Bayesian calibration of the Community Land Model using surrogates

J. Ray[1], K. Sargsyan[1], Z. Hou[2] and M. Huang[2]
jairay [at] sandia [dot] gov

[1]Sandia National Laboratories, Livermore, CA,
[2]Pacific Northwest National Laboratory, Richland, WA

Funded by DoE/Office of Science
SAND2012-10275C

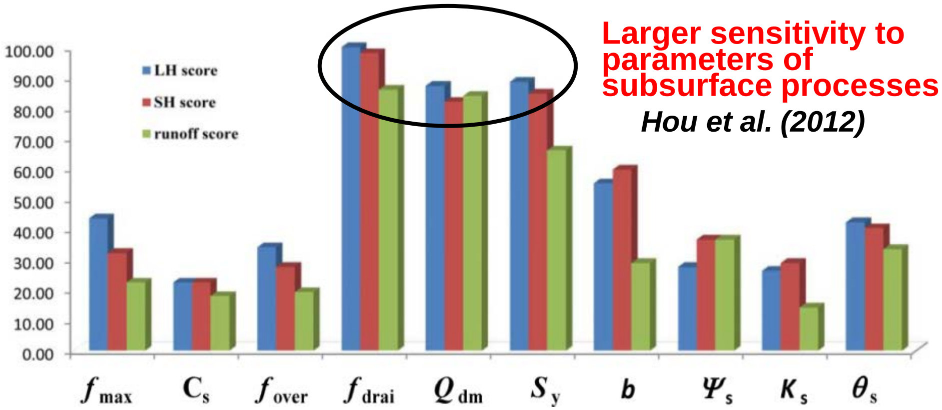
Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.
PNNL is operated for the US DOE by Battelle Memorial Institute under Contract DE-AC06-76RLO1830

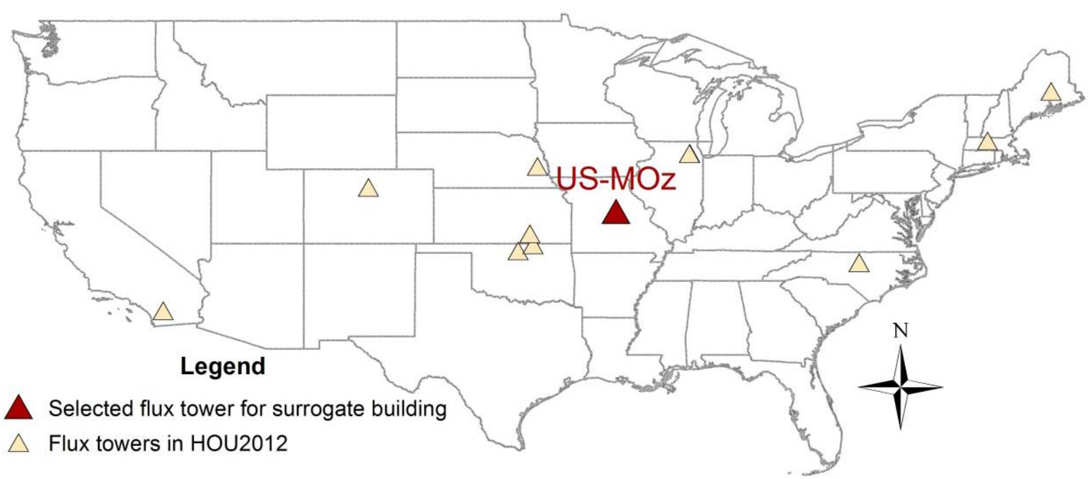
What is this talk about?

- Aim: Can the calibration of CLM be achieved using surrogates?
  - With quantified uncertainty
- Difficulty:
  - Bayesian calibration require 10^3 runs of CLM4SP, and 10^4-10^5 runs of CLM4CN;
  - One obvious solution: a quick-running surrogate model of CLM
- Technical challenge:
  - What does the model look like?
  - How many CLM runs does it require to make the model?
  - Are there complexities in the calibration? Artifact of the surrogate?

Study site and data

Ranks of significance of input parameters in CLM4





- Calibrate the three significant hydrological parameters of the CLM with quantified uncertainty
  - Data: measurements of Latent Heat (LH), 1997-2004, US-MOz site
  - Bayesian calibration – develop a joint distribution of the most sensitive CLM parameters (F_{drai}, $\log(Q_{dm})$, S_y) with LH data

