Region-wide calibration of 3PG using data assimilation

Annika Jersild & R. Quinn Thomas
Virginia Tech

Many others from the 3PG team

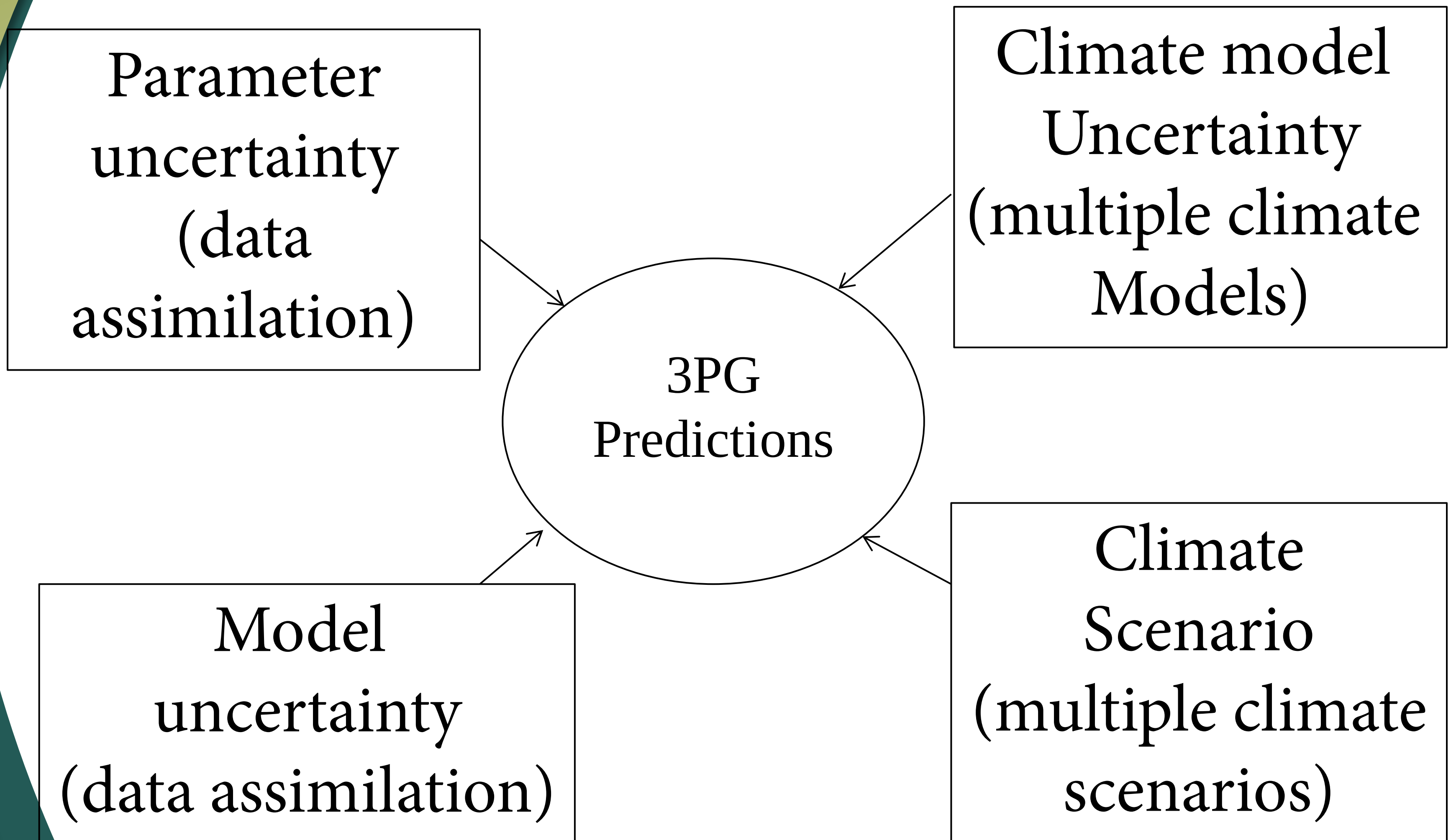
PINEMAP is a Coordinated Agricultural Project funded by the
USDA National Institute of Food and Agriculture, Award # 2011-68002-30185.

