SAND2015-0706C

Applicability of surrogate-based MCMC-Bayesian inversion of CLM at flux tower sites with various field conditions

[1]Hou Z., [2]Ray J., [1]Huang M., [2]Swiler L., [1]Ren H., [1]Liu. Y

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

Motivations

- Surface fluxes (e.g., latent heat flux LH) are sensitive to major hydrologic parameters in CLM4 at various flux tower sites.
- To evaluate the feasibility of developing surrogate models as alternative to computationally demanding CLM simulator for model parameter estimation
- To evaluate applicability of several MCMC-Bayesian inversion strategies for model calibration at flux tower sites with various field conditions
- To evaluate parameter transferability
- inverting hydrologic parameters in CLM4 using surface flux and streamflow observations .

Study sites and parameterization

- Ten hydrologic parameters were selected because of their significant impacts on surface and subsurface runoff, latent and sensible fluxes, and soil moisture.
- Parameter screening with an UQ framework that integrates quasi-Monte Carlo sampling, minimum-relative-entropy theory for defining priors, and statistical parameter significance tests.

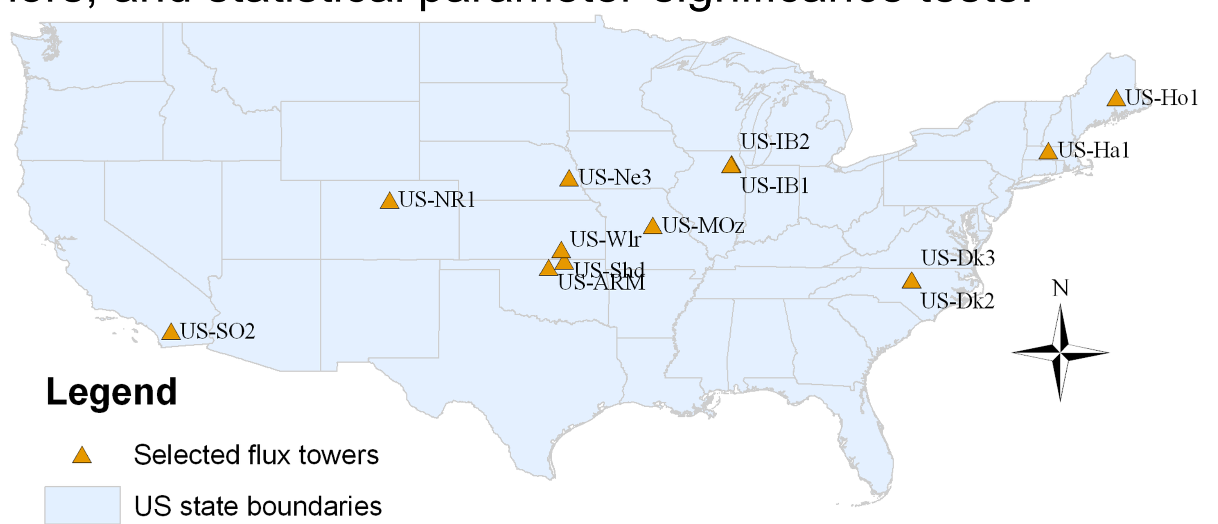


Figure 1. Geographic locations of the selected flux towers.