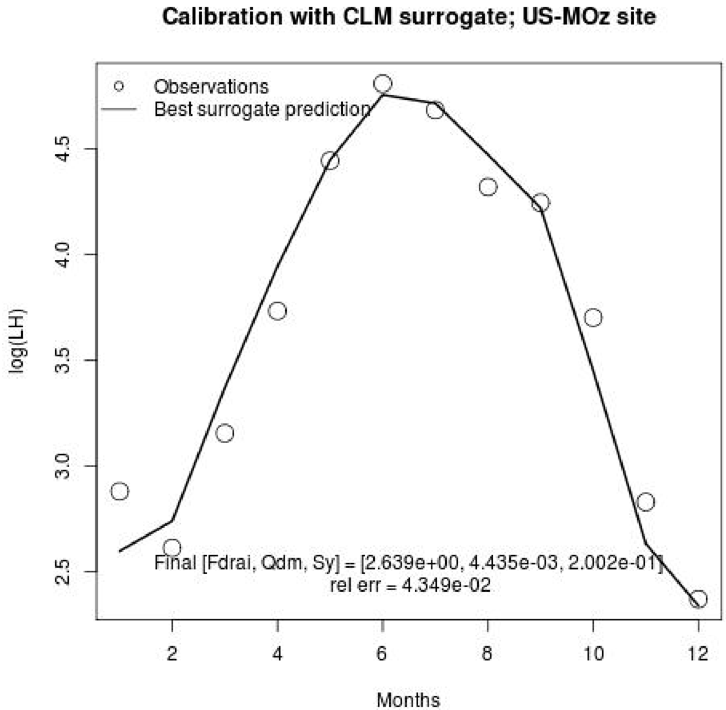
Sandia National Laboratories

Steps in calibration process

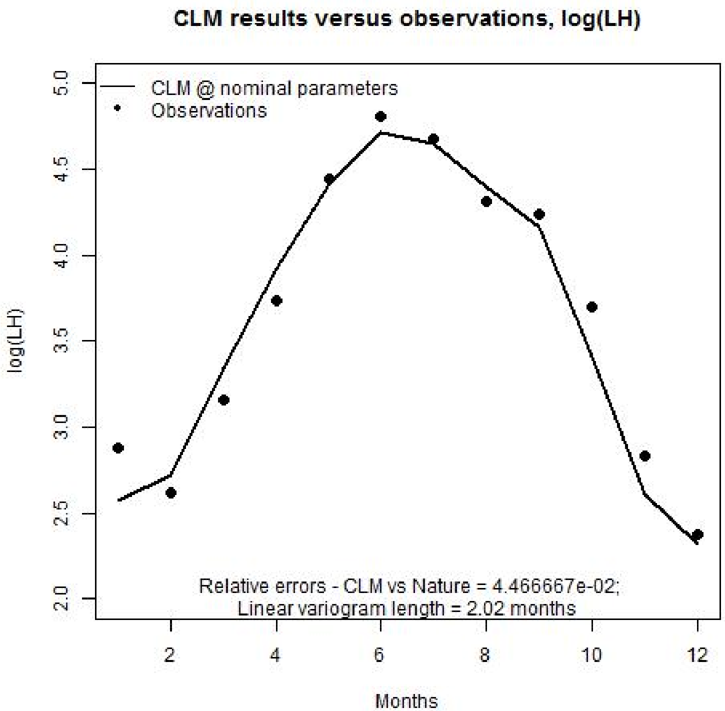
- Construction of surrogate models
  - Run CLM for 282 values of $\{F_{drai}, Q_{dm}, S_y\}$, sampled from the parameter space (lower & upper bounds are known); output LH(t)
  - Postulate competing polynomial models $LH(t) = g(F_{drai}, \log(Q_{dm}), S_y; t)$
  - Fit to data: model selection based on Bayesian Info. Criterion
- Calibration – is there a unique set of $\{F_{drai}, \log(Q_{dm}), S_y\}$ that explains LH observations at US-MOz?
  - Perform optimization-based model fitting using surrogate models
- Bayesian calibration
  - Fit surrogate to US-MOz data using Markov Chain Monte Carlo
  - Check if results are sensitive to the surrogate model
    - If the surrogates were made with half the CLM runs, would $\{F_{drai}, \log(Q_{dm}), S_y\}$ be different?