USDA United States Department of Agriculture National Institute of Food and Agriculture

Overall Goal

Make predictions for loblolly pine productivity with a known uncertainty that is consistent with region-wide data and prior knowledge

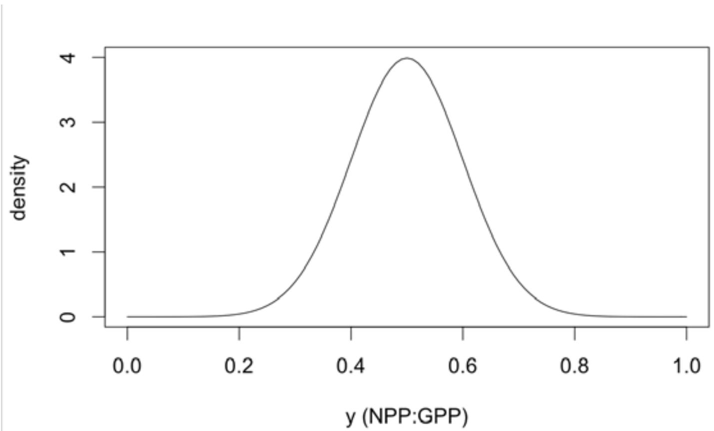
Known Unknowns



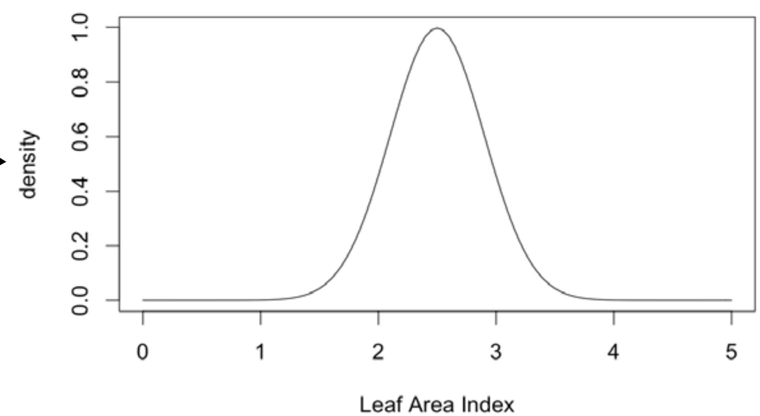
Other uncertainty: observation uncertainty

What is Data Assimilation?

- Formal integration of data into model parameterization, in our case, using Bayesian techniques
- Development of probability distributions for parameters based on data and prior knowledge
- Used to develop uncertainty predictions based on the parameter uncertainty
- Allows us to utilize multiple different data types



Parameter uncertainty



Prediction uncertainty

Metropolis-Hastings Markov-Chain Monte Carlo Method (a practical view)

1. Start at initial values for parameters
2. Draw new parameters by randomly selecting from a normal distribution (centered around current value) with a defined standard deviation (defined as jump variable)
3. Run 3PG model at multiple sites using the newly selected parameters
4. Calculate likelihood of output (summed across plots and data types) using a normal distribution for each data type
5. Calculate probability of the priors based on a uniform distribution
6. Multiply likelihood times prior
7. If current set of parameters is more probable than previous set, accept parameters. If not, accept in proportion to how much worse.
    - Accept parameters as a batch, not a single parameter, to develop covariance
8. Repeat 2-7 for desired number of iterations

Bayes Rule (simplified)

