PECAn:
workflow management for data assimilation and forecasting

David LeBauer, Dan Wang, Xiaohui Feng, and Michael Dietze

Energy Biosciences Institute

bp

ILLINOIS

Overview

Goals

Improve ecological forecasts

Investigate and reduce uncertainty

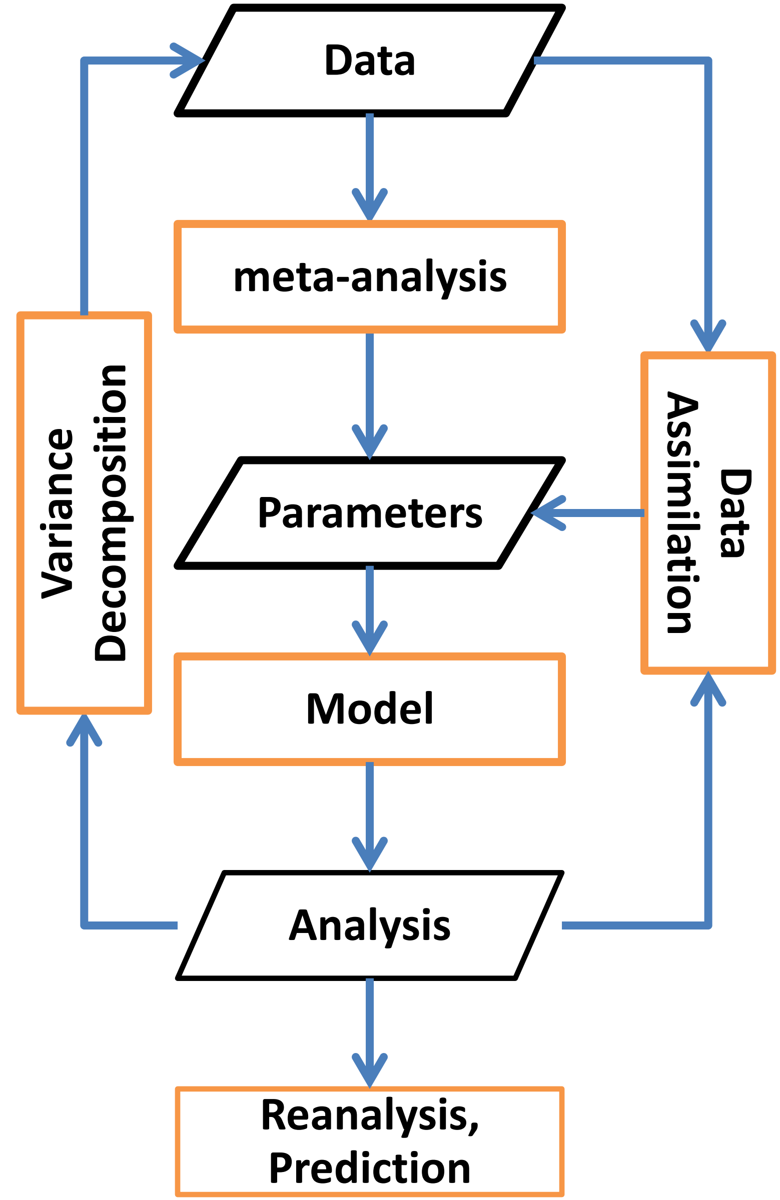
Reproducible computation

Ecosystems

Biofuel Agriculture

- Perennial grass
- Woody crops
- LIHD prairie

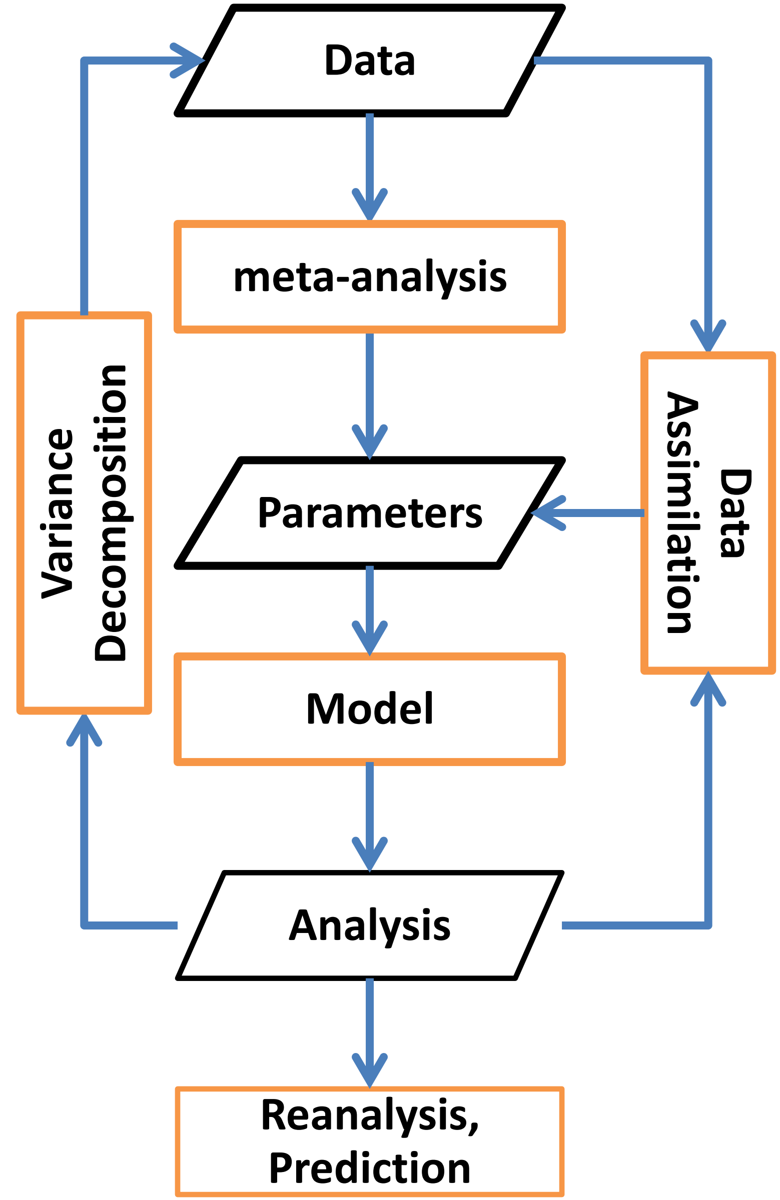
Temperate forests



Overview

Approach

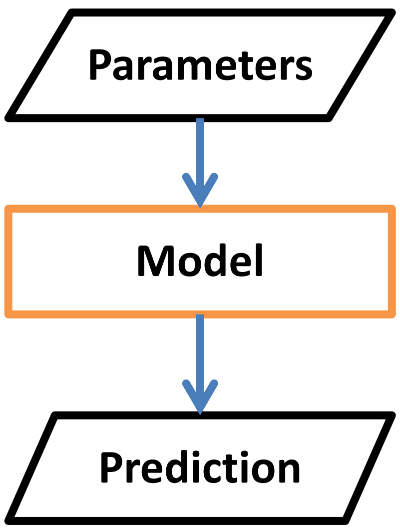
Bayesian meta-analysis
Data Summary
ED2 Ecosystem Model
Vegetation Model
Data Assimilation
Parameter Constraint
Variance Decomposition
Understand Uncertainty
Workflow management (PECAn)
Efficient Computation



