Calibrating soybean parameters in CLM-Crop using an MCMC approach

Beth Drewniak, Xiaoyan Zeng, and Emil Constantinescu

Work supported by DOE BER through the Climate Science for a Sustainable Energy Future (CSSEF) project

CLM-Crop Development

- The CLM-Crop model is a agricultural module that relies on a suite of parametric inputs that govern plant growth under a given atmospheric forcing and available resources. (Drewniak et al., 2013)
- CLM-Crop development used measurements of GPP, NEE, and carbon from AmeriFlux data to choose parameter values that optimize crop productivity in the model.

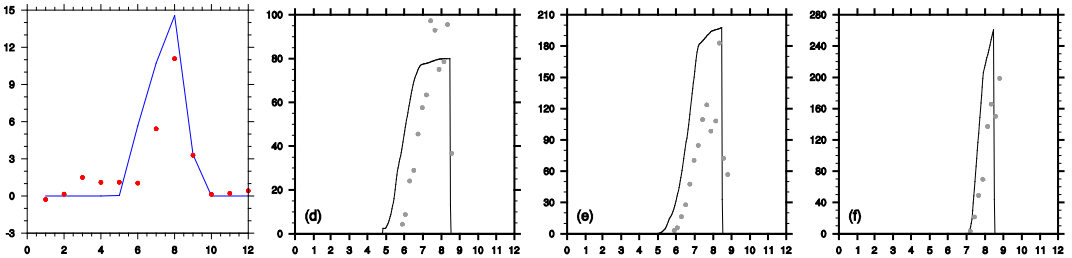
Photo courtesy of David Cook

- Components of CLM-Crop integrated into the CLM4.5
  - Separate organs pool
  - Fertilization
  - Soybean nitrogen fixation
  - Nitrogen retranslocation
  - New carbon nitrogen ratios

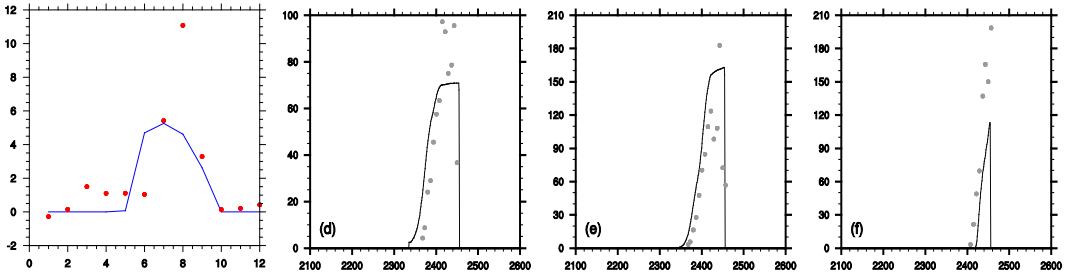
Motivation

- Differences between CLM3.5 and CLM4 crop models
- Major concern is crops are not as productive in CLM4
- Need to calibrate the parameters governing growth
- But over 100 parameters
  - Carbon nitrogen ratios :
    - Leaf (pre- and post-grain fill)
    - Stem (pre- and post-grain fill)
    - Organ
- Focus on soybean crop first
- Use data from Bondville, IL AmeriFlux site

GPP | Leaf Carbon | Stem Carbon | Organ Carbon



CLM3.5-Crop
vs.
CLM4.0-Crop



Calibration Strategy

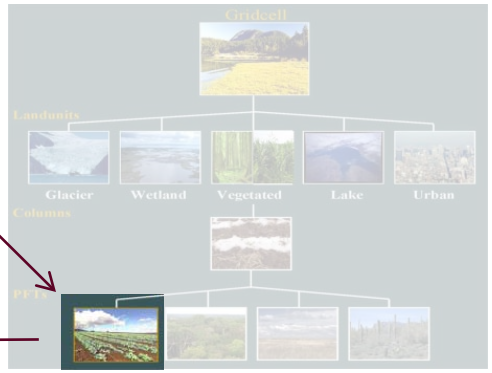
- Parallel Markov chain Monte Carlo (MCMC) strategy for the CLM-Crop model: multiple chains are run in parallel and exchange information as they evolve