Sandia National Laboratories

Deterministic model fits to LH observations



Predictions with calibrated surrogate

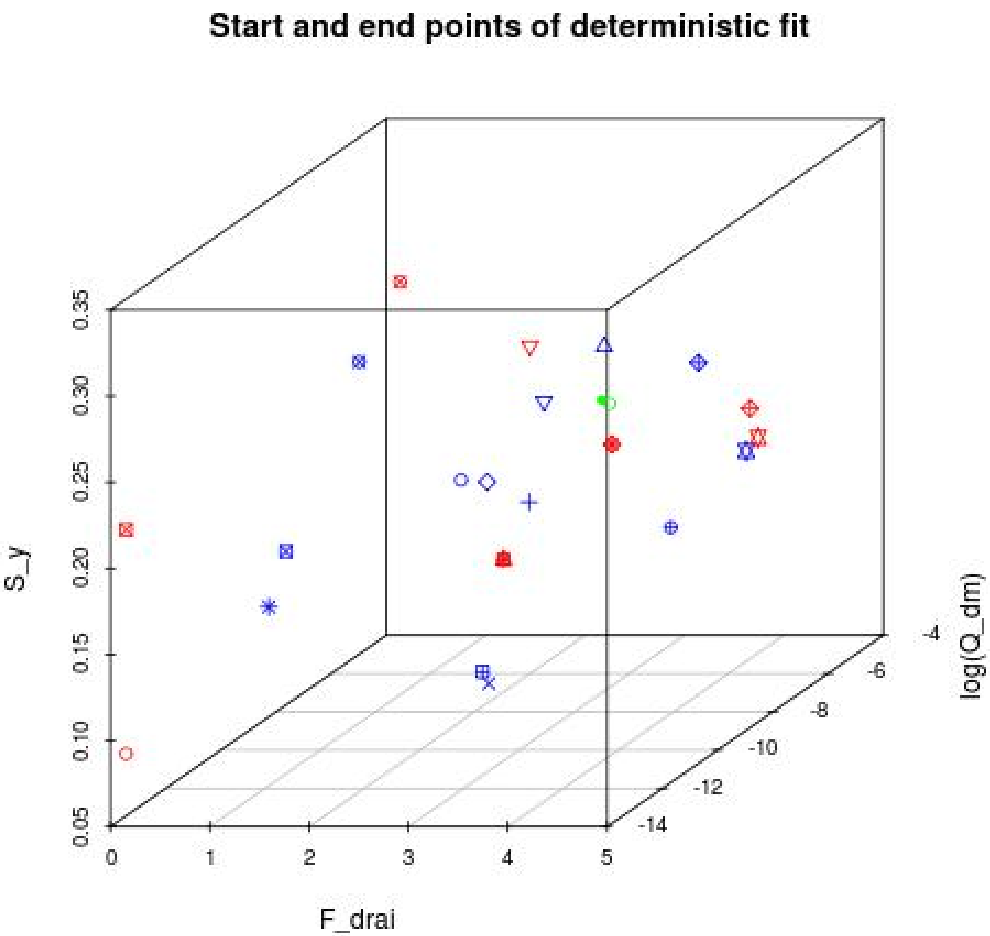


CLM predictions with nominal values

- Deterministic fit (w/ surrogate) and "nominal values" look similar
  - But errors sum to zero in the surrogate case

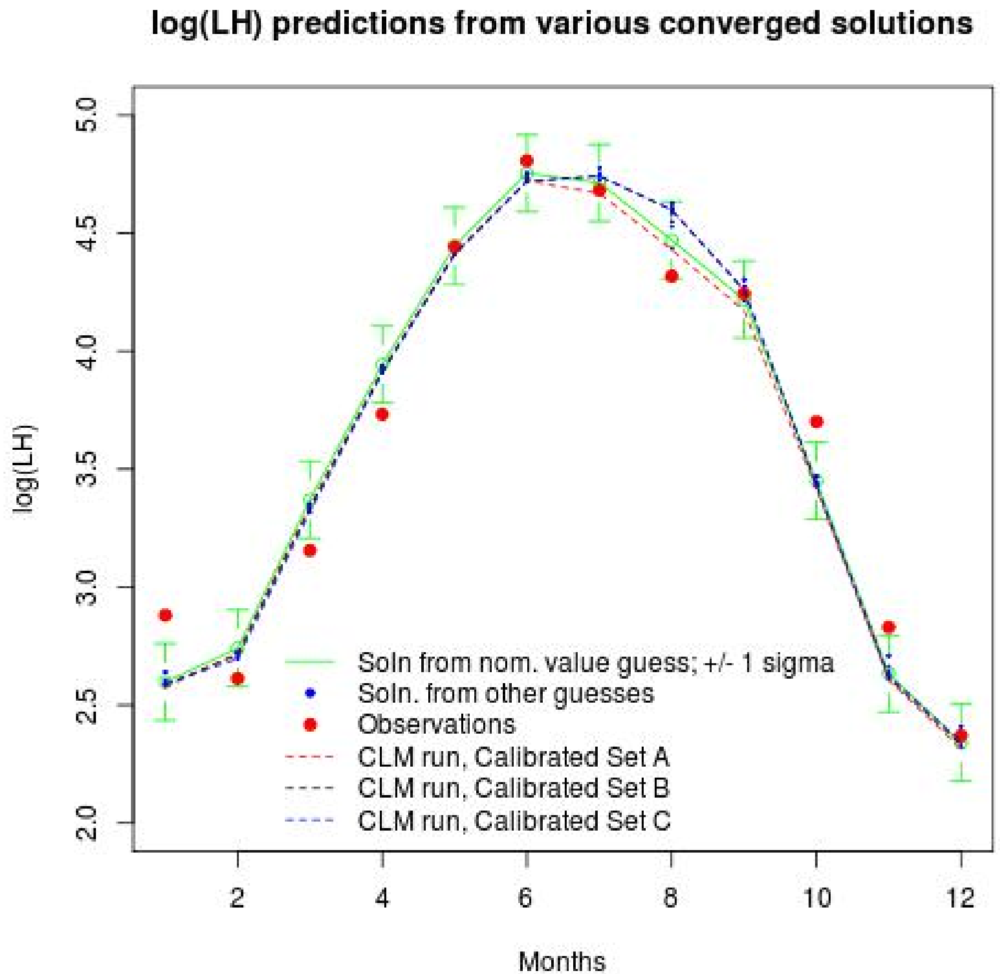
Calibrated values are non-unique!