Overview: Modeling approaches

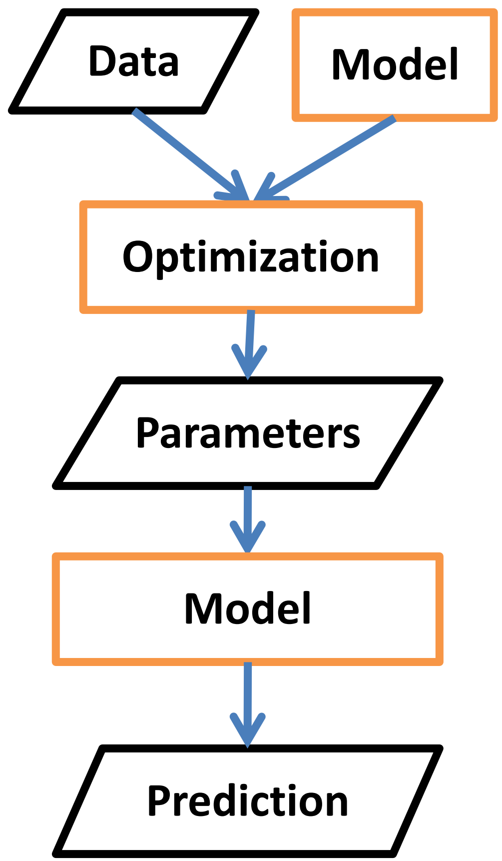
Conventional

No formal data integration



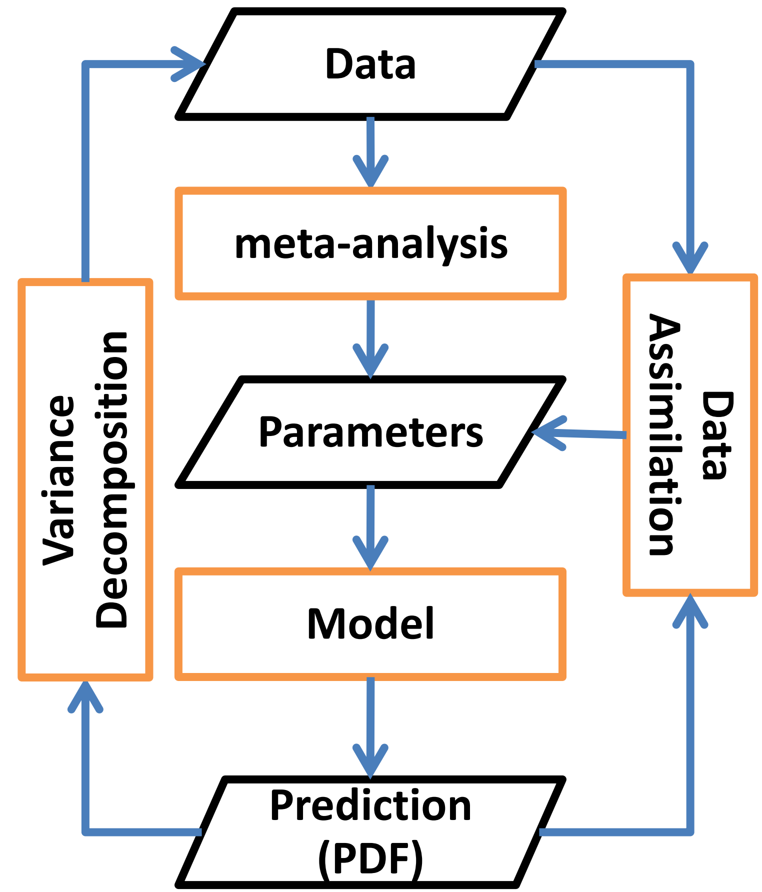
Inversion

Computationally expensive
No PDF of prediction

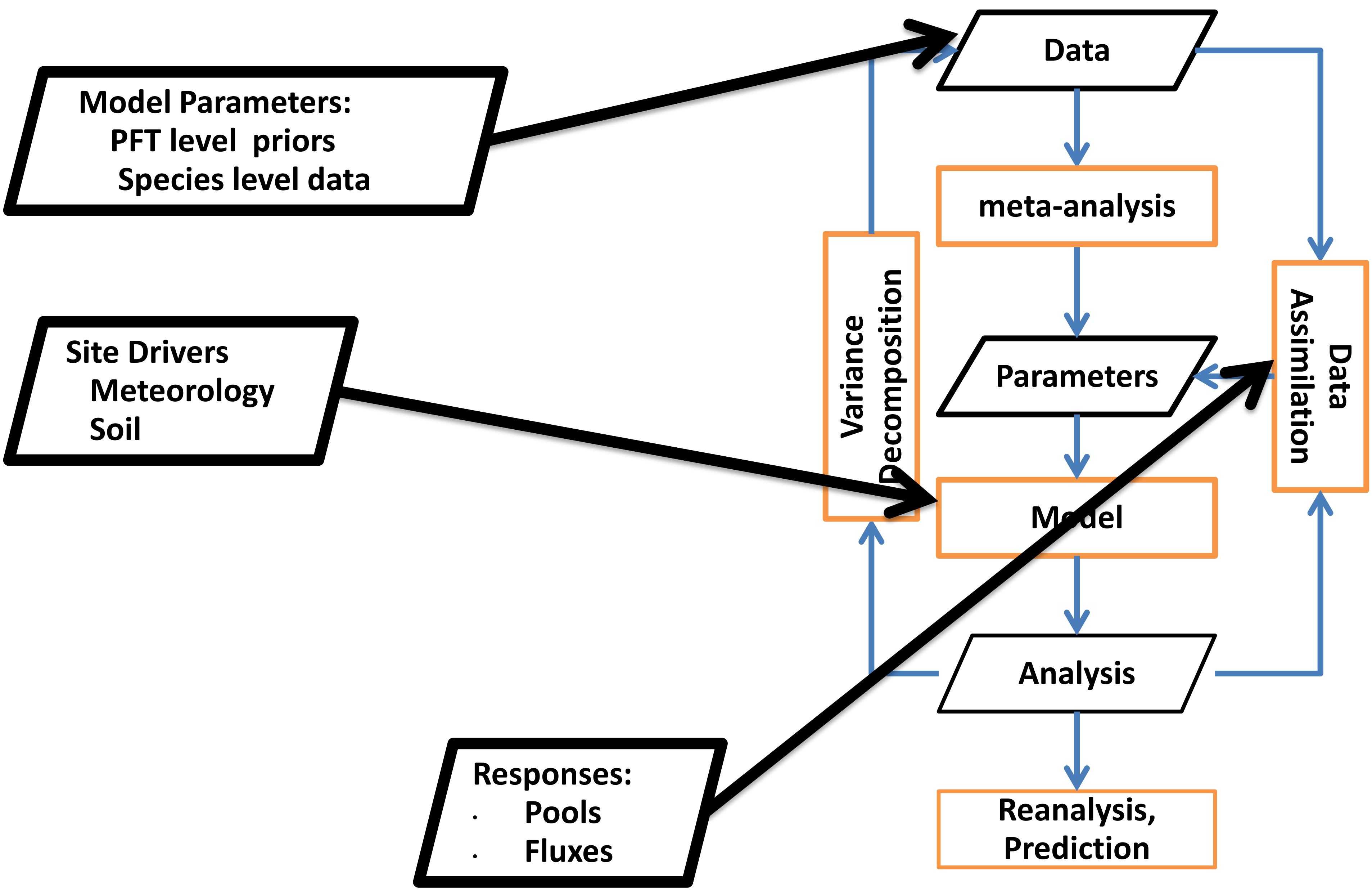


This Study

Probabilistic output
Computationally efficient
Multiple feedbacks to constrain uncertainty



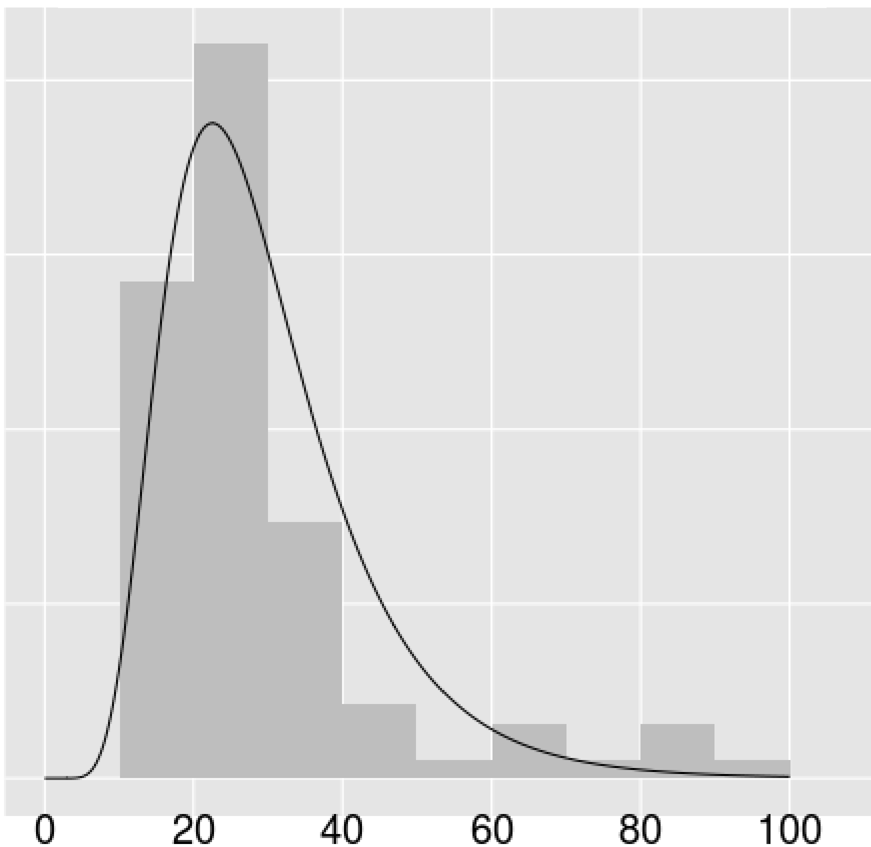
Data



Model Parameters:
PFT level priors
Species level data
Site Drivers
Meteorology
Soil
Responses:
Pools
Fluxes
Data
meta-analysis
Parameters
Model
Analysis
Reanalysis, Prediction
Variance Decomposition
Data Assimilation