- CLM-Crop model calibration setup:

parameters: θ
C, N ratios/plant

spin-up:
C, N, litter pools

observables: y
NEE, GPP, plant C

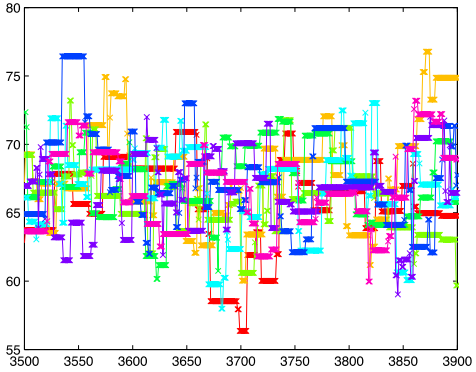


sample posterior distribution for model parameters

$$\pi(\theta|y) \propto \pi(y|\theta)\pi(\theta)$$

parallel MCMC chains

evolution of live stem C/N value in 8 Markov chains



- Performed soy carbon nitrogen ratio calibration; explored C, N & litter pools as calibration parameters

MCMC Implementation

- Parallel adaptive MCMC strategy to determine the posterior distribution of parameters given observations
- Assume prior and likelihood:

$$\log(\pi(\theta)) = -\frac{1}{2}(\theta - \bar{\theta})^T \Sigma_\theta^{-1} (\theta - \bar{\theta})$$

$$\log(\pi(y|\theta)) = -\frac{1}{2}(F(\theta) - H(y))^T \Sigma_{\text{obs}}^{-1} (F(\theta) - H(y))$$

- compute posterior using Bayes rule: $\pi(\theta|y) \propto \pi(y|\theta)\pi(\theta)$

- use a parallel MCMC implementation to accelerate and diagnose the convergence by running several Markov chains in parallel while adjusting a Gaussian proposal distribution according to their spread

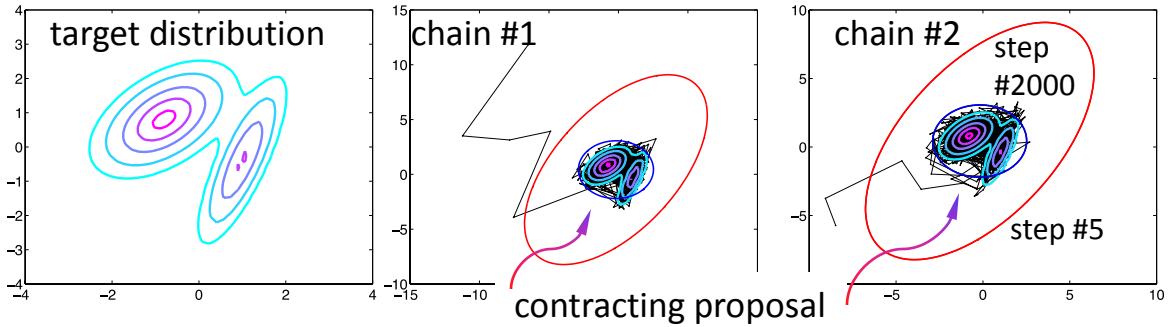
Standard MCMC algorithm:

Given the target density $\pi(x)$, pick a symmetric **proposal density** $Q(x'|x_t)$, start with x_0, $t = 0$ then proceed as follows:

1. generate a proposed new sample value from $Q(x'|x_t)$
2. calculate $\alpha = \frac{\pi(x')}{\pi(x_t)}$
3. pick u from uniform 0 and 1, and if $u < \alpha$ set $x_{t+1} = x'$, else set $x_{t+1} = x_t$; set $t = t + 1$, go to 1.

