Crop Parameter Evaluation in CLM-Crop

Xiaoyan Zeng, Beth Drewniak, Azamat Mametjanov, Mihai Anitescu, Paul Hovland, Boyana Norris, and Veerabhadra R. Kotamarthi

Motivation

- Calibration and uncertainty quantification of CLM-Crop
- Initially focus on six parameters:
  - Leaf initial and final carbon nitrogen ratios (leafcn and fleafcn)
  - Stem initial and final carbon nitrogen ratios (stemcn and fstemcn)
  - Fine root carbon nitrogen ratios (frootcn)
  - Organ carbon nitrogen ratios (organcn)
- Two approaches – Intrusive uncertainty and Automatic Differentiation (AD)
- Calibration performed using PTCLM at one location: AmeriFlux site at Bondville, IL
  - Corn and soybean rotation
  - Carbon measurements of various plant components
- Additional runs were performed after a 5% perturbation (+/-) for each parameter

Maximum Likelihood (Optimization) Approach

- The Observation Model:

  $o(t_i) = F(\boldsymbol{a}, t_i) + \epsilon_i$

- Solve the Optimization Problem:

  $$\min_{a} \phi(\boldsymbol{a}) := \sum_{i=1}^{N} \big(o(t_i) - F(\boldsymbol{a}, t_i)\big)^T \big(o(t_i) - F(\boldsymbol{a}, t_i)\big)$$

- A complete likelihood based UQ approach:

1. To solve the maximum likelihood problem and determine best guess parameter a* we need:
   - The gradient of "φ". A simple way is with finite divided difference: perturb $\boldsymbol{a}$, rerun the model and compute the ratios.
   - Gradient-based optimization method to solve the optimization problem (BFGS)
2. Compute the informational matrix at the best guess parameter $\boldsymbol{a}$ * – use divided difference for approximating Hessian.
3. Produce uncertainty bounds for parameter $\boldsymbol{a}$ based on confidence intervals.

Simplifications

- Approximate F with a linear function of $\boldsymbol{a}$
- Observation data modeled as:

$$o_l(t_i) = \sum_{j=0}^{p} \beta_{l,j} x_j(t_i) + \epsilon_{l,i}, l = 1, \dots n$$

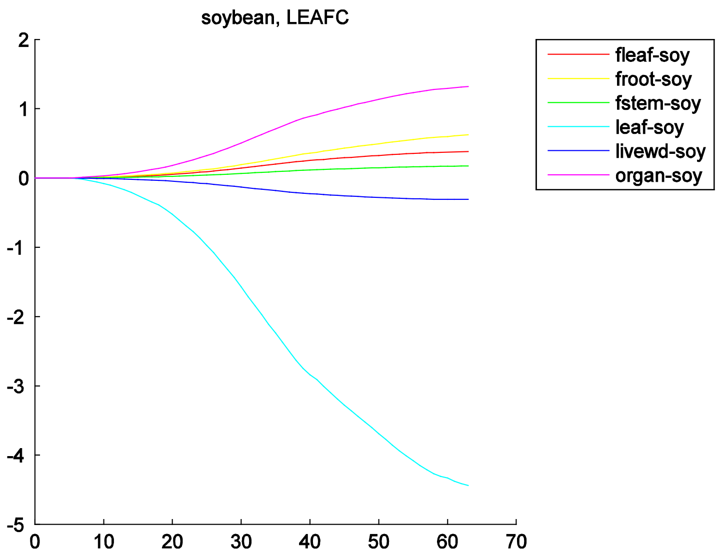
- Using initial guess of parameters $\boldsymbol{a}$, we take the first order truncation of the Taylor expansion of F and compute the finite difference using our perturbation runs, and after some math (which I will skip for everyone's sake), the new optimization to solve is:

$$\min_{a} \frac{1}{M} \sum_{j=1}^{M} \sum_{l=1}^{n} (\hat{f}_l(a, t'_j) - h_l(t'_j))^2$$