P(parameters | observations) ~

P(parameters) ← Priors

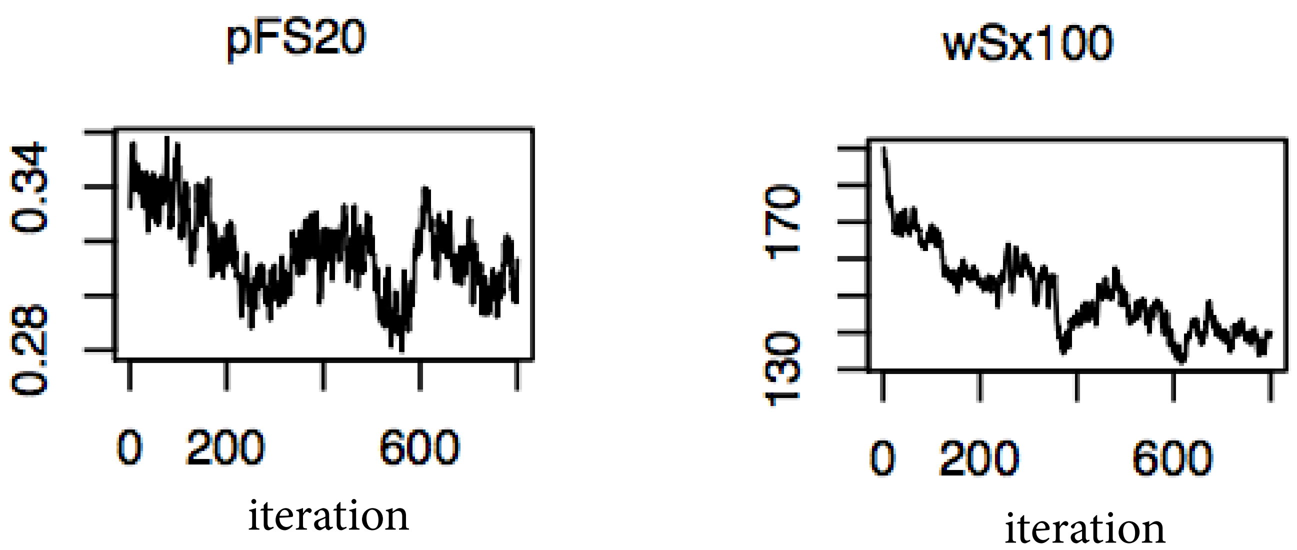
x

Likelihood

P(observations | parameters)

Example Chains

- A "fuzzy caterpillar"
- A converged chain still has variance around a mean, trying multiple values but overall jumping around the most likely value for a parameter
- Sets inherently have covariance between parameters



Overview: Our Code

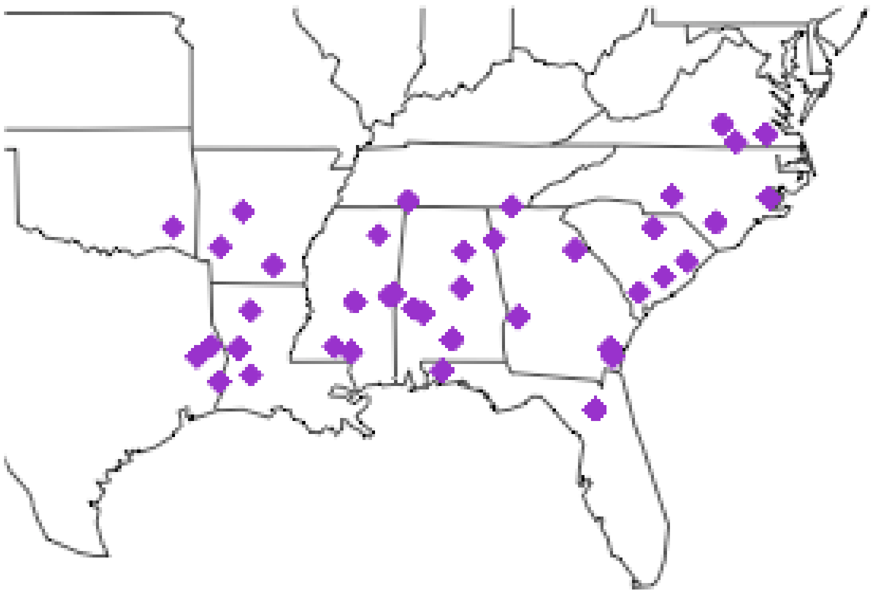
- The code works and we have tested up to 100,000 iterations
- We have 30 sites with 58 total parameters
  - 49 specific to 3PG
  - 3 specific to FR model
  - 6 are the standard deviation for likelihood calculation of each observation type
- R runs the MCMC and calls the Fortran 3PG code. Different plots within an iteration are run in parallel
  - Currently using Flavor C but can be switched to a different Flavor
- Using Santosh FR model form but parameters for FR model are fitted
- User can specify number of iterations, number and location of plots