Validation of the Calibration Strategy

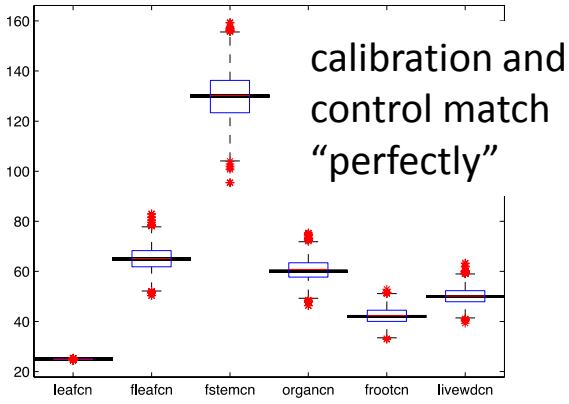
- Simple synthetic example:
  - proposal is auto-tuned
  - chains are run in parallel
  - *[Craiu et al 2009]; [Zeng et al 2013]*



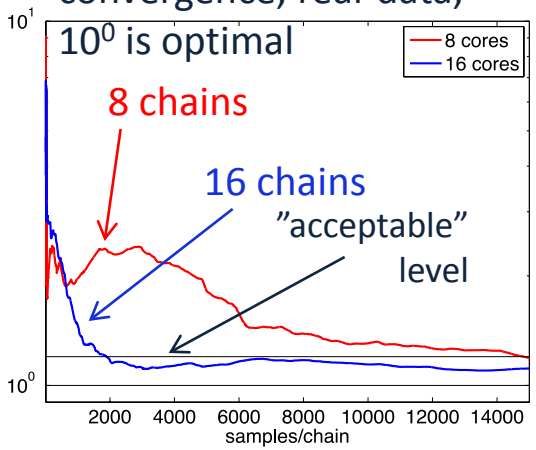
- CLM-Crop: complex, relatively inexpensive:
  2 min/simulated year

- Parallel chains:
  - Convergence is accelerated by adapting the proposal
  - Diagnostics are obtained from (inter-/intra- chain variance)
  - *[Zeng et al 2013]*

twin experiment

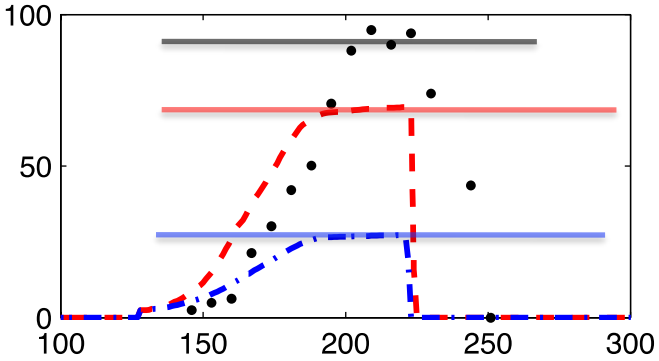


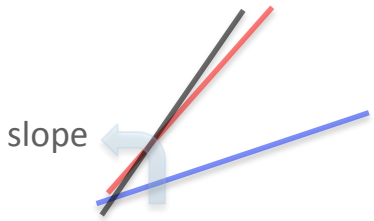
convergence, real-data,



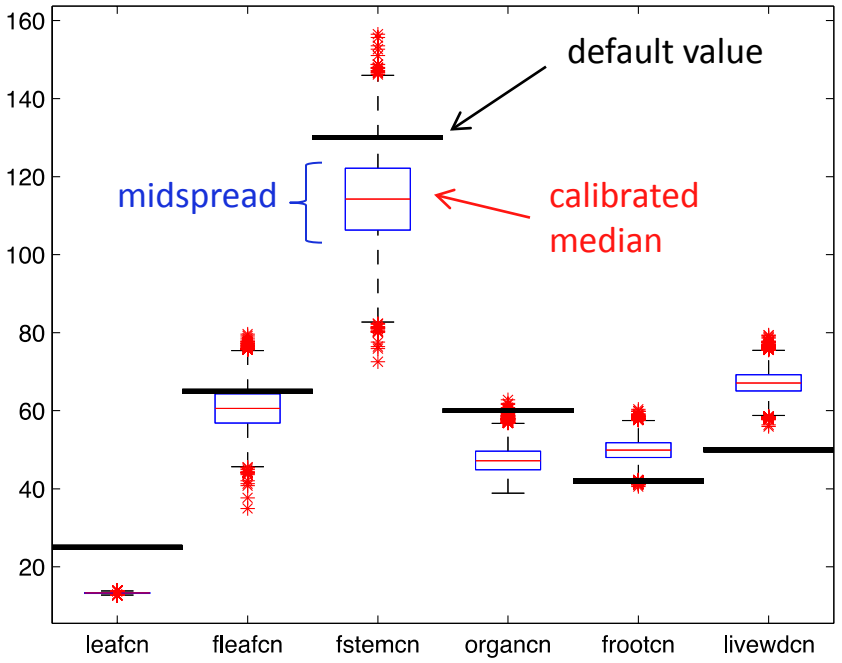
CLM-Crop Model Calibration: Results using Bondville site observations

