Estimating a model discrepancy term for the Community Land Model using latent heat and runoff observations

J. Ray, L. Swiler, M. Huang and Z. Hou

Contact: jairay@sandia.gov

Sandia National Laboratories, CA & NM and Pacific Northwest National Laboratory.

SAND2015-1458C

Acknowledgments: Funded by DoE/SC/ASCR

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.
Introduction

- **Aim**: Calibrate the Community Land Model (CLM) using time-series measurements of latent heat and runoff
  - Bayesian calibration of 3 hydrological parameters w/ uncertainty
  - Estimate structural error (model-form error)
  - Compare with calibration done with each data type individually
- **Site**
  - US-MOz (latent heat) and MOPEX site # 7186000 (runoff)
- **Why?**
  - Structural error impairs a model’s ability to reproduce all observables well
- **Challenges**
  - CLM is expensive – 45 minutes/invocation per site;
  - No. of model invocation needed for Bayesian calibration = $O(10^4)$
What is CLM?

• A model for biogeochemical & hydrological processes
• Used in Earth system models; coupled to an atmosphere & ocean model
• Can be used in global (gridded) mode or locally for a site ("bucket" mode); can be driven by real meteorology
• Distributed by NCAR; has hard-coded parameters ("nominal values") which are meant to provided good global predictions
• When used in local mode, the parameters have to re-calibrated to be representative of local hydrological and biogenic processes
  – But it is not known whether calibrating to 1 data stream (e.g. latent heat) makes it predictive for all other observables
  – This is a type of structural error
What is structural error?

• The fundamental inability of a model to reproduce observations
  – Caused by missing physics in the model
• Previous work\(^1\) has shown that calibrating to latent heat (LH) observations makes CLM predictive for LH
  – And it has a modest structural error that can be modeled as i.i.d. Gaussians
  – This does not show if the calibrated model can reproduce other observables like runoff
• This study
  – Calibrate using runoff; see what parameter estimates are like and how well we reproduce runoff
  – Then calibrate jointly on runoff and LH and see whether we still reproduce observations well
  – And how far the parameter estimates are from nominal values

\(^1\)Ray et al, Bayesian calibration of the Community Land Model using surrogates, SIAM J. Unc. Quant., accepted January 2015
The observations

• Data covers 2004-2007, 48 months
• Latent heat (LH) observations, $Y_{LH}^{(obs)}$
  – Obtained from US-MOz – a site in Missouri Ozark mountains
  – Averaged monthly, and then climatologically averaged to provide a 12-month time-series
• Runoff observations, $Y_{WPC}^{(obs)}$
  – Very noisy and not very useful as-is
  – We take a wavelet transform and use the amplitude-squared (wavelet power) at each time-scale as the observations
    • Called wavelet power curve (WPC)
  – We retain time-scales between 21 days and 4 years for calibration
• Sensitivity analysis showed that LH and WPC are most sensitive to 3 hydrological parameters – $p = \{F_{drai}, Q_{dm}, S_y\}$
  – These will be our calibration variables
Bayesian inference

- Model parameters $p = \{F_{drai}, \log_{10}(Q_{dm}), S_y\}$ estimated with the model errors
  - $Y^{(obs)}_{LH} = M_{LH}(p) + \varepsilon_{LH}$, $\varepsilon_{LH} \sim N(0, \sigma^2_{LH})$
  - $Y^{(obs)}_{WPC} = M_{WPC}(p) + \varepsilon_{wpc}$, $\varepsilon_{wpc} \sim N(0, \sigma^2_{wpc})$
  - $\sigma^2_i$, $i \in \{LH, WPC\}$ is a crude measure of structural error in CLM

- Our prior beliefs (PDFs) for each parameter in $\{F_{drai}, \log_{10}(Q_{dm}), S_y\}$ are independent, uniform distributions with prescribed upper & lower bounds

- Posterior distribution $P(p, \sigma^2_{LH}, \sigma^2_{WPC} \mid Y^{(obs)}_{LH}, Y^{(obs)}_{WPC})$

\[
P(F_{drai}, \log_{10}(Q_{dm}), S_y \mid Y_{LH}^{obs}, Y_{WPC}^{obs}) \propto \exp \left( - \left( \frac{Y_{LH}^{obs} - M_{LH}(p)}{\sigma^2_{LH}} \right)^2 - \left( \frac{Y_{WPC}^{obs} - M_{WPC}(p)}{\sigma^2_{WPC}} \right)^2 \right) \pi(p)
\]

- Solved using an adaptive Metropolis algo – DRAM
Surrogate models

- The inverse problem needs about 50K invocations of CLM
  - Can’t be done today, so we make surrogates