The minimizer is the zero solution of the gradient, found with least squares

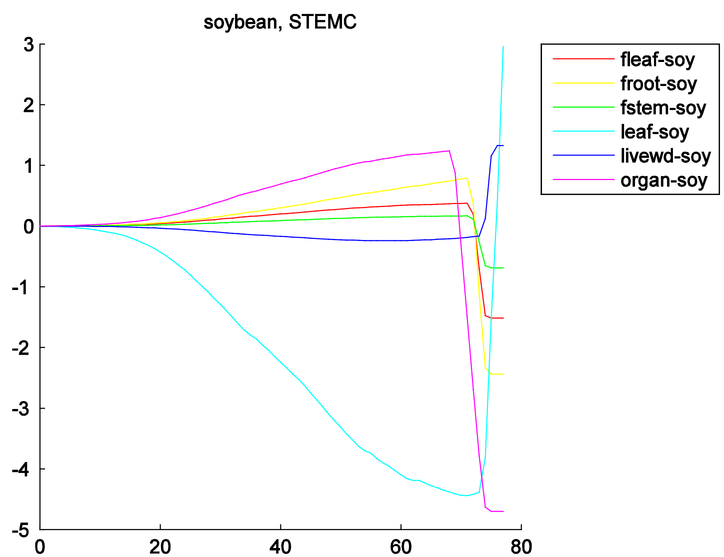
So, results…

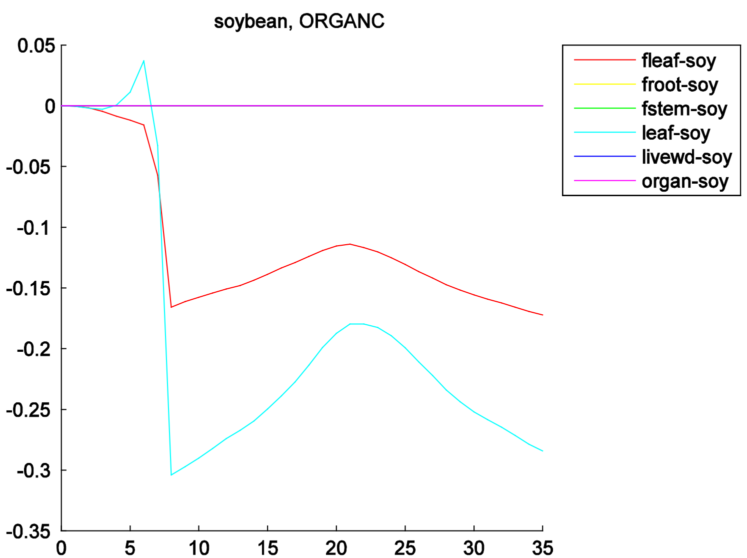
Finite Difference of Carbon



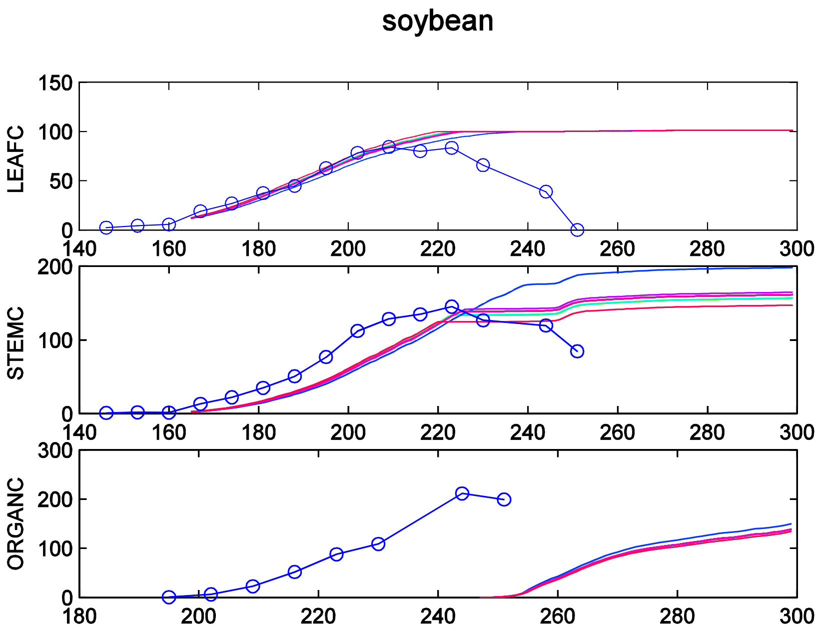
Derivatives with respect to leaf carbon and stem carbon vary most with leafcn and organcn.

Dervitives with respect to organ carbon vary most with leafcn and fleafcn.