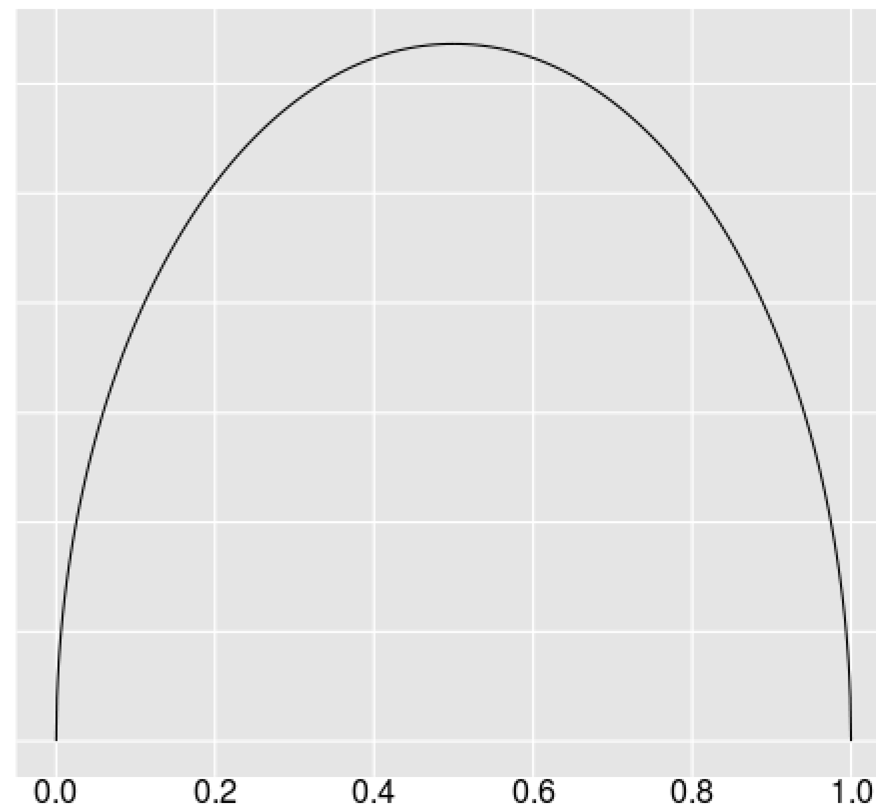
Meta-analysis: Priors for 17 Traits

Leaf C:N (grass)



good constraint:
data from GLOPNET
N=95
c2n~logN(3.3,0.4)

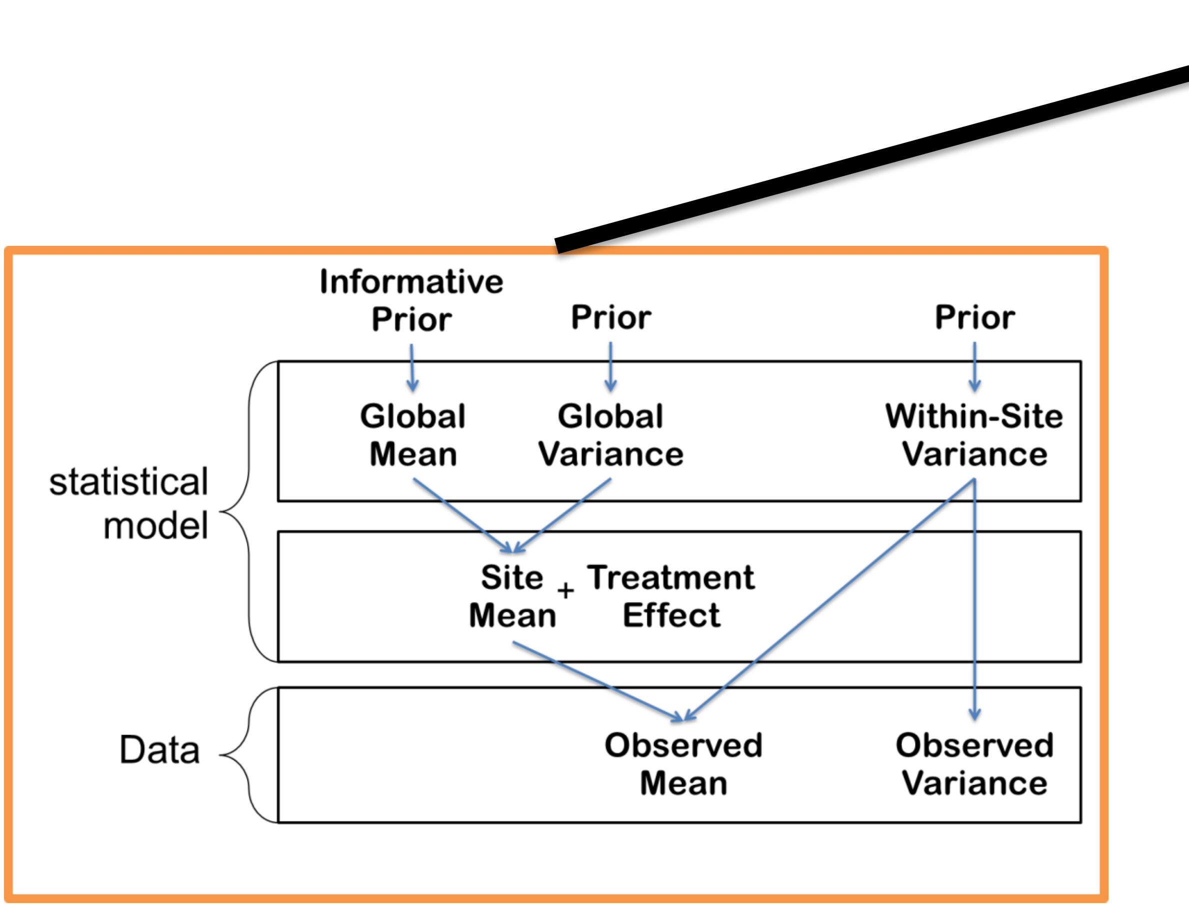
litter labile fraction

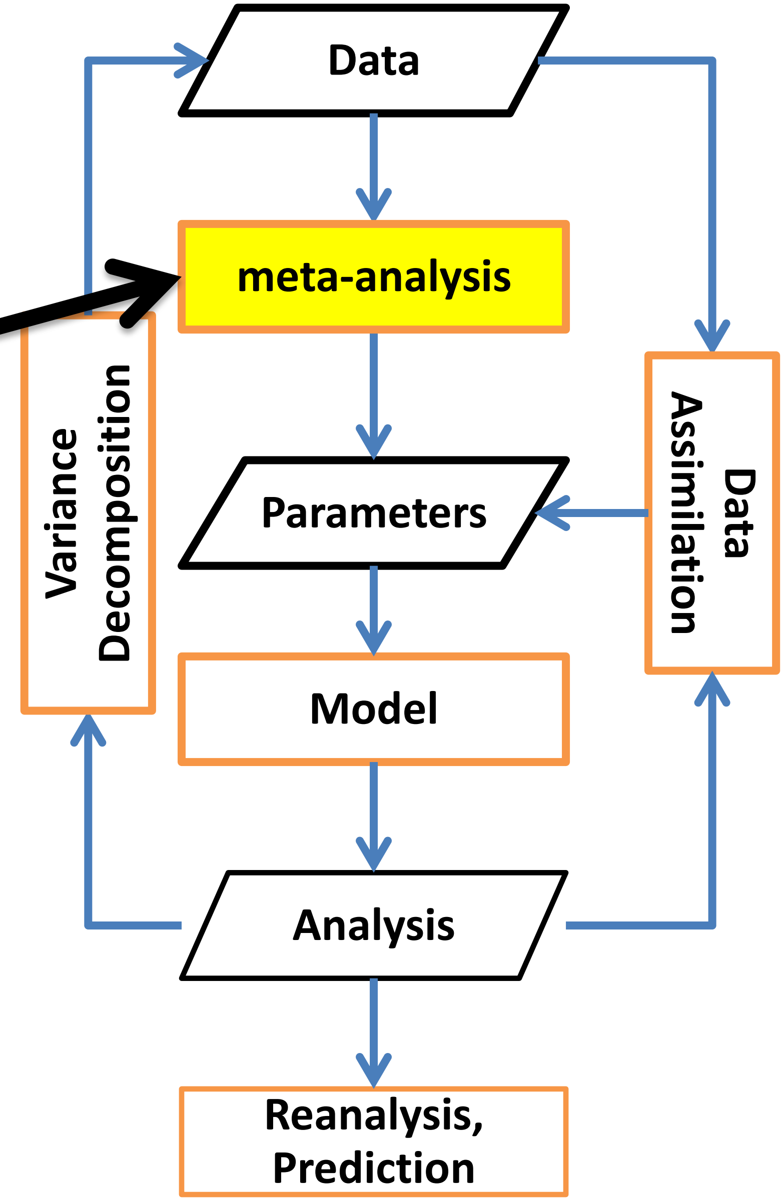


poor constraint:
f_{labile}~beta(1.5,1.5)