- Surrogate details
  - We sample the \((F_{\text{drai}}, \log_{10}(Q_{\text{dm}}), S_y)\) space with 282 points chosen via a quasi Monte Carlo space-filling method
  - CLM is run at these points; we save climatologically averaged predictions of LH and runoff
  - This is our training set

- Models are curve fits – express \(Y = M(p)\)
  - You have to pick \(M\) and fit to data
  - You need some way to check against overfitting - cross-validation, AIC etc.
  - Invariably latent heat or runoff needs to be transformed before being able to fit \(M\)
Latent heat surrogate models

• Transformations
  – 48 months of LH data is climatologically averaged, then log transformed

• Proposed a 5th order polynomial for $M_{LH}(p)$
  – Used AIC to simplify the model down to quadratic
  – Use randomized subsample validation tests to check for overfitting
  – Separate model for each month
  – The final fitted polynomial model has 10% - 20% errors – not good enough

• Regression kriging
  – Used quadratic as a mean/trend model and stationary Gaussian Process model around it (to combat 10%-20% discrepancy)
  – All models’ errors dropped below 10%
Runoff surrogate models

- **WPC surrogates**
  - Surrogate could only be made for a subspace of \((F_{\text{drai}}, \log_{10}(Q_{\text{dm}}), S_y)\) space
  - Computed the MSE of each training run wrt observations & discarded the worst 25%
  - The retained parameters covered a region \(\mathcal{R}\) of the parameter space
  - Within \(\mathcal{R}\), \(M_{\text{WPC}}(p)\) could be modeled using quadratic polynomials
- **We redefine our prior**
  - \(\pi(p) = 1, p \in \mathcal{R}, 0\) otherwise

\(\mathcal{R}\) defined using a SVM classifier trained on selected & discarded runs in the training set
Calibration with LH data only

\[ P(\cdot|\cdot) \propto \exp \left( -\frac{(Y_{LH}^{\text{obs}} - M_{LH}(p))^2}{\sigma_{LH}^2} - \frac{(Y_{WPC}^{\text{obs}} - M_{WPC}(p))^2}{\sigma_{WPC}^2} \right) \pi(p) \]

- The PDFs are not very well defined (bi-modal etc.)
- There is not much support for the nominal values of parameters
Reproducing observations

- Pick 100 samples from the posterior density
- Run CLM for each
- Plot ensemble of predictions
- The variation in log(LH) predictions is tiny
  - Can’t see the error bars around the circles
  - Explains why it was so difficult to find a sharp posterior distribution
Calibration with WPC data only

\[ P(\psi|\lambda) \propto \exp \left( -\frac{(Y_{\text{LH}}^{\text{obs}} - M_{\text{LH}}(\psi))^2}{\sigma_{\text{LH}}^2} - \frac{(Y_{\text{WPC}}^{\text{obs}} - M_{\text{WPC}}(\psi))^2}{\sigma_{\text{WPC}}^2} \right) \pi(\psi) \]

- The PDFs are simple
  - Not much support for nominal \( F_{\text{drai}} \)
- PDFs very different from the ones estimated using LH data only
  - First indication that it takes very different estimates of the parameters to match LH and WPC data
Reproducing observations

- Lots of scatter in predictions at small temporal scales
- But calibrated model’s predictive skill better than nominal values of the parameters
Joint calibration to LH and WPC data

- Huge change in PDFs
  - $F_{drai}$ affects LH much, so joint and LH-only calibration are similar
  - $Q_{dm}$ affects WPC much, and so joint and WPC-only calibration similar
- $S_y$ – well, your pick

- And the structural error is 6x larger
  - We simply can’t be very predictive
Reproducing LH observation with CLM

- Hardly any variability
  - LH predictions not at all sensitive to posterior
  - No wonder we could not get useful PDFs our of LH-only calibration
Reproducing WPC observation with CLM

• Good variability
  – Observations contained in inter-quartile “error bars”

• Big improvement in predictions at monthly timescales
  – Should be enough to resolve seasonal variations
Conclusions

• CLM can be calibrated to reproduce a given datastream well
  – The model form error so obtained is too optimistic
  – And the parameters estimates are wrong
• Joint calibration with 2 types of data uncovers a second type of structural error
  – Its inability to reproduce multiple observations stream accurately
  – The parameter estimates obtained from 2 data streams have some resemblance to their nominal values
• Related talks
  – L. Swiler, MS 164, Room 251, Monday, 2:20pm – 2:50pm [On the perils of parameter estimation using surrogates of CLM]
  – Z. Hou, CP 16, Room 254B, Wednesday, 9:25am – 9:35am [On the applicability of parameter estimated from one site, to other similar sites; called “transferability”]