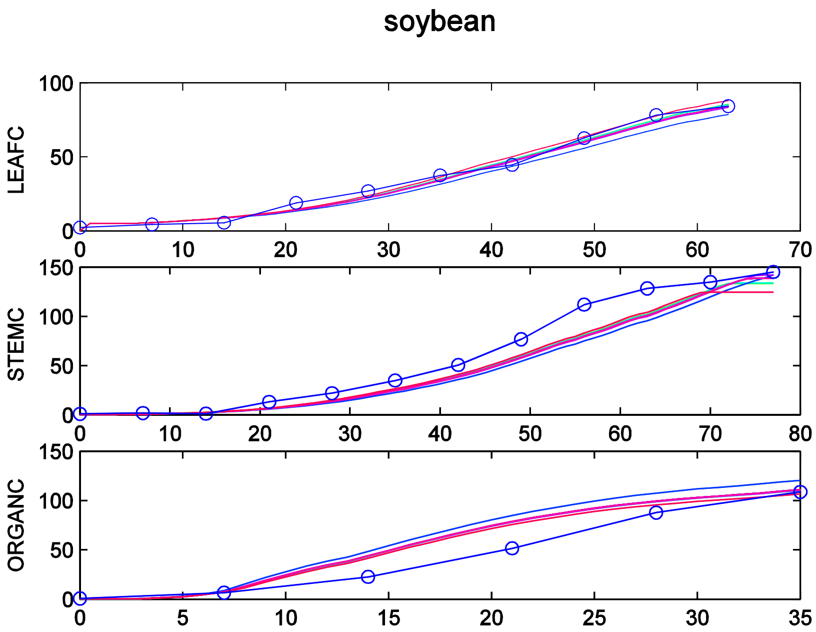


Data vs. Model

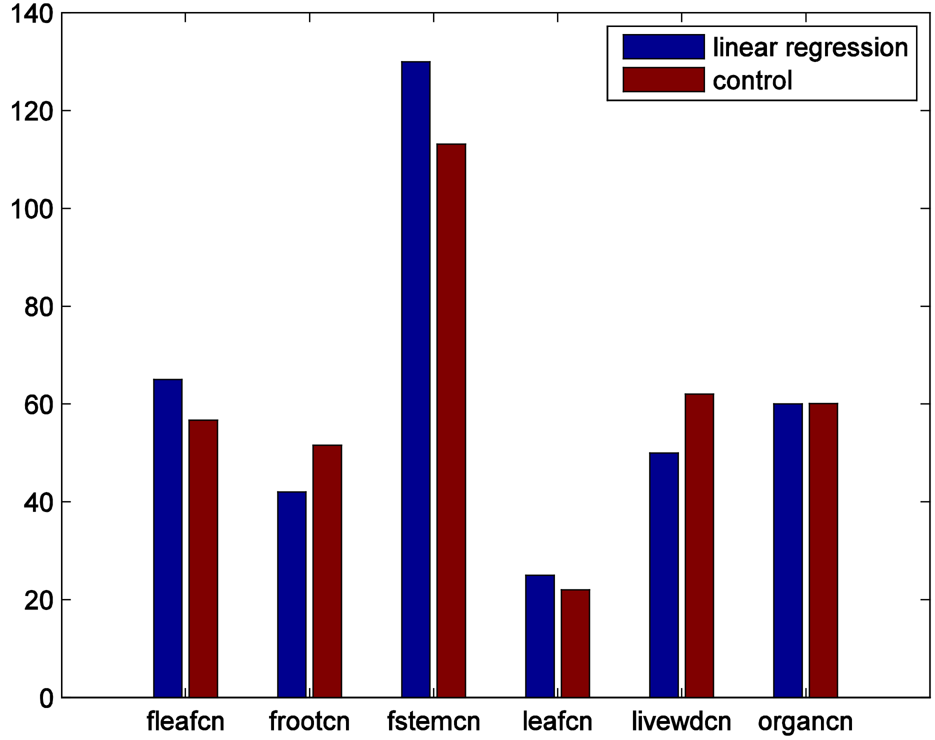


Raw data: Gaps between observations (circled line) and the model output (solid lines), most likely from planting date and harvest timing.

Massaged data: Line up observations with model and apply a second-order polynomial fit.