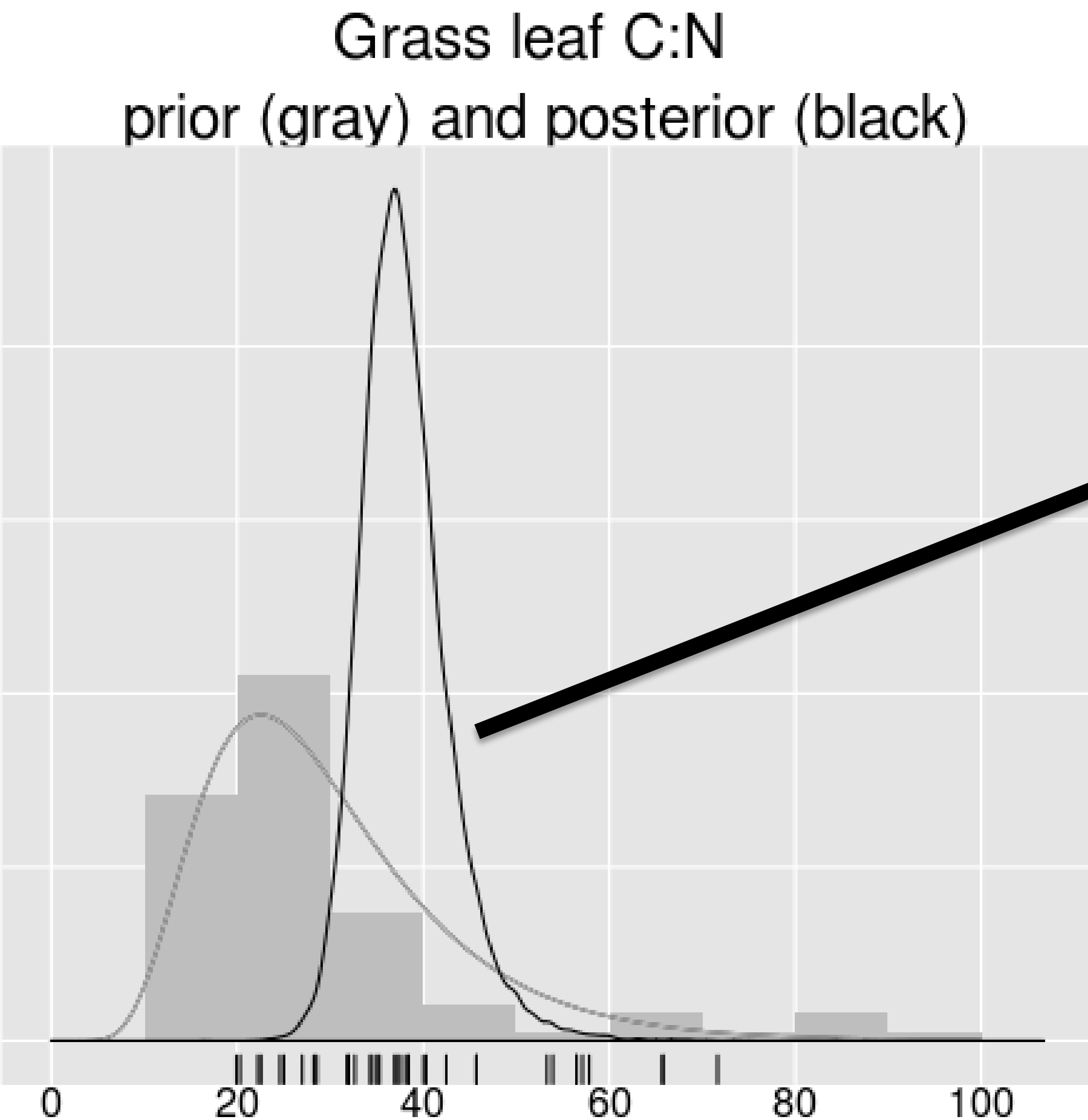
Meta-analysis Posteriors = Model Parameters

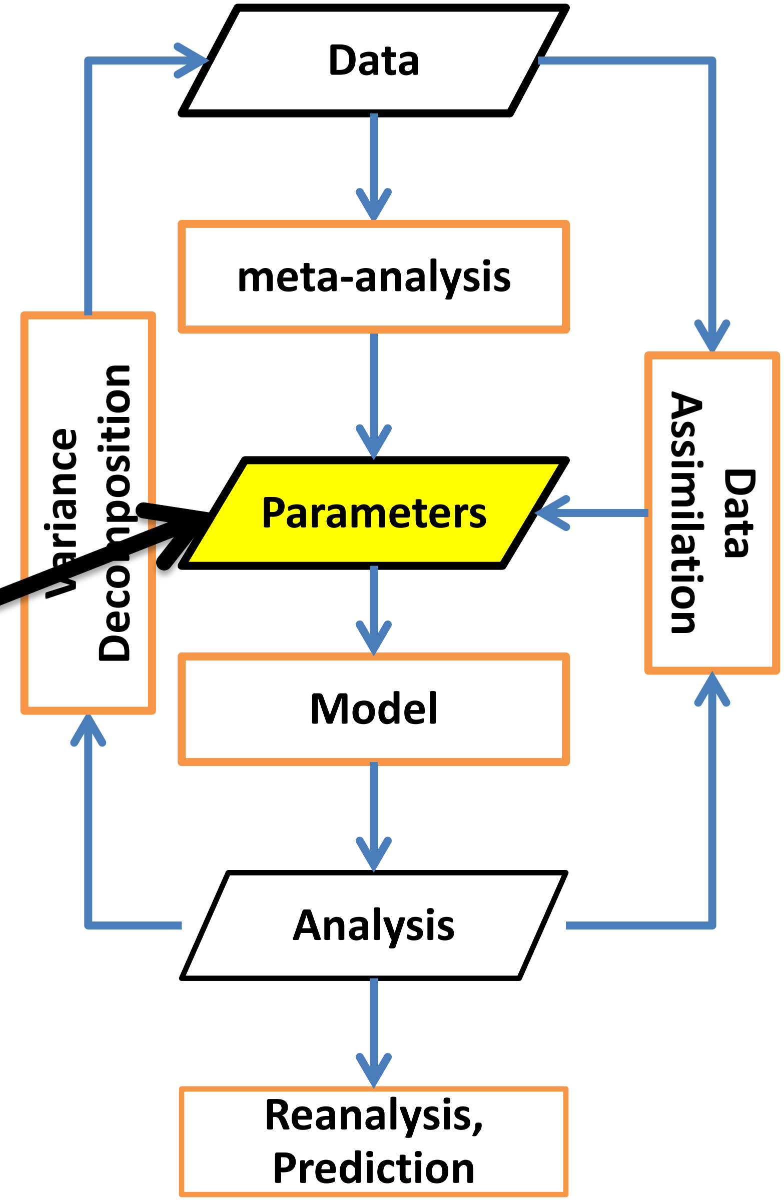
PDF of model parameter combines data with prior