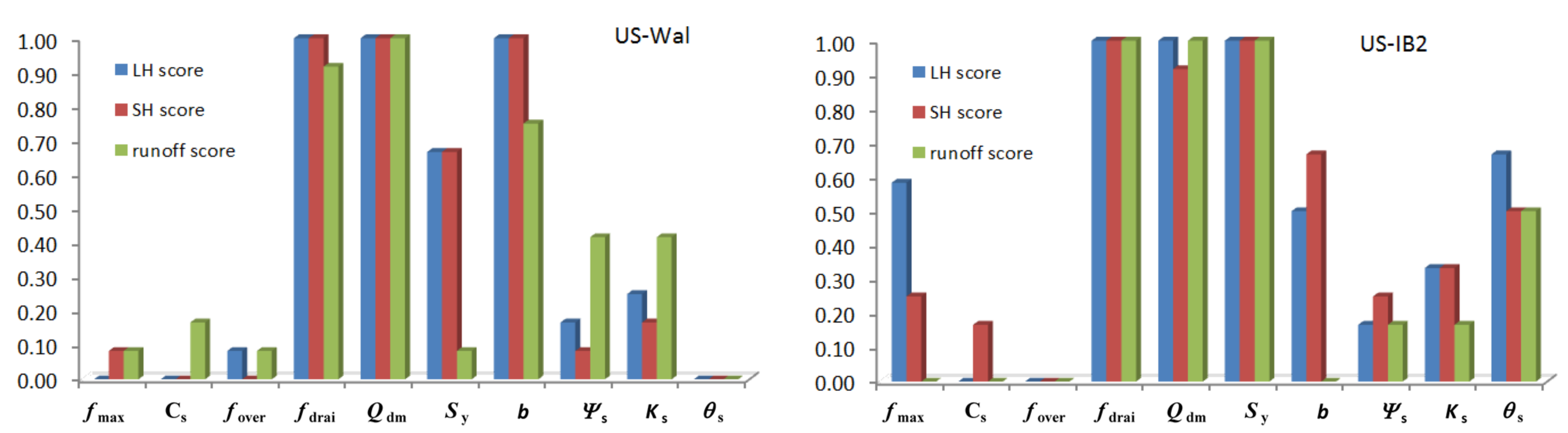


Figure 2. parameter significance scores at different flux tower sites.

Surrogate development

Tested up to 5th order polynomial models

$$\log(LH)=\sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{k=0}^{5}\sum_{l=0}^{5} a_{ij}\, p_i^k p_j^l, \quad (k+l)\le 5$$

given CLM-simulated LH with 256 parameter sets generated using quasi Monte Carlo sampling. A separate LH surrogate is created for each month. Cross-validation is conducted by separating the data into training and testing sets. Finalized surrogates are quadratic.

Inversion strategies

- MCMC-Bayesian approach for generating posterior samples
- Surrogate evaluation: surrogates to be included as the forward model need to pass the cross-validation (e.g., RMSEs of fitting for both randomly chosen training and testing sets < 15%)
- Modifications when surrogate development is difficult for certain data points:
  - Uses a composite model by add a kriging component to the fitting errors, that is, to constructs surrogate models of quadratic + kriging, then set up the likelihood and the prior and uses adaptive MCMC as before
  - makes a surrogate (e.g., quadratic) model for a subset of the parameter space close to the ‘true’ parameter set. A classifier is needed (e.g., using treed linear models) to define such a ‘good’ subspace for generating posterior samples.
  - Assume uncorrelated vs time-correlated errors.

Validation of surrogate models

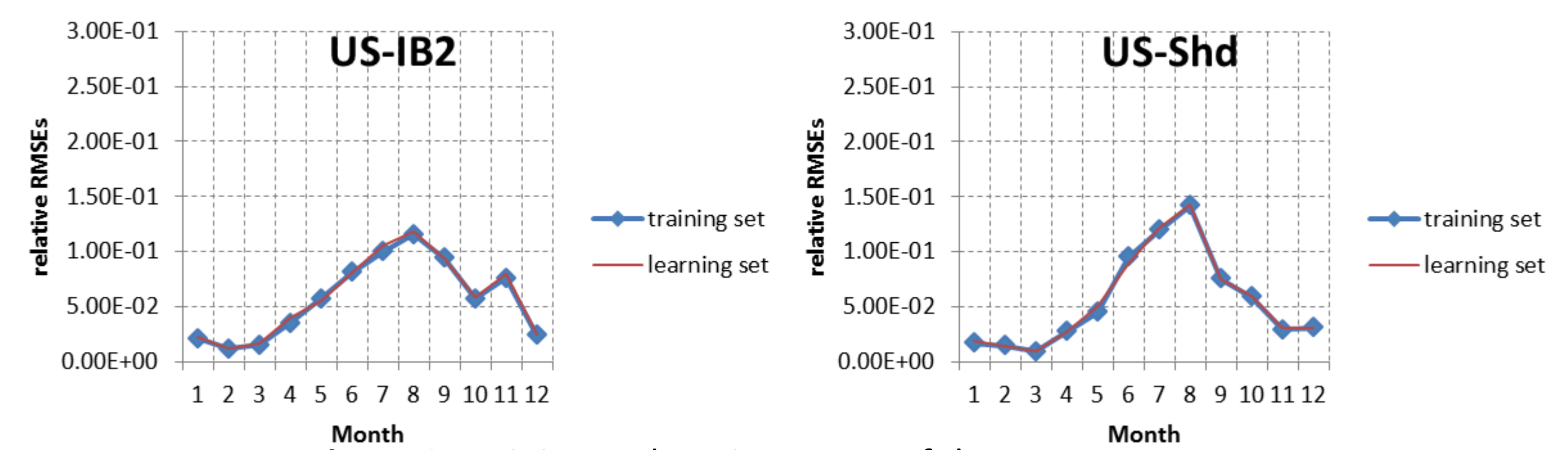


Figure 3. training and testing errors of the surrogates .