- Preliminary calibration results for soy carbon nitrogen ratio with validation
- Performance metric: match the filtered peak and slope during emergence



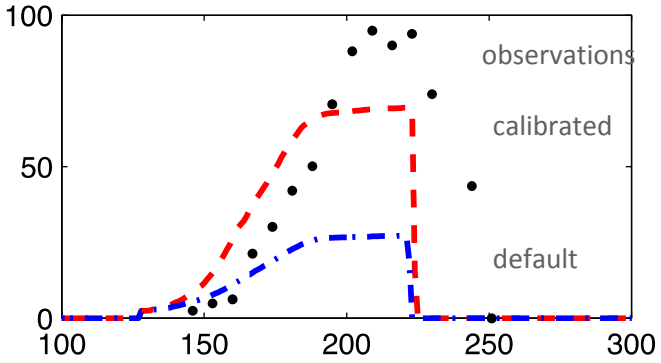


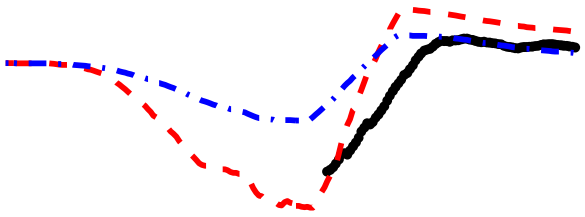
Default and calibrated parameter values



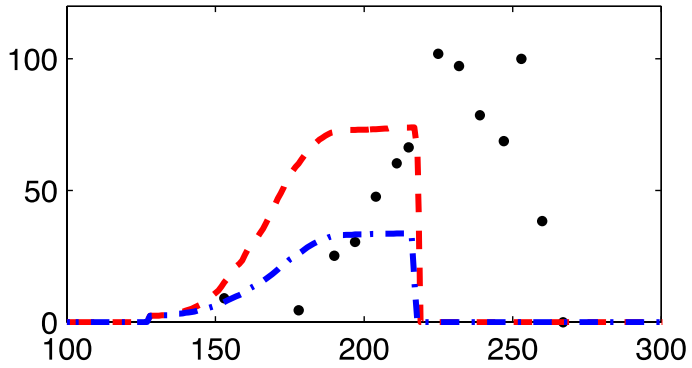
- Obtained new calibrated parameters for soy that give significantly better fit with data

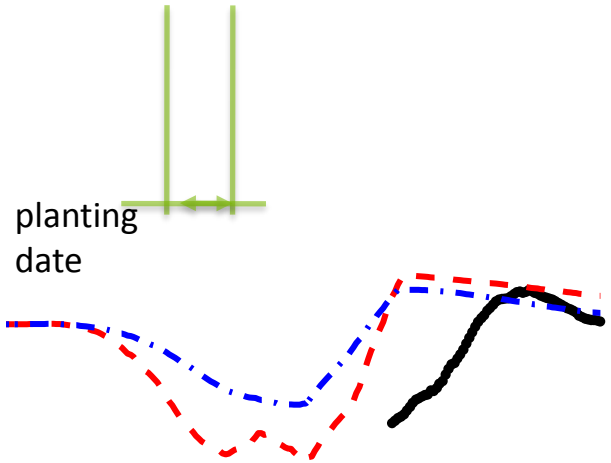
Calibration results: soybean (2004)





Validation results: soybean (2002)





Moving Forward

- Need to expand
  - Number of parameters
  - Crop types
- Computationally expansive - takes too long to converge
- Work on developing surrogate models for parameter calibration

Photo http://landcovertrends.usgs.gov/east/eco70Report.html

This work was supported by the Office of Biological and Environmental Research, U.S. Department of Energy, under contract KP-17-03-02-0 through the Climate Science for a Sustainable Energy Future (CSSEF) project.