Meta-analysis Posteriors & Model Parameters

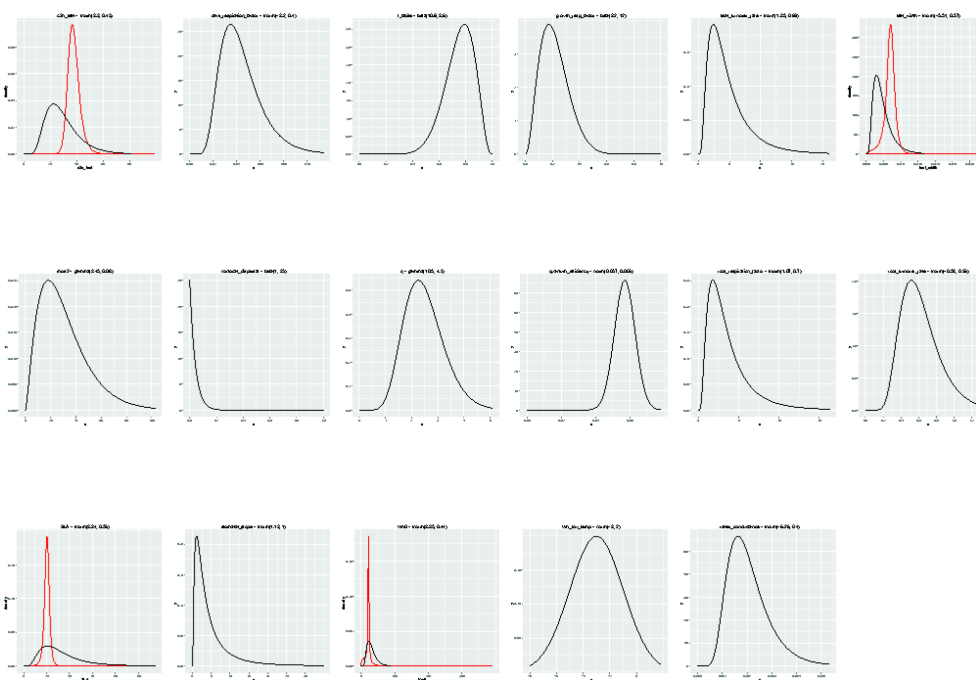




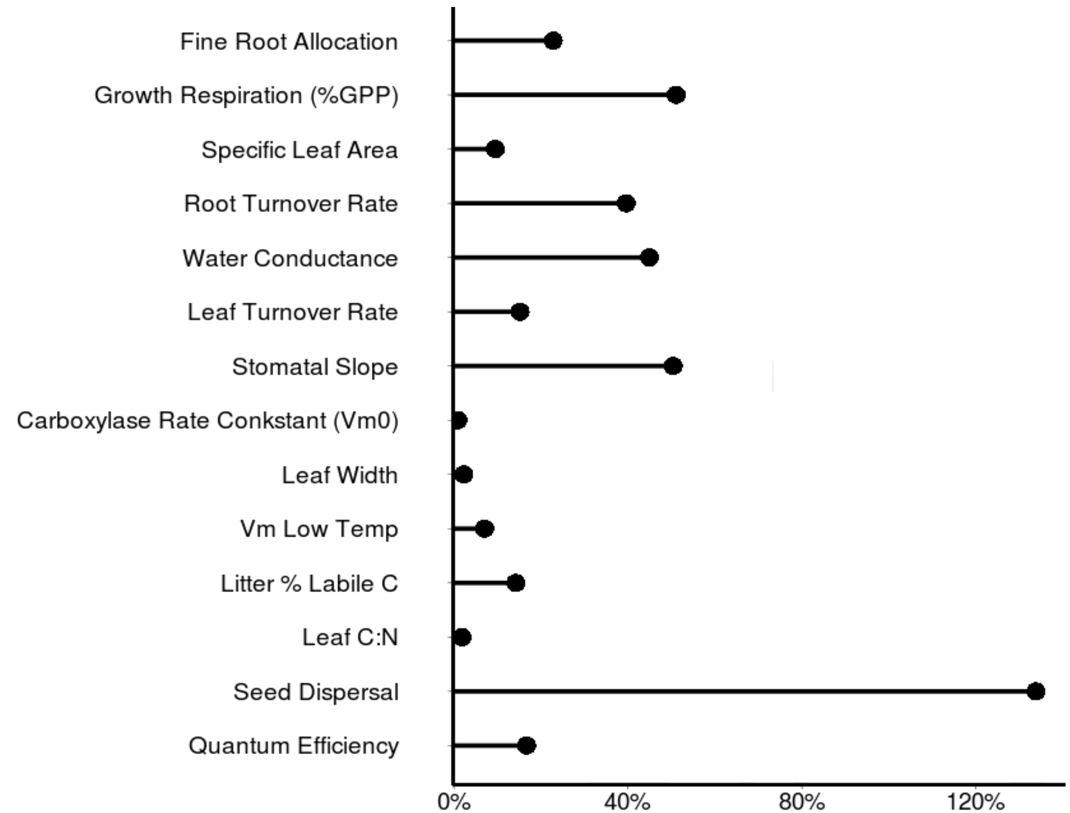
Meta-analysis Posteriors & Model Parameters

