence yield. Histogram of re-

1. sults from prior ensemble runs (dashed), posterior ensemble runs (solid line),
2. and the spline ensemble (gray line). The gray box on the left represents non-
3. viable ensemble members (≤ 2Mg/ha, see text). Horizontal bars provide a
4. summary of yields, from top: a three year trial at the modeled site (Heaton
5. et al., 2008), all 1902 observations included in a recent meta-analysis (Wang
6. et al., 2010), viable runs from the ED2 ensemble based on prior and poste-
7. rior parameterizations. Diamonds indicate the median; thick and thin lines
8. indicating the 68% and the 95% CI, respectively. Histogram-style plots pro-
9. vide comparison of the distributions of observations and model runs. For
10. clarity, non-viable and viable runs are plotted with different width bins. . 66 1289 6 **Univariate relationships between parameters and 2004-2006 aver-**
11. **age modeled yield.** Parameter values are on the x-axis and biomass is on
12. the y-axis while runs centered around the prior median are in gray and those
13. centered around the posterior median are in black. The univariate responses
14. were estimated using a cubic spline to fit model output at the median and
15. ±[1*,*2*,*3]*σ* quantiles of each parameter while holding other parameters to the
16. median value. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 67

## 1296 7 Partitioning of variance by parameter results from variance decom-

1. position conducted before (grey) and after (black) updating parameter es-
2. timates with species-level data in the meta-analysis. From left to right,
3. panels present: a) the uncertainty associated with each parameter (coeffi-
4. cient of variation, CV = *σ/µ*). The degree to which some parameters have
5. been constrained by data is indicated by the reduction in CV in the black
6. compared to the grey bars; sample sizes are indicated in Table 2. b) the
7. sensitivity of modeled aboveground biomass to each parameter presented
8. as elasticity (elasticity is normalized sensitivity, and an elasticity of one in-
9. dicates that model output will double when the parameter value doubles).
10. Sensitivity is the slope of the line at the median in Figure 6). Parameters
11. with larger bars have greater influence on model output. c) Partial variance
12. is the contribution of each parameter to explained variance. This is a func-
13. tion of both the parameter variance and sensitivity. Parameters with both
14. large CV and elasticity contribute the most to uncertainty in model output. 68 Figure 1



Specific Leaf Area

●

|  |  |  |
| --- | --- | --- |
| ● |  | ● |
|  |  |  |

1. 10 20 30 40 50 m2 kg−1

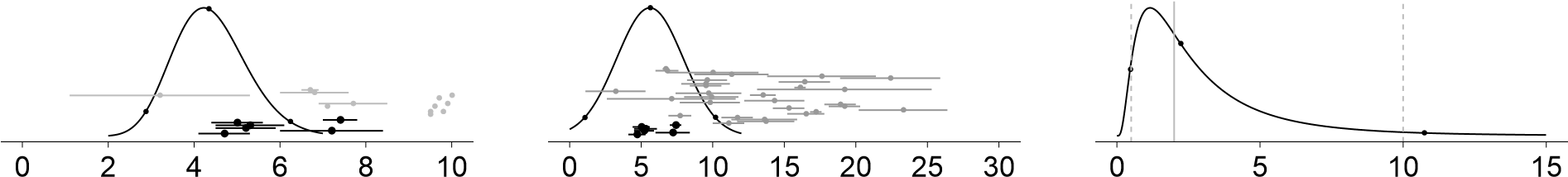
Quantum Efficiency

●

|  |  |  |
| --- | --- | --- |
| ● |  | ● |
|  |  |  |

0.04 0.05 0.06 0.07 0.08

fraction

Leaf Width

mm

Seed Dispersal

●

●

●

0.0

0.1

0.2

0.3

0.4

0.5

fraction

Seedling Mortality

●

●

●

0.4

0.6

0.8

1.0

fraction

Leaf Turnover Rate Root Turnover Rate

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | | ● | | ● | ● | | | | | |  | | | | | |  |  |  |  |  |  |  |  | | |  |  |  | | --- | --- | --- | | ● | ● | ● | |  |  |  | |