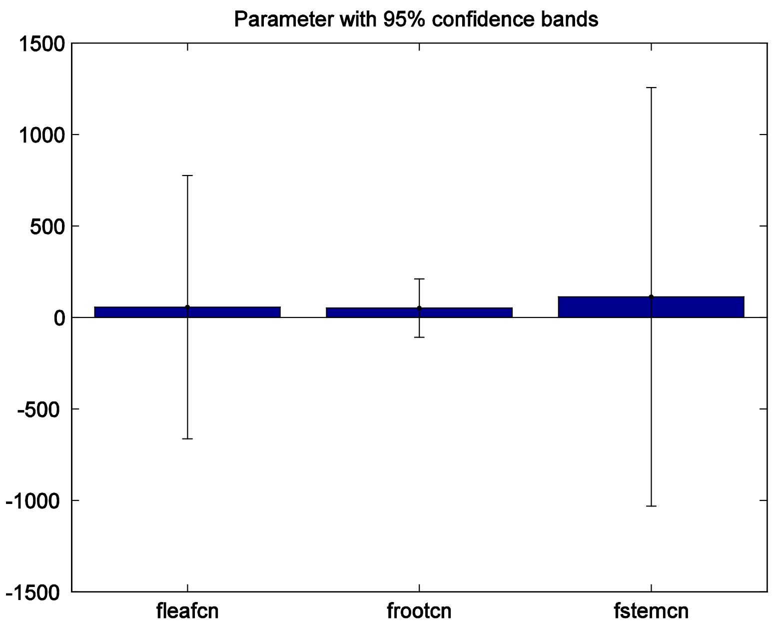


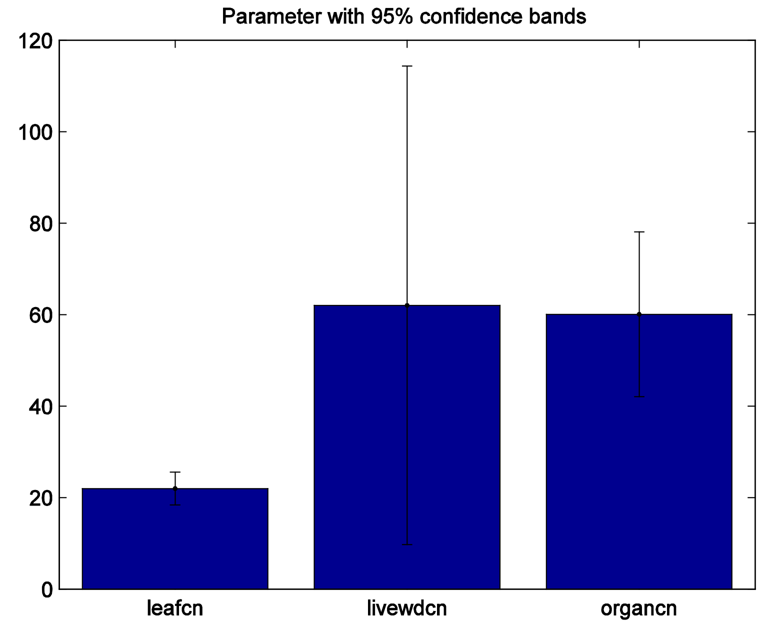
Best Guess Parameters



Looks good, the initial guess is close to the calibrated parameters…but

Uncertainty of parameters

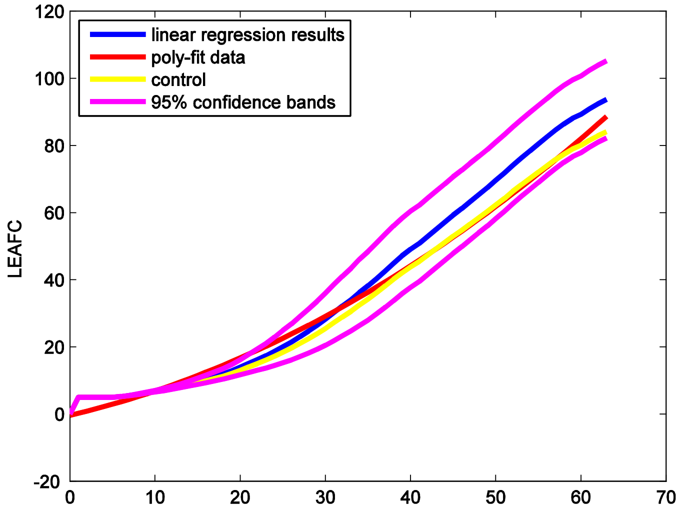


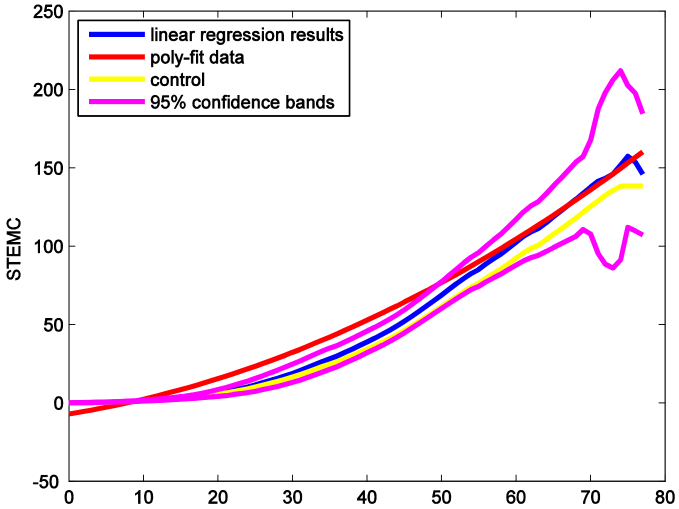


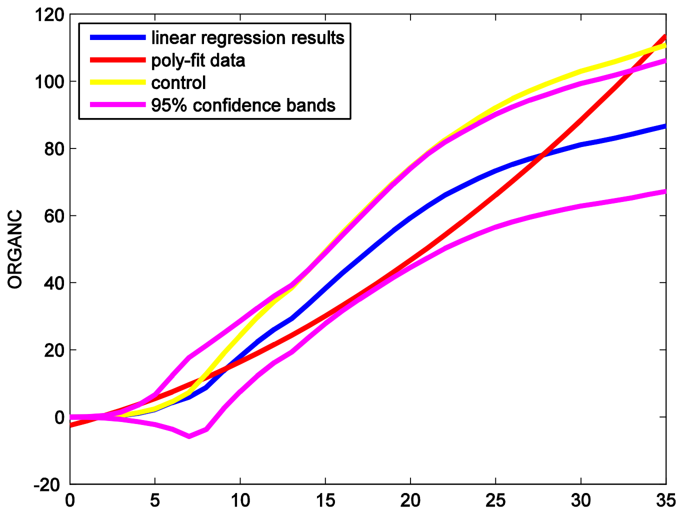
Ouch, 95% confidence bands fall in negative zone – more effort needed to fix these!

Better, these parameters have more reasonable uncertainty estimates based on the data.

Propagation







Most of the observations fit within the 95% confidence intervals (not fully tested yet).

What does this mean?
The high uncertainty in the model parameters doesn't influence the output very much.

Automatic Differentiation

- Automatic Differentiation Approach
  - View model as composition of elemental differentiable functions
  - Compute derivatives of elementals and propagate them using chain rule
  - Compute gradients to determine sensitive parameters
- Status
  - Too difficult to perform AD on entire model (at least for now), so settle for performing AD on a portion of the model
  - Can get the derivative of the fluxes with a portion of the state vector and parameters, which can be used in the max likelihood approach to determine gradient and compare with finite difference method