500-1000 parameter combinations
pseudo-random samples from PDFs used in model

Parameter PDFS



Parameter CVs

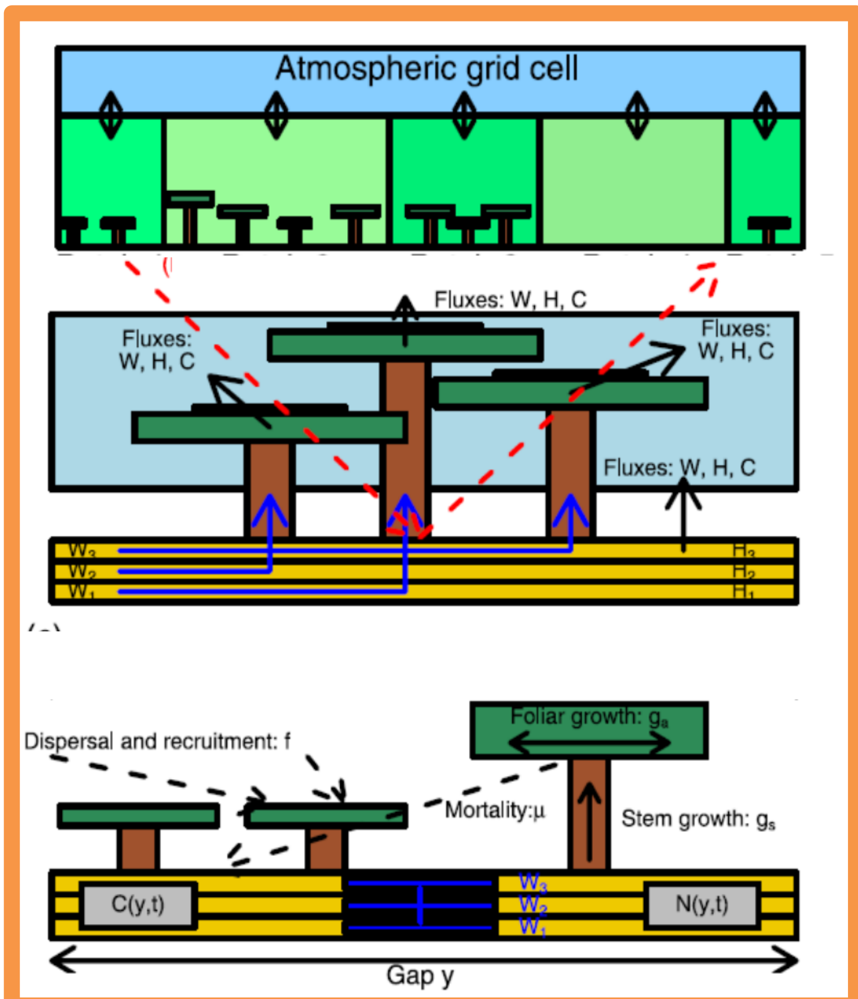


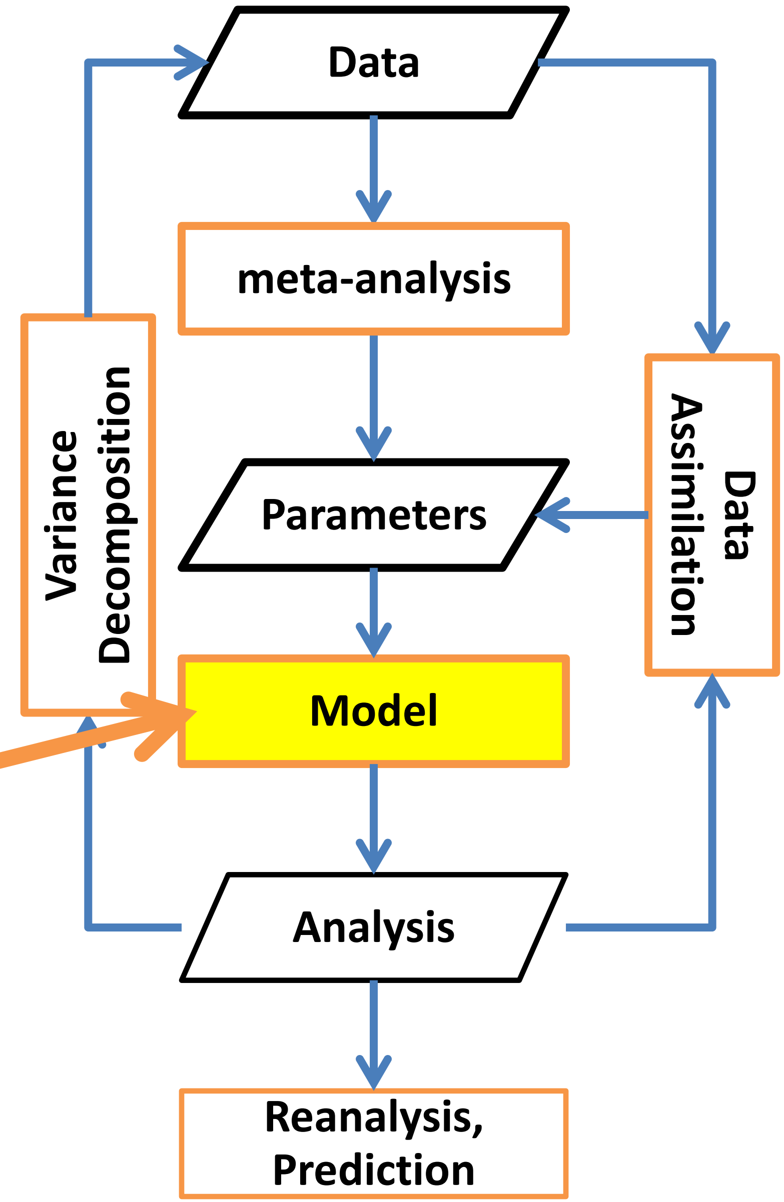
Ecosystem Demography Model (ED)

biophysical ecosystem model

ED2.2 = ED2.1 + Agriculture

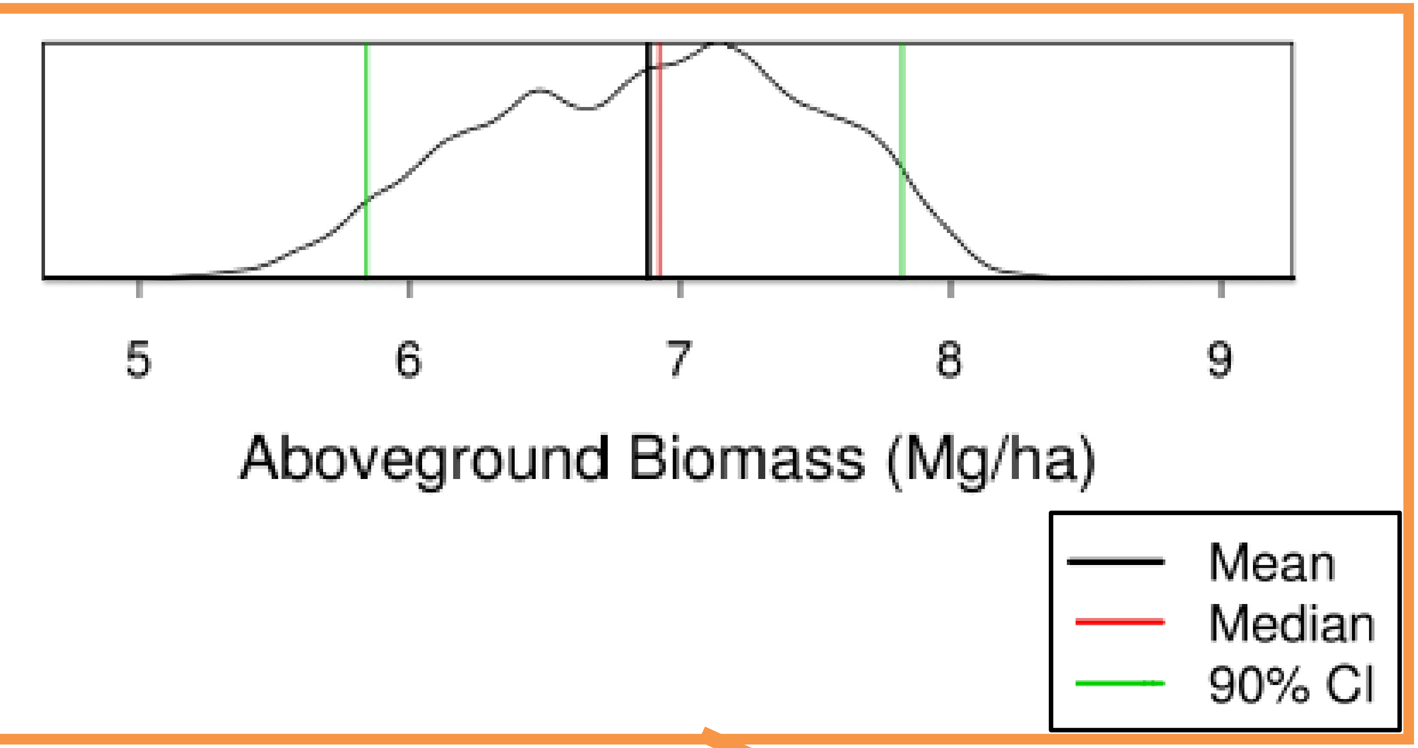
- -planting and harvest
- -dynamic allocation and phenology
- -fertilization