0 5 10 15 20 0.0 0.5 1.0 1.5 2.0 yr−1 yr−1

|  |  |
| --- | --- |
| umol CO2 kg−1 s−1 | ratio |
| Photosynthesis min temp | Growth Respiration |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ● |  |  |  |  | ● |
| ● |
|  |  |  |  |  |  |

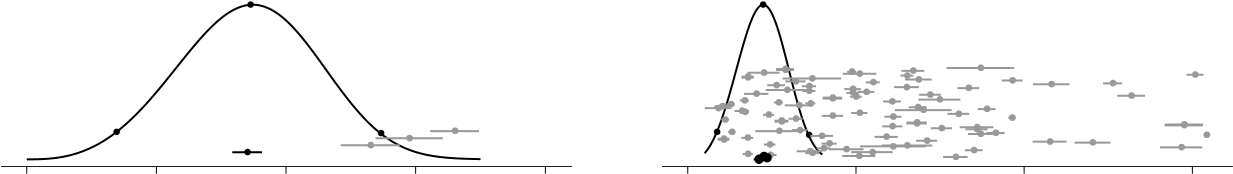
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | ● | ● |  | ● |
|  |
|  |  |  |  |  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 6 | 8 10 12  Celsius | 14 | 0.0 | 0.2 0.4 0.6 0.8 fraction | 1.0 |
|  | Mortality Coefficient |  |  | Reproductive Allocation |  |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ● | |  | | ● | | | | |
| ● | |
|
|
|
|  |  |  |  |  |  |  |  |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| ● | ● |  | |  | | |
|
| ● | | |
|  |  | |  | |  |  |

Stomatal Slope Vcmax



|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 0 | 2 4 6 ratio | 8 | 0 | 50 100  umol CO2 m−2 s−1 | | 150 |
|  | Root Respiration Rate |  |  | Fine Root Allocation | |  |
| 0 50 100 dimensionless | | 150 | 0.0 0.2 0.4 0.6 fraction | | 0.8 | 1.0 | |