  - Have obtained derivatives of the carbon and nitrogen fluxes by performing AD on the module responsible for calculating the carbon and nitrogen allocation
  - Also have obtained derivatives of the state update functions

Derivative results from AD approach

	CORN		WHEAT		SOY	
	C	N	C	N	C	N
LEAF						
f lea f cn	7.0353917	-93.030506	5.0544004	-100.241099	3.2190047	-59.18986
f rootcn	4.6136744	-46.257848	-10.0406592	-72.813487	2.7559661	-22.607132
f stemcn	1.9315305	0.1931531	2.679161	0.1786107	0.6794228	0.0271769
STEM						
f lea f cn	14.455417	0.2891083	57.2387074	1.1447741	34.5605004	0.69121
f rootcn	9.4894822	-12.986593	-116.169733	-35.86494	26.2942184	-20.578508
f stemcn	3.9686602	-25.901916	30.3402387	-52.7061099	7.2945506	-12.37426
ORGAN						
f lea f cn	1277.0877	25.541754	949.694509	23.7423627	589.0504	9.8175067
f rootcn	809.026384	16.180528	-1938.70403	-48.4676006	436.621367	7.2770228
f stemcn	350.617841	7.0123568	503.399874	12.5849969	124.32858	2.072143

Ongoing Work

- Solve the full nonlinear model
  - Use AD for derivative input to intrusive model
- Continue to improve crop model, test with new best guess parameters
- Evaluate other crop parameters (including those that aren't crop specific)