- Blue symbols: starting points
- Red symbols: Converged/stopping points
- Green symbol: starting & ending points, starting from nominal values
- Next step: Project on 2D and see



…. And they are all equally good!

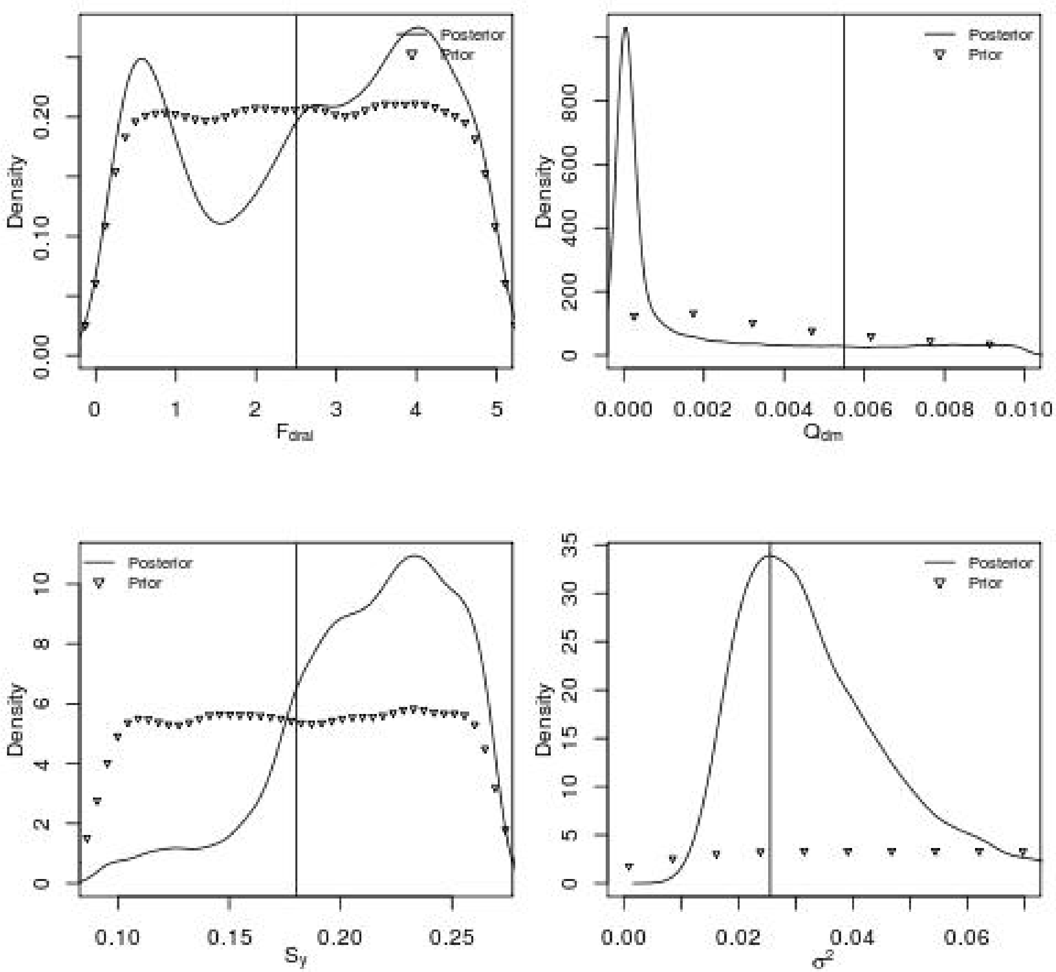
- Green: converged solution, starting from nominal values
- Blue: all other 13 converged solution
  - No noise added!
- Error bars: 1σ error between green and observations
- Repeat runs with a set of 3 calibrated parameter with real CLM
  - Dashed lines
- Bottomline: Variation caused by various converged values negligible compared to CLM – observation misfit



