Bayesian calibration and evaluation of transferability of hydrologic parameters in CLM

Maoyi Huang, Zhangshuan Hou, Yu Sun, Huiying Ren, Ying Liu, Yinghai Ke, Teklu Tesfa, Guoyong Leng, Jiali Guo, Hongyi Li, L. Ruby Leung

Pacific Northwest National Laboratory

CESM joint LMWG-BGCWG-UQ meeting, Mesa Lab, Boulder, CO
21 February 2013
Outline

▶ An uncertainty quantification framework designed for CLM
  ■ Results at selected flux towers and MOPEX basins

▶ Bayesian inversion/calibration of the hydrologic parameters
  ■ MCMC inversion of CLM4
  ■ MCMC inversion using surrogates (Ray et al., this meeting)

▶ Toward understanding the parameter transferability
  ■ Classification of MOPEX basins
  ■ Future work plan
An uncertainty quantification framework designed for CLM

Parameterization - candidate parameters of interest

- Prior information (mean, variance, bounds)
- Entropy method
  - Minimum Relative Entropy (MRE)
  - Closed-form Prior probability density functions (pdfs)
- Quasi Monte Carlo sampling
- Realizations of parameter sets
- CLM forward modeling
- Output responses:
  - Latent heat fluxes (LH)
  - Sensible heat fluxes (SH)
  - Total runoff

Entropy method formula:

\[ f_{MRE}(m|I) = \prod_{j=1}^{p} \frac{\beta_j e^{-\beta_j m_j}}{e^{-\beta_j m_j} - e^{-\beta_j \mu_j}} \]

Symbol Definition

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>f_{max}</td>
<td>Max fractional saturated area, from DEM</td>
</tr>
<tr>
<td>C_s</td>
<td>Shape parameter of the topographic index distribution</td>
</tr>
<tr>
<td>f_{over}</td>
<td>Decay factor (m^{-1}) for fsat</td>
</tr>
<tr>
<td>b</td>
<td>Clapp and Hornberger exponent</td>
</tr>
<tr>
<td>K_s</td>
<td>Hydraulic conductivity (mm s^{-1})</td>
</tr>
<tr>
<td>\theta_s</td>
<td>Porosity</td>
</tr>
<tr>
<td>\Psi_s</td>
<td>Saturated soil matrix potential (mm)</td>
</tr>
<tr>
<td>f_{drai}</td>
<td>Decay factor (m^{-1}) drainage</td>
</tr>
<tr>
<td>q_{drai,max}</td>
<td>Max drainage (kg m^{-2} s^{-1})</td>
</tr>
<tr>
<td>S_y</td>
<td>Average specific yield</td>
</tr>
</tbody>
</table>

Output statistics:
1. Natural variability of LH/SH/runoff
2. Propagation of uncertainty of input parameters

Multi-variate generalized linear model analysis and significance statistical tests:
1. Rank of significance of input parameters
2. Relationships between LH/SH/runoff and inputs

Model reduction and adaptive sampling
Selected flux towers and MOPEX basins

Legend
- Selected flux towers
- Selected MOPEX basins
- USGS stations

*Hou et al. (2012), JGR*
*Huang et al. (in review), JHM*
Efficient sampling of the parameter space

QMC samples in probability space

Prior PDF of model parameters

QMC samples in parameter space
Surface fluxes and runoff are very sensitive to model parameters
Larger sensitivity to parameters of subsurface processes
Inversion workflow for CLM4

Inversion setup
- Default
- Mean
- PEST estimates
  - Choices of initial values
  - Choices of proposal distribution width
  - Reference acceptance probability

Inversion methods
- Bayesian updating
  - Metropolis-Hasting sampling methods

Parameterization
- CLM(4.0)
  - Observation (Heat flux and runoff)

Posterior distribution of parameters
- Misfits
- Heat flux and runoff responses

Improvements through Bayesian model averaging

Sun et al., to be submitted to HESS
Selected sites for inversion
Inversion results

Inversion based on 10 parameters

Inversion based on 3 parameters
Model reduction and parameter transferability

- Although parameter calibration could be used to improve model simulations and quantify uncertainties, it suffers from issues associated with high-dimensionality in number of parameters or large spatial dimensions (e.g., a global domain);

- To reduce the dimensionality, we need to cluster sites/grid cells into classes that share similar characteristics;

- A PCA+MCLUST-based approach is used for classifications;

- Different classification systems were explored:
  - C-Class: Climate-based
  - S-Class: sensitivity-based;
  - H-Class: hydrologic indices-based
431 MOPEX basins over the US

C-Class: Koppen climate classes
Parameter sensitivity over 431 MOPEX basin

Sensitivity of LH

Sensitivity of runoff
Overall significance of the hydrologic parameters across the MOPEX basins
Determine number of classes

- Bayesian Information Criterion (BIC) for determining number of classes
- It is reasonable to choose 8 classes for the best compromise between class model complexity and likelihood of the model
H-Classes of the MOPEX basins

1. Runoff ratio
2. Humidity index
3. Baseflow index
4. Horton index
5. Drainage density
Summary and future work

- An uncertainty quantification framework has been implemented with CLM4; Application of the framework to selected sites suggests that CLM4 show the largest sensitivity to subsurface runoff generation parameters;
- The global sensitivity analyses provides guidance on reduction of parameter set dimensionality and parameter inversion framework design for CLM4;
- To reduce the dimensionality, we classify sites into classes that share similar characteristics;
- MCMC-Bayesian calibrations (using CLM or its surrogates) will be conducted at representative sites within each class to test/confirm the parameter transferability among all MOPEX basins.
Acknowledgement

- DOE: Climate Science for a Sustainable Energy Future

- PNNL: Platform for Regional Integrated Modeling and Analysis (PRIMA) initiative
Summary of significance over flux towers

**Site-wise**
- Latent heat flux
- Sensible heat flux
- Total runoff

**Warmer = More Significant**

**Seasonal**
- Latent heat flux
- Sensible heat flux
- Total runoff
potential of inversion:
sensitivity of correlation with observed hydrographs over the MOPEX basins

Peak flow
Center of mass flow rate (CT)

Month in water year

Oct Nov Dec Jan Feb Mar Apr May Jun Jul Aug Sep

peakflow score
CT score
COR score

\(f_{\text{max}}\)
\(C_s\)
\(f_{\text{over}}\)
\(f_{\text{deth}}\)
\(Q_{\text{dm}}\)
\(S_y\)
\(b\)
\(\Psi_s\)
\(K_s\)
\(\theta_s\)
Posterior distributions based on the 3-parameter inversion at US-MOz