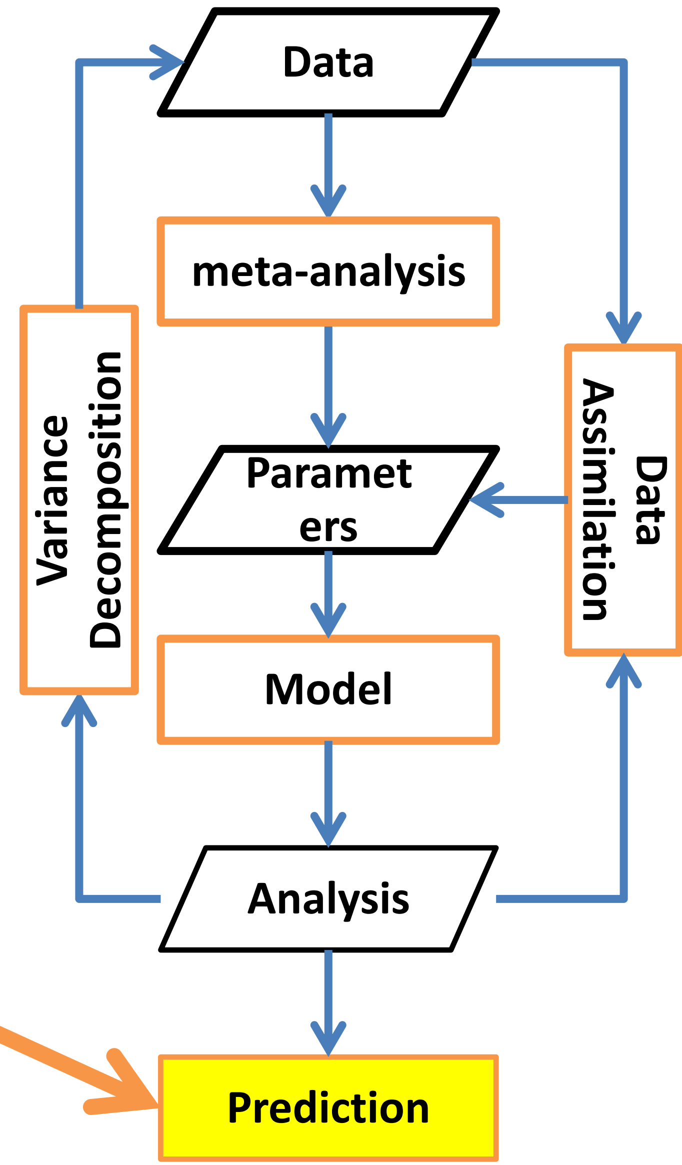




Moorcroft et al. 2001
Medvigy et al. 2009

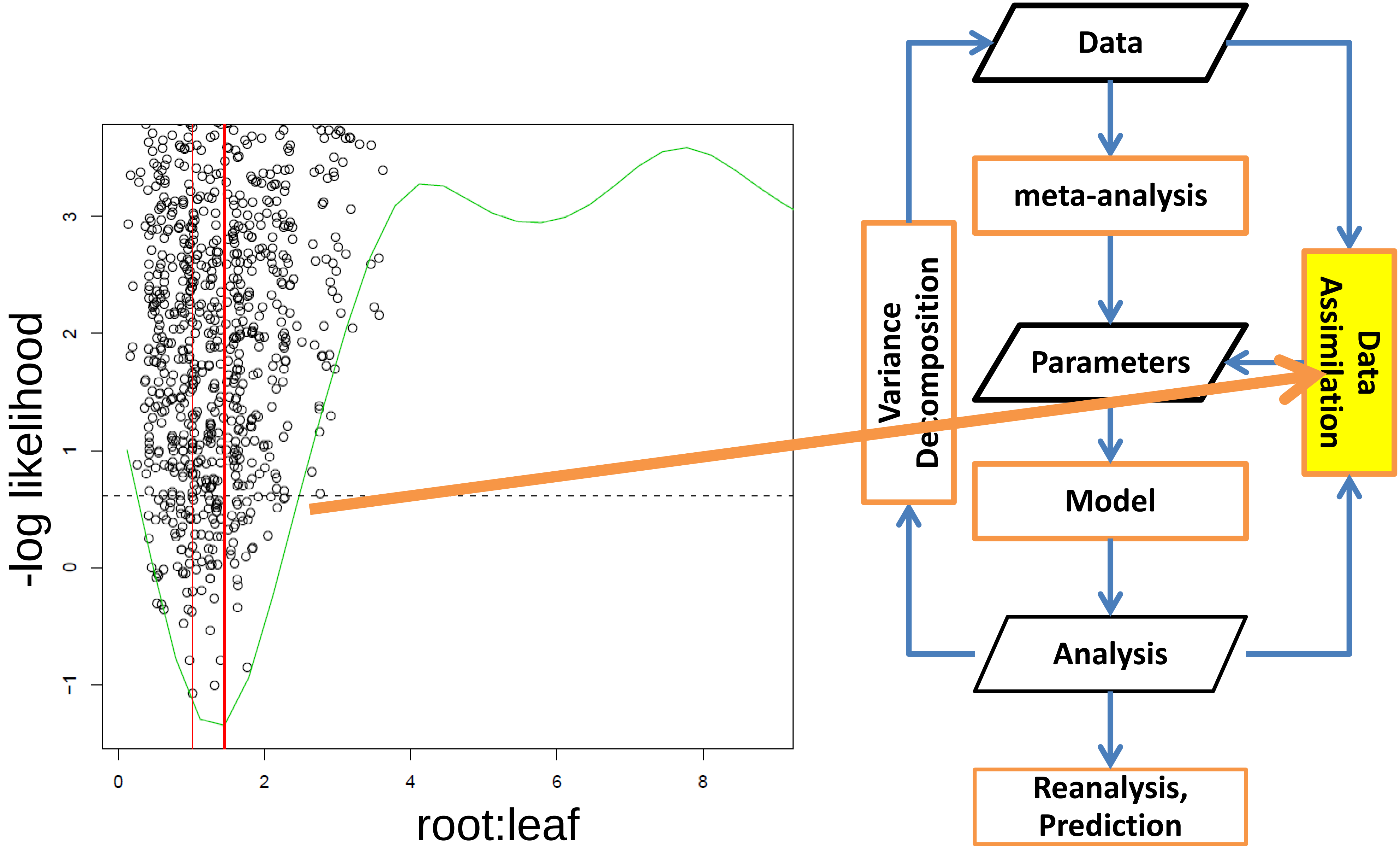
Model Ensemble Posterior



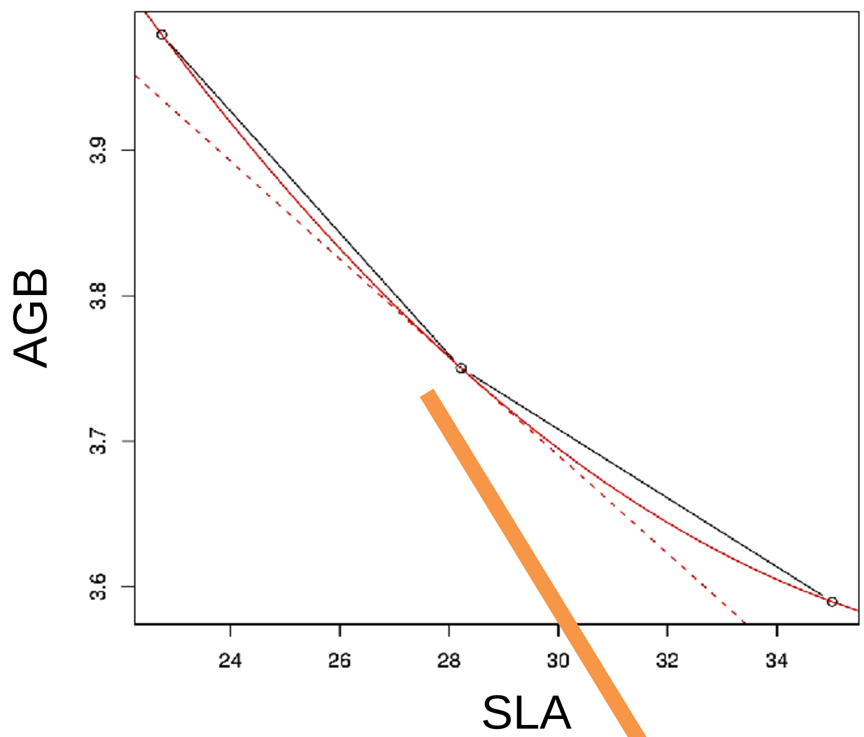


Data Assimilation

$\mathcal{L}$=f(Model output – Observation)



Variance Decomposition

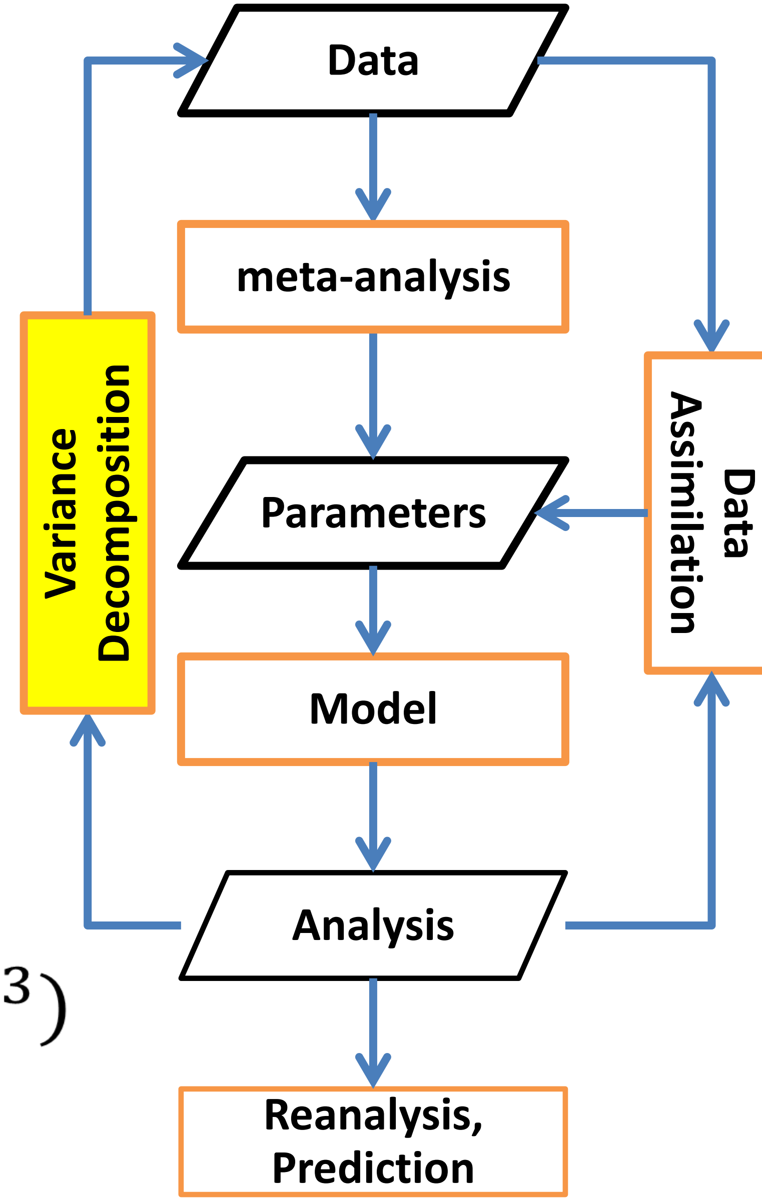


Ensemble Variance

Sensitivity Analysis

$$Var(Y) = \sum Var[\Theta_i] \left(\frac{dY}{d\Theta_i}\right)^2 + O(\Theta^3)$$