MCMC calibration

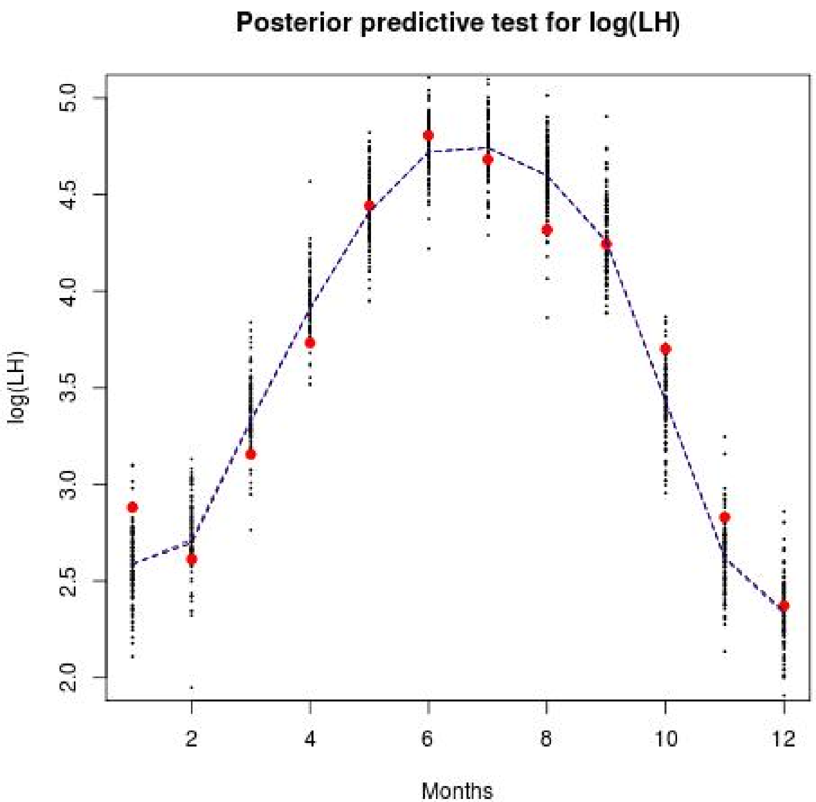
- Since there seem to be multiple, equally good, calibrated values of $\{F_{drai},Q_{dm},S_y\}$, is there a distribution that we should target?
  - What does this distribution look like?
- Construct distribution via MCMC
  - Have 3 different starting points and estimate $\{F_{drai},Q_{dm},S_y\}$
  - Model the structural error as i.i.d. Gaussian; estimate it
  - See if they provide (1) converged distributions or (2) converged summary statistics like 25^{th}, 50^{th} and 75^{th} percentiles of
- How do summary statistics compare with
  - Nominal/default values of $\{F_{drai},Q_{dm},S_y\}$ in CLM
- What big is the model – data error (model structural error)?

Posteriors and nominal values

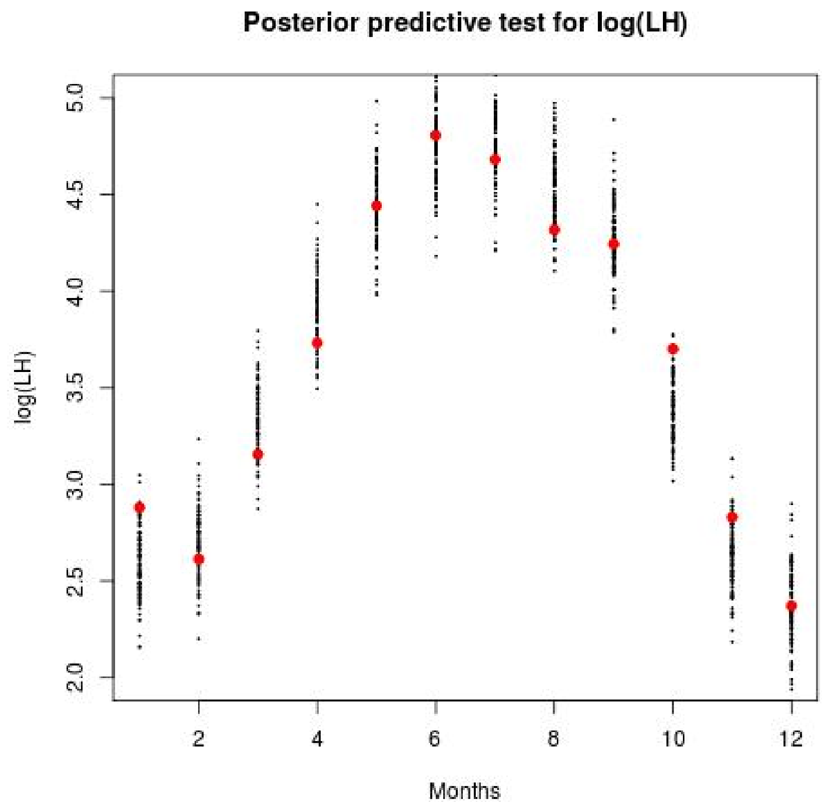


- MCMC required 10^5 model evaluations to converge
- Vertical lines are nominal values
- Nominal value for σ^2 is from the deterministic fit of surrogate
- Surrogate model constructed from 282 CLM runs