Figure 2

Figure 3

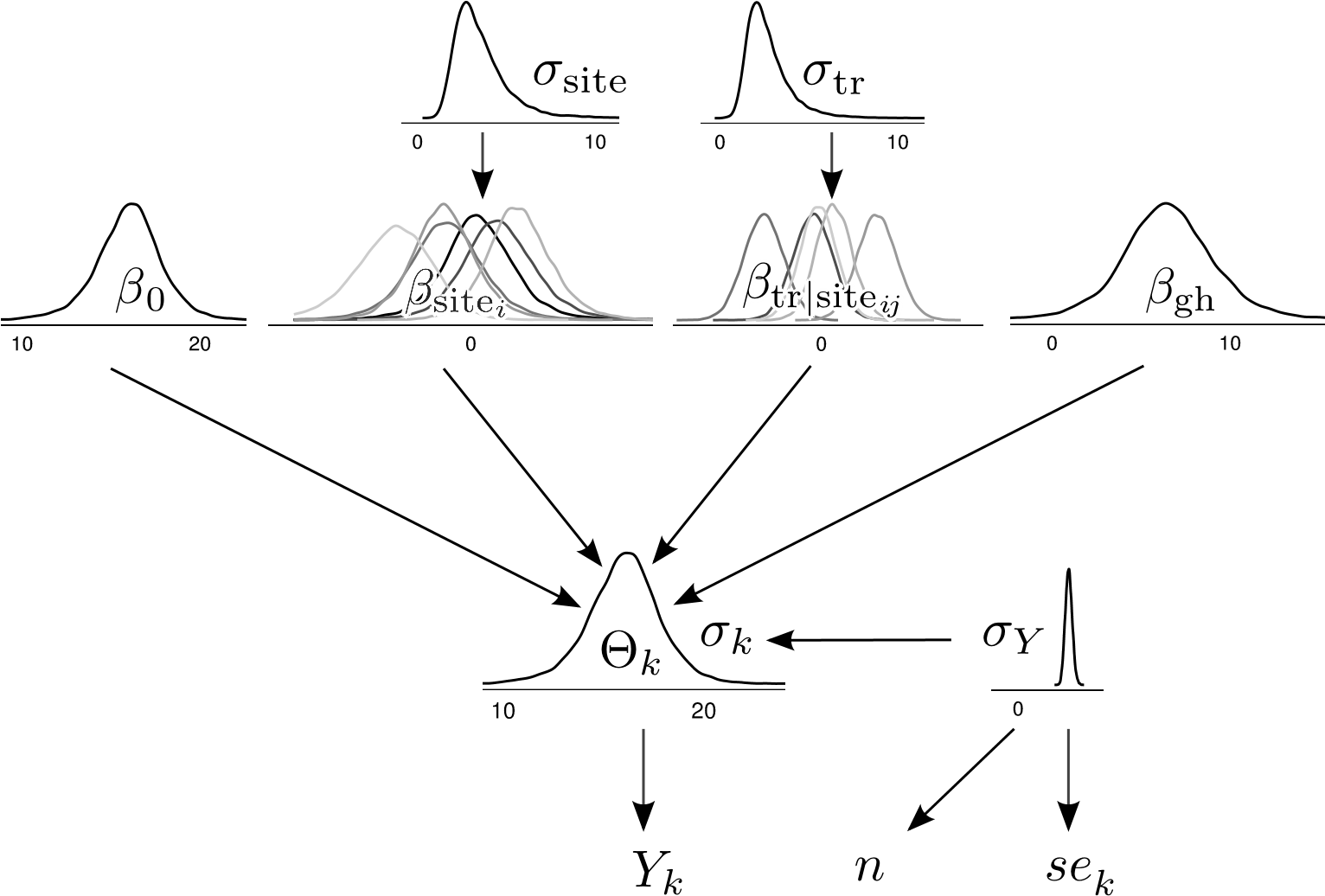


Figure 4

|  |  |
| --- | --- |
| mm | ratio |
| Root Turnover Rate | Stomatal Slope |

Leaf Width Fine Root Allocation

●

●

●

●

●

●

●

~~●~~

~~●~~

~~●~~

●

~~●~~

~~●~~

~~●~~

~~●~~

~~●~~

~~●~~

~~●~~

~~●~~

●

~~●~~

~~●~~

●

~~●~~

●

~~●~~

●

●

●

●

●

~~●~~

●

●

~~●~~

●

●

0

2

4

6

8

10

12

●

●

●

●

0

2

4

6

8

10

|  |  |
| --- | --- |
| ●  ~~●~~ | |
| ~~●~~ | ● |

0 2 4 6 8

●

0

0.5

1

1.5

2

|  |  |
| --- | --- |
| yr−1 | ratio |
| Root Respiration Rate | Specific Leaf Area |

●

0

5

10

15

●

●

●

●

●

●

●

●

●

●

●

●

●

●

●

●

●

●

●

●

●

●

●

~~●~~

0

10

20

30

40

umol CO2 kg−1 s−1 m2 kg−1

Vcmax

●

●

●

●

0

10

20

30

40

50

umol CO2 m−2 s−1

Figure 5

Modeled and Observed Switchgrass Yield

Aboveground Biomass, Mg/ha

density

non−viable plants

spline ensemble

ED2 prior ensemble

ED2 ensemble

Wang et al. 2010

Wang et al. 2010, n = 1902

Heaton et al. 2008, n = 4

ED2 ensemble

ED2 prior ensemble

0

20

40

60

80

100

Figure 6

Sensitivity of Aboveground Biomass (Mg/ha) to Fifteen Plant Traits

Mortality Coefficient Growth Respiration Leaf Turnover Rate

|  |  |  |
| --- | --- | --- |
|  | 40  20 | ●  ● ●  ●  ●  ●  ●● ●●  ● |
|
| ● ●● ● ● ●  ●● ● ● ● |
| ●  ●  ●  ● |
| ● ●● ●● |