Posterior Parameter Variance

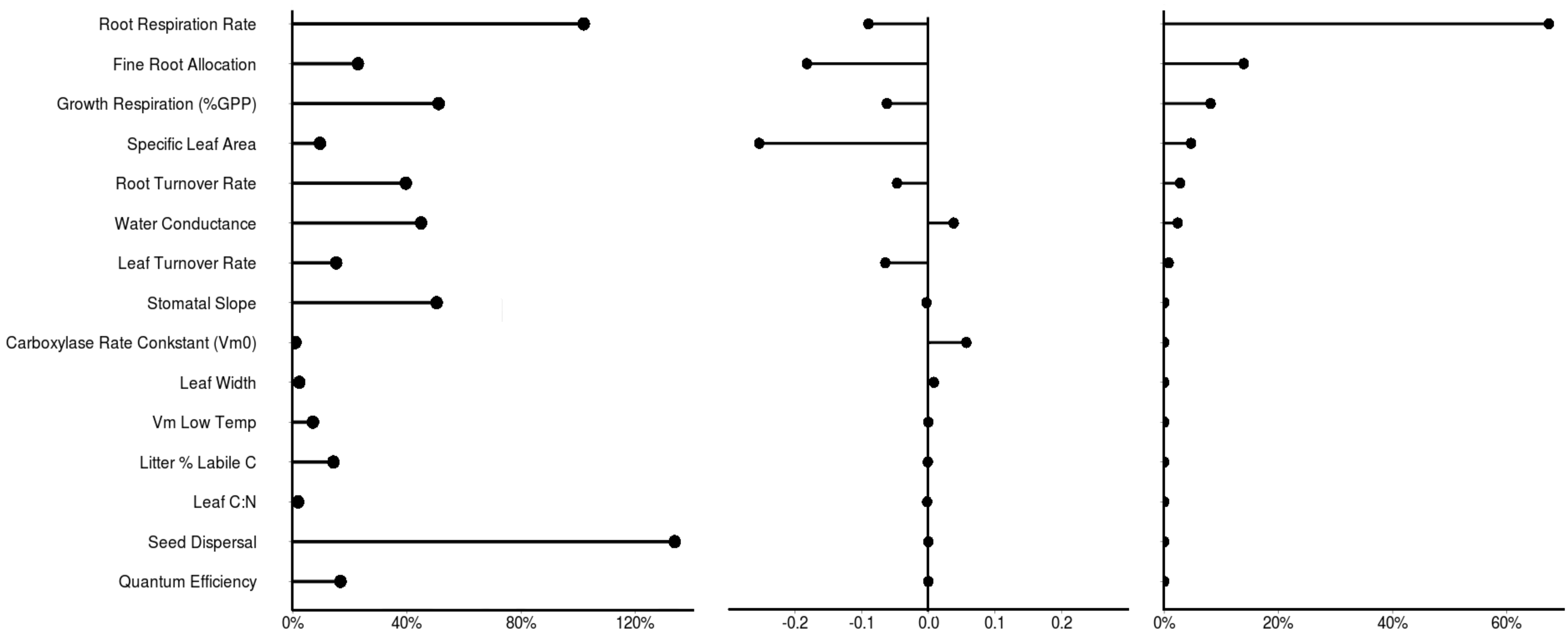


Variance Decomposition

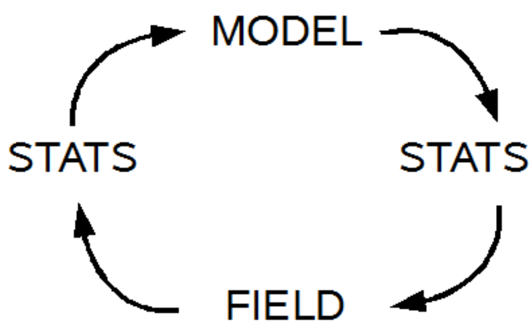
$$Var(Y) = \sum Var[\Theta_i]\left(\frac{dY}{d\Theta_i}\right)^2 + O(\Theta^3)$$

CV | Elasticity (Normalized Sensitivity) | % Variance

$\sqrt{Var(\Theta_i)}/\bar{\Theta}_i$ | $\frac{dY}{d\Theta_i}\Big/\frac{\bar{Y}}{\bar{\Theta}}$ | $Var(\Theta_i)\Big/\sum Var(\Theta)$



Model-data feedback

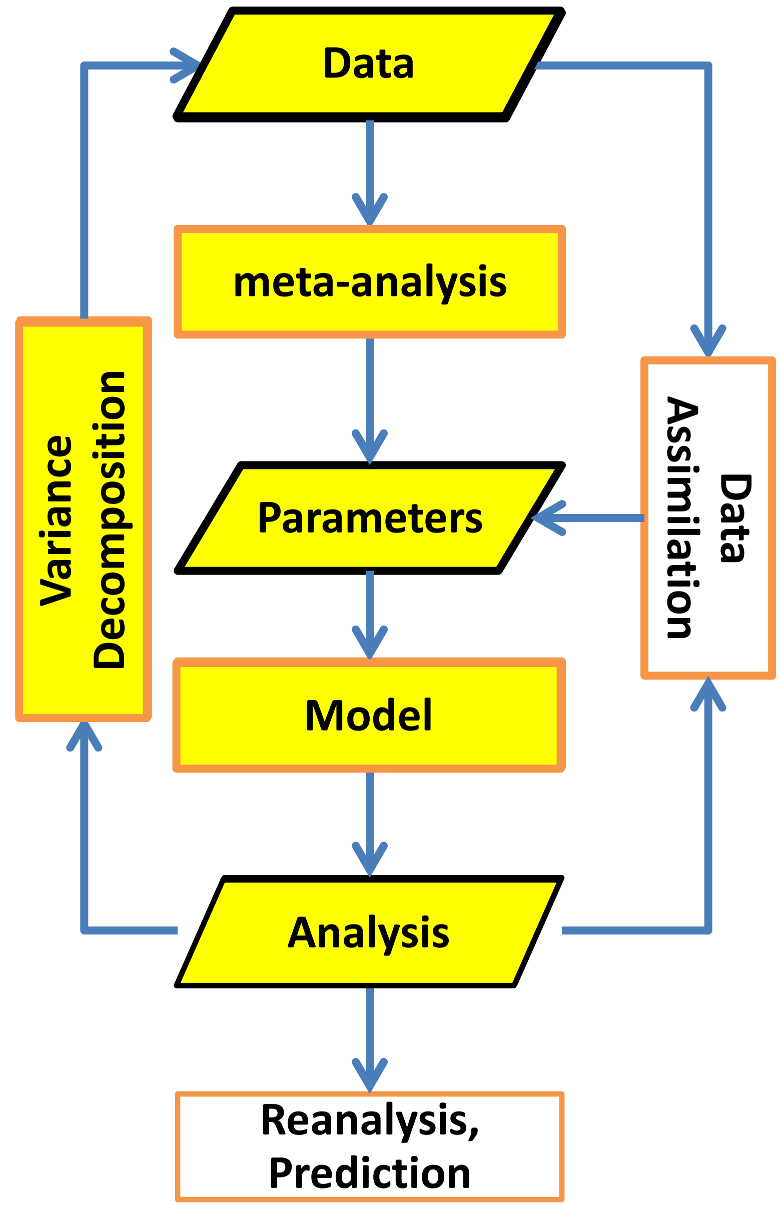


Reducing Uncertainty:

- given dY/dΘ, n, cost/sample
- identify target data

E.G. for switchgrass

- root respiration data top priority
- would need n=800 to reduce var(SLA) by ½
- for new species, SLA is 1st priority



Next steps

- Real time forecasting and data-assimilation
  - 10 day weather
  - 9 month climate
  - daily flux data
  - Yield, Spectral, other data as available
- Operational estimates of yield, other ecosystem functions
- Requires automation, thus PECAn

PECAn:
the Predictive Ecosystem C-cycle Analyzer

1. Workflow Manager
   - Directs computational steps
   - Automates and documents workflow
2. Applications
   - Biofuel crops
   - Temperate forests
   - Assimilation of data and models across scales - Fluxnet, NEON

Conclusions

- Integrate diverse available data into forecasts and reanalysis that reflect state of scientific understanding
- Uncertainty propagation and investigation
- Kriged Likelihood surface improves efficiency
- PECAn automates and documents computation
- PECAn makes modeling accessible and reproducible

Acknowledgements

Cyberintergrator Developers:
Jim Myers, Rob Kooper, NCSA

Webmaster: Patrick Mulrooney

Undergraduate Techs:
Andy Tu, Jenny Tang, Sarah Virani

Funding:

Energy Biosciences Institute