Posterior predictive test (282 v/s 128 runs)



282-run calibration

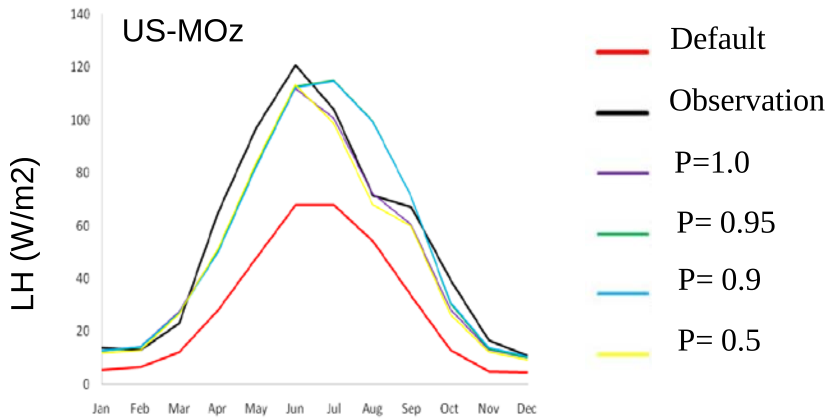


128-run calibration

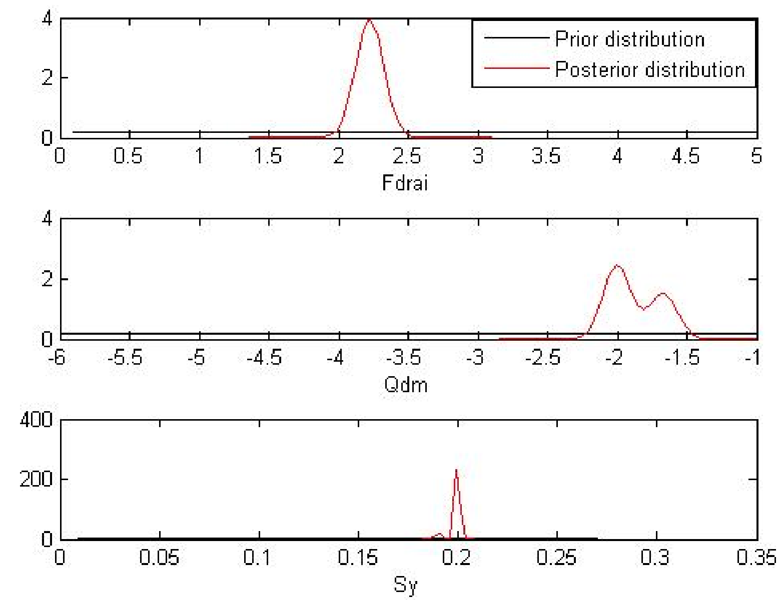
- Surrogate model constructed with < ½ the runs has similar predictive skill
- Dashed lines are 3 runs done with CLM
  - the predictions are not artifacts of the surrogate model

Sandia National Laboratories

MCMC inversion with CLM (not surrogate)



CLM simulations of LH using default parameters and MCMC-inverted posteriors
P — reference acceptance probability



Posterior distributions of the three significant parameters through MCMC inversion

GC43D-1052: Y. Sun; Z. Hou; M. Huang; F. Tian; L. Leung, Inverse Modeling of Hydrologic Parameters Using Surface Flux and Streamflow Observations in the Community Land Model, 1:40 PM-6:00 PM, Hall A-C (Moscone South), session GC43D. Interpretation and Uncertainty Quantification of Climate and Integrated Earth System Models IV Posters

Sandia National Laboratories

Conclusions

- CLM has non-unique solutions
  - Requires calibration in the form of distributions
- Bayesian methods allow us to estimate parameters as distributions
  - Even for expensive models like CLM
  - Allow probabilistic predictions, that enable us to quantify risk of failure / error in predictions
- Surrogate models often require significant sophistication to construct
  - Using sparsity-enforced model fitting to find the most parsimonious polynomial model
  - And adding a kriging component for local interpolation/structure in CLM's behavior in F_{drai}-Q_{dm}-S_y space