Priors

- Where do they come from?
  - Expert opinion
  - Prior published values
- What do we need?
  - Distributions for each parameter
  - Currently: uniform priors (max and min possible)
- Prior in analyses shown here are from an initial literature review or broad bounds

Likelihood and Data

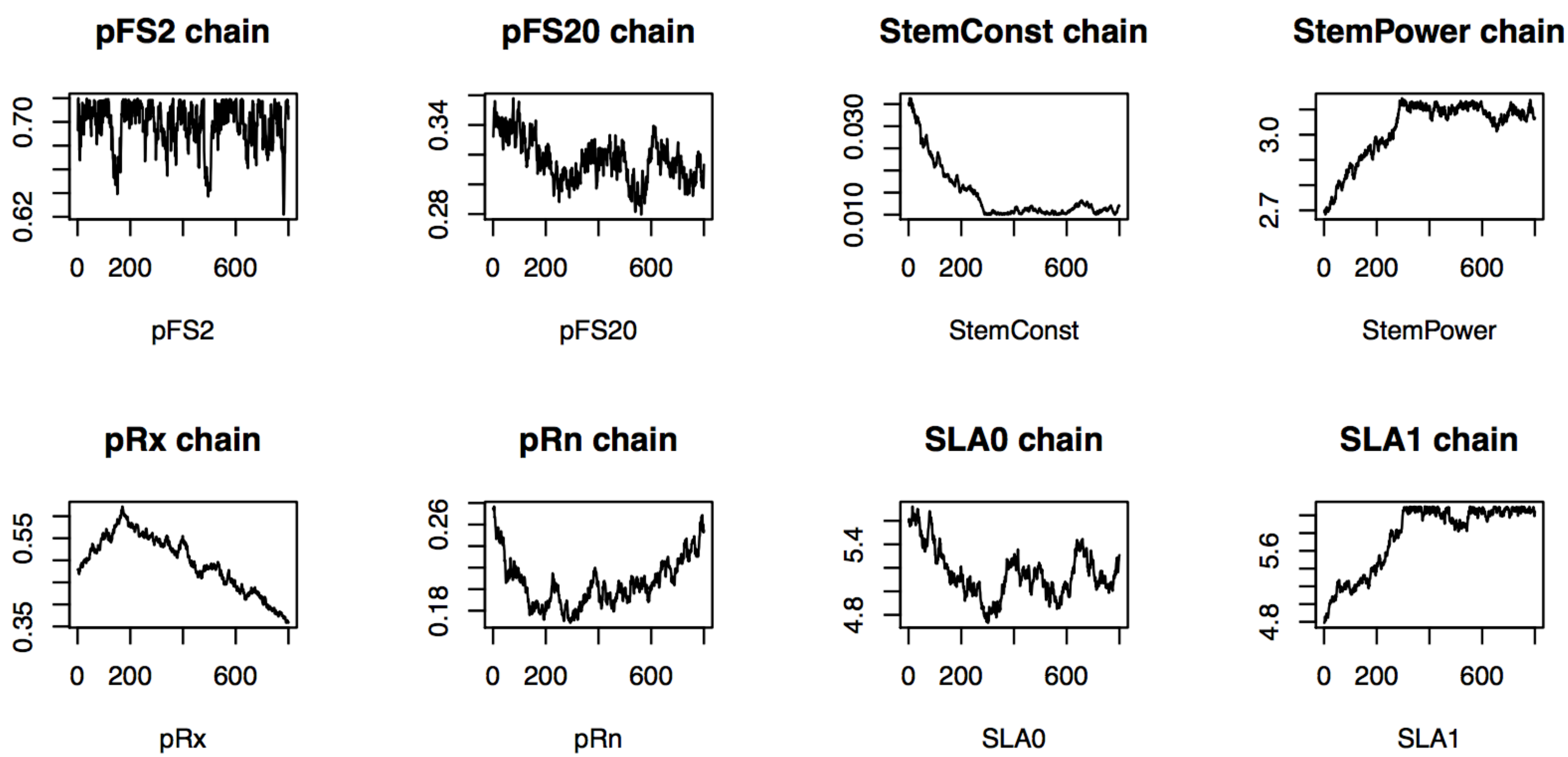
- Used plots from Carlos
  - Data used: stem number, foliage biomass, stem biomass, roots biomass, basal area, volume outside bark
  - Inputs: climate (Idaho), age, available soil water, planting year, seedling size, soil class, latitude, site index
- Likelihood of data given the model is normally distributed