4040

●

●

●

●

●

●

●

●

●

●

●

●

●

●

●

●

2020

1. 0 0

0 50 100 150 0.0 0.2 0.4 0.6 0.8 0 5 10 15 20 dimensionless fraction yr−1

Leaf Width

mm

0

20

40

●

●

●

●

●

●

●

●

●

●

●

●

●

●

●

●

0

5

10

15

Seed Dispersal

fraction

0

20

40

●

●

●

●

●

●

●

●

●

●

●

●

●

●

●

●

0.0

0.1

0.2

0.3

0.4

Fine Root Allocation

ratio

0

20

40

●

●

●

●

●

●

●

●

●

●

●

●

●

●

●

0

5

10

15

20

25

30

Root Turnover Rate

yr−1

0

20

40

●

●

●

●

●

●

●

●

●

●

●

●

●

●

●

●

0.0

0.5

1.0

1.5

2.0

2.5

Seedling Mortality

fraction

0

20

40

●

●

●

●

●

●

●

●

●

●

●

●

●

●

0.0

0.2

0.4

0.6

0.8

1.0

Stomatal Slope

ratio

0

20

40

●

●

●

●

●

●

●

●

●

●

●

●

●

●

●

●

0

2

4

6

8

|  |  |  |
| --- | --- | --- |
| ●  ●  ●  ●● ● ●  ●●  ●  ● ● ● ● | 40  20 | ● |
| ●  ●  ●  ●● ● ●  ●● |
|
| ●   * ● * ● ● |

0.0 0.2 0.4 0.6 0.8 1.0 0 5 10 fraction umol CO2 kg−1 s−1

|  |  |
| --- | --- |
| Specific Leaf Area | Vcmax |

40

20

|  |  |  |  |
| --- | --- | --- | --- |
| 0  0 10 20 30  m2 kg−1 | 40 50 | 0  0 10 20 30  umol CO2 m−2 s−1 | 40 |

●

Photosynthesis min temp

Reproductive Allocation

0

20

40

●

●

●

●

●

●

●

●

●

●

●

●

●

●

●

●

Root Respiration Rate

0

20

40

●

●

●

●

●

●

●

●

●

●

●

●

●

●

●

●

15

Celsius

0

20

40

●

●

●

●

●

●

●

●

●

●

●

●

●

●

●

●

0

5

10

15

Quantum Efficiency

0

20

40

●

●

●

●

●

●

●

●

●

●

●

●

●

●

●

●

50 0.00 0.02 0.04 0.06 0.08

fraction

Figure 7

Parameter CV (%) Elasticity Root Variance (SD)

|  |  |  |
| --- | --- | --- |
|  |  |  |
| ● ● | ● ● | ●● |
| ● ●● | ● ● |
| ● ● | ● ● | ● ● |
| ● ● | ●●  ●  ●●  ●  ● ● | ● ● |
| ● ● ● | ● ● |
| ● | ●● |
| ● ●●  ● | ●●  ●●  ●●  ●●  ●  ●● |
|
| ● |
| ● ● ●● |
| ● ● |
| ● ● ● |
| ● ● ● | ● ●  ●● |
| ● ● ● |

Growth Respiration ● ● ● ●

Fine Root Allocation

Leaf Turnover Rate

Specific Leaf Area

Vcmax

Seedling Mortality

Mortality Coefficient

Reproductive Allocation

Photosynthesis min temp ●

Quantum Efficiency

Root Turnover Rate

Seed Dispersal

Root Respiration Rate

Stomatal Slope Leaf Width

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 0 | 0 | 20 | 40 | 60 | 80 100 | −2 | 0 | 2 | 4 | 0 | 5 | 10 | 15 | 20 |