Sandia National Laboratories

BONEYARD

Surrogate model and calibration

- Model: Log(LH) = quadratic function of (F_{drai}, Q_{dm}, S_y) + correction
  - Quadratic model coefficients calculated from 232 CLM runs (Learning set); serves as a "trend" in a kriging model
  - Correction obtained by kriging interpolation from 232 data points
  - Prediction error ~ 5-10%, calculated from Testing Set (50 runs)
- Other models (linear & and higher order) investigated and rejected using BIC
- Calibration first done with L-BFGS, driving surrogate mode
  - To investigate the nature of the calibration problem
- Calibration redone in a Bayesian setting
  - Use MCMC to develop joint distribution of (F_{drai}, Q_{dm}, S_y)
  - Quantify uncertainty in calibrated values

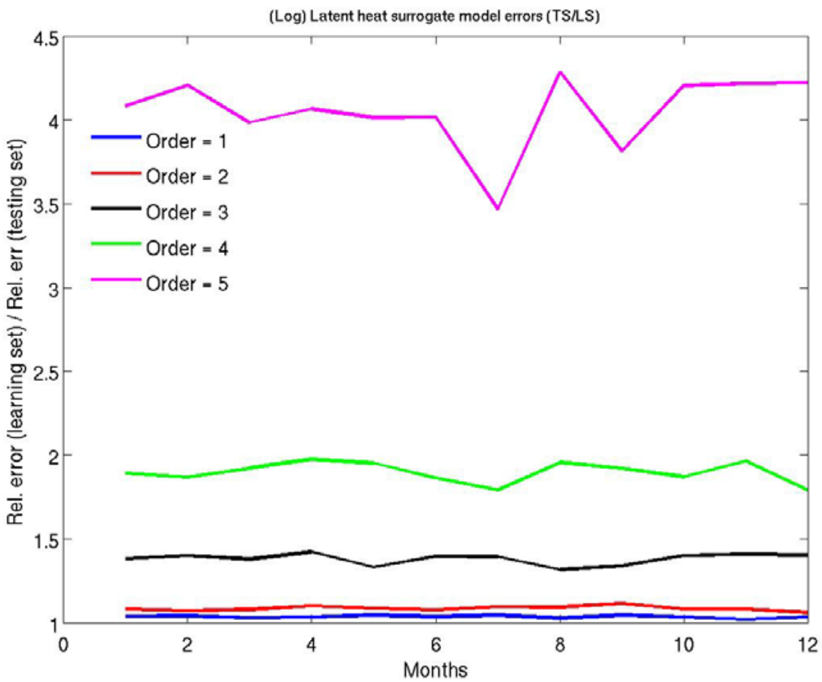
Making a polynomial fit

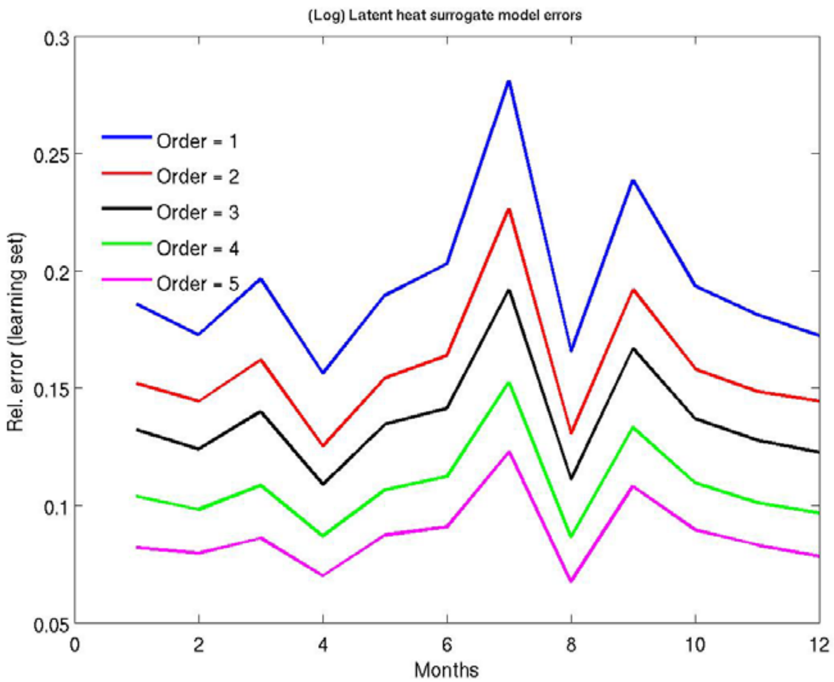
- Propose multiple polynomial models of different orders

$$\log(LH) = \mathbf{y}^{(\mathbf{clm})} = \sum_{i=1}^{3} \alpha_i p_i + \sum_{i=1}^{3}\sum_{j=i}^{3} \beta_{ij} p_i p_j + \sum_{i=1}^{3}\sum_{j=i}^{3}\sum_{k=(i+j)}^{3} \gamma_{ijk} p_i p_j p_k + \ldots..$$