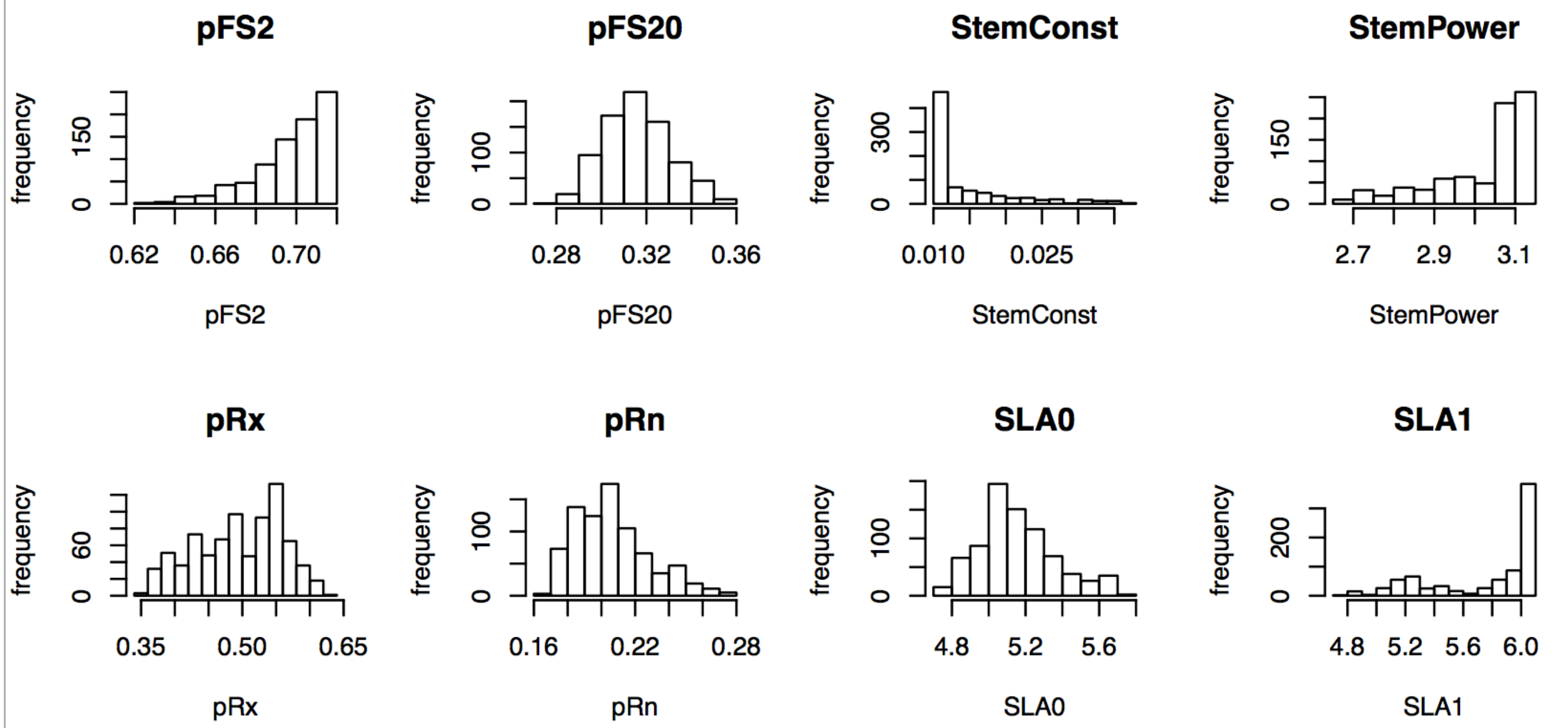
Example of parameter searching

Single plot

Preliminary Results: Trace Plots



Preliminary Results: Histograms



Prediction

- Use the parameter distributions and covariation among parameters to run 3PG
  - Sample from accepted parameter sets
  - Run 3PG model with each different sampled set
  - Develop model output with uncertainty based on our parameter uncertainty given the region-wide data and priors
- We can use climate change predictions for regions with our parameter distributions and develop growth predictions with uncertainty