Composite surrogates vs subspace surrogates

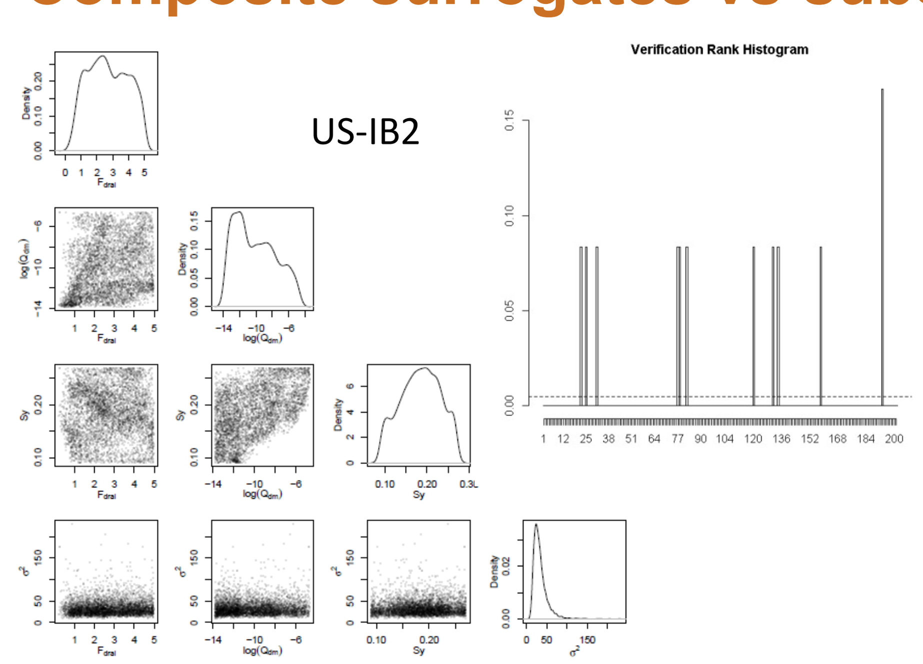


Figure 4. posterior distributions of parameters with composite models, and observation rank for validation.

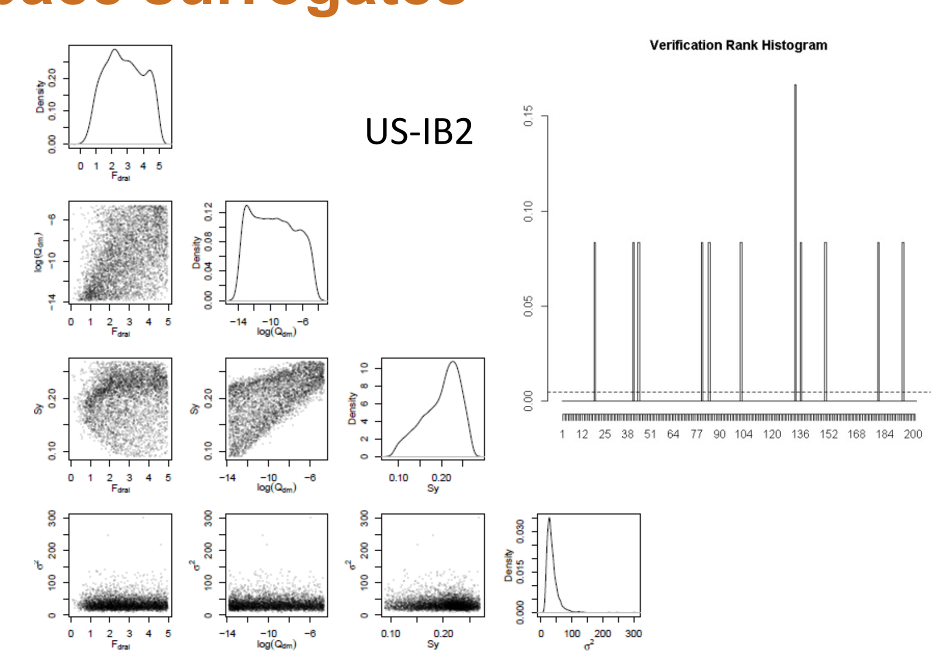