- Separate 282 CLM runs into 500 (Learning-Set/Testing-Set pairs)
  - The testing set has 50 runs in it
- Fit polynomial to Learning Set using sparsity-enforced fitting
  - Called Bayesian compressive sensing (BCS)
  - Calculate error of fit in the Learning Set
- Use the fitted polynomial model to predict the LH for the $\{F_{drai}, Q_{dm}, S_y\}$ values in the Testing Set
  - Calculate error in Testing Set
- We expect that polynomial models are equally predictive in the Learning and Testing Sets

Visualizing relative errors across months



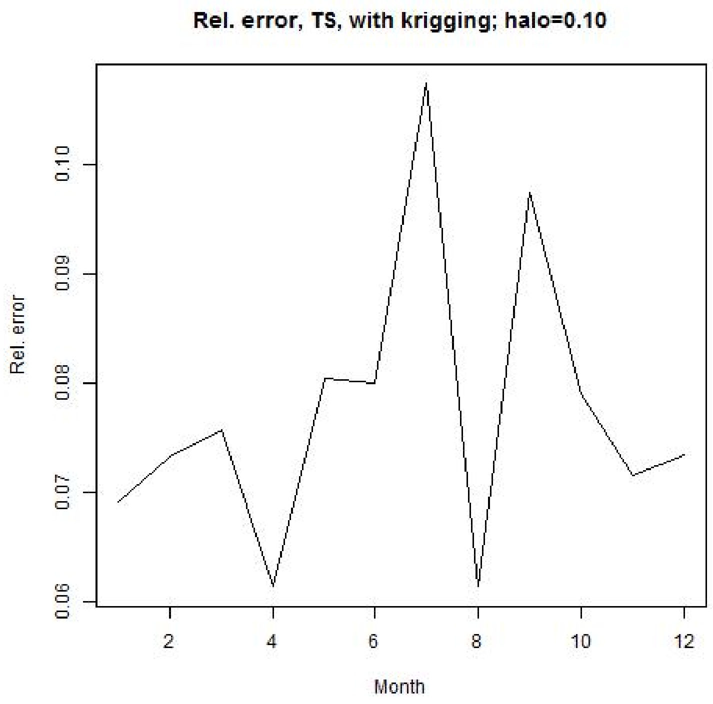


- Linear and quadratic models have similar errors for LS and TS
  - No overfitting here
- But quadratic model has lower errors overall, so choose it.

Sandia National Laboratories

Augmenting the quadratic model

- Quadratic model has pretty large error (~17%)
  - Because it captures no more than the trend of log(LH) in **p**-space
- $\mathbf{y}^{(\mathbf{surr})}(\mathbf{p}) = \mathbf{y}^{(\mathbf{quad})}(\mathbf{p}) + \mathbf{c}(\mathbf{p})$, **c** is a correction
  - It is smooth (correlated) function of **p**
  - Model c(**p**) as a multivariate Gaussian
- With c(**p**) model, we can evaluate $\mathbf{y}^{(\mathbf{surr})}(\mathbf{p})$ at arbitrary **p**
  - Includes a quadratic prediction
  - And a correction interpolated from the 128 runs



Augmented model give max 10% error

Summary statistics from 3 MCMC runs

Parameters	Run 1 Mean (25^{th}, 50^{th}, 75^{th}) PC	Run 2 Mean (25^{th}, 50^{th}, 75^{th}) PC	Run 3 Mean (25^{th}, 50^{th}, 75^{th}) PC	Best deterministic model
Start Point	{2.5, 5.5e-3, 2.0e-1}	{3.0, 1.0e-3, 2.5e-1}	{1.5, 7.5e-3, 1.5e-1}	{0.945, 2.28e-4, 0.26}
F_{drai} (2.5)	2.69; (1.25,2.8,3.72)	2.71; (1.3, 2.94, 3.96)	2.71; (1.26, 2.94, 3.97)	2.64e-01
Q_{dm} (5.5e-3)	1.66e-3; (6.9e-6, 9.6e-5,2.0e-3)	1.76e-3; (6.8e-6, 8.6e-5, 2.4e-3)	1.76e-3; (7.0e-6, 8.6e-5, 2.2e-3)	4.889e-03
S_y (0.2)	2.14e-1; (1.9e-1, 2.2e-1,2.4e-1)	2.14e-1; (1.9e-1, 2.2e-1, 2.4e-1)	2.14e-1; (1.9e-1, 2.2e-1, 2.4e-1)	2.698e-01
σ^2	0.035; (0.024, 0.031, 0.043)	0.036; (0.024, 0.032, 0.043)	0.036 (0.024, 0.032, 0.043)	0.0255

- Means & quantiles of posterior samples from the 3 MCMC runs do not vary much
- They don't deviate much from the nominal values either, except for Q_{dm}
- The best deterministic run gives a smaller model-data error

Sandia National Laboratories