Priors

- Where do they come from?
  - Expert opinion
  - Prior published values
  - Measurements from PINEMAP
- As a group we can develop a value set of priors

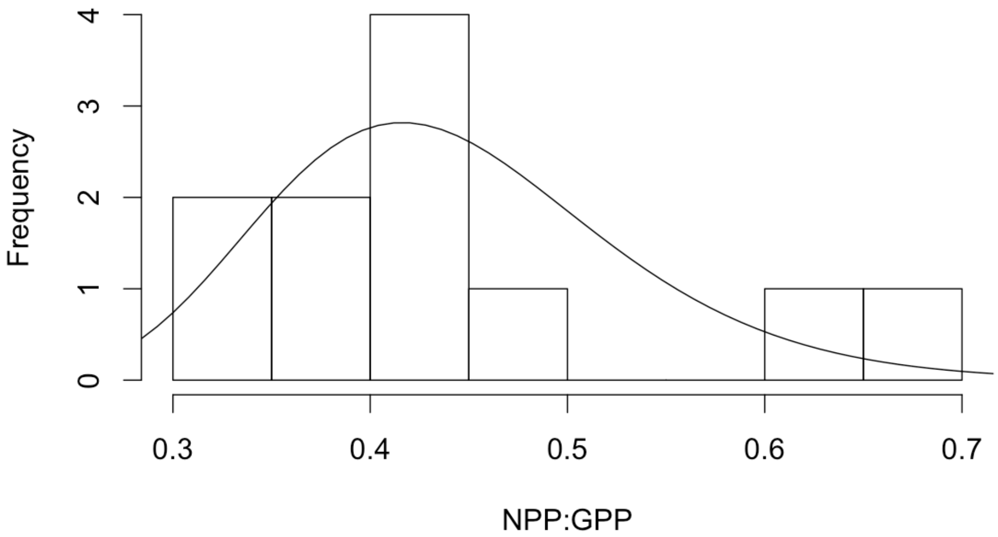
Light extinction coefficient	specific leaf area	quantum yield	leaf life-span	root life-span
allocation patterns	ratio of bark to non-bark stem	thinning curve exponent	NPP:GPP ratio	Foliage biomass on trees that die

Priors example (NPP:GPP)

Global Change Biology (2007) **13**, 1157-1167, doi: 10.111/j.1365-2486.2007.01365.x

Forest carbon use efficiency: is respiration a constant fraction of gross primary production?

Evan H DeLucia, John E. Drake, Richard B Thomas, and Miquel Gonzalez-Meler



Priors: continued

- Ecological rules where model parameter combinations are rejected when determined ecologically "nonsensical"
    - relative parameters: i.e. Tmax > Tmin
    - fracBB0 > fracBB1

Biogeosciences, 12, 1299-1315, 2015
www.biogeosciences.net/12/1299/2015/
doi: 10.5194/bg-12-1299-2015

Constraining ecosystem carbon dynamics in a data-limited world: integrating ecological "common sense" in a model-data fusion framework

A.A. Bloom and M. Williams

Plans & Goals

- Fine-tune priors to further constrain parameters
- Incorporate Duke FACE site to constrain CO_2 response, AMERIFLUX (Duke and Coastal NC) sites to constrain photosynthesis and evaporation
- Run 5 million iterations for 30 sites
- Develop region-wide assessments and predictions
- Compare to sites not used in calibration
- Compare different types of uncertainty
    - Parameter (parameter distribution)
    - Model (estimated error terms)
    - Climate model (20 different climate models)
    - Climate scenario (2 different scenarios – RCP 4.5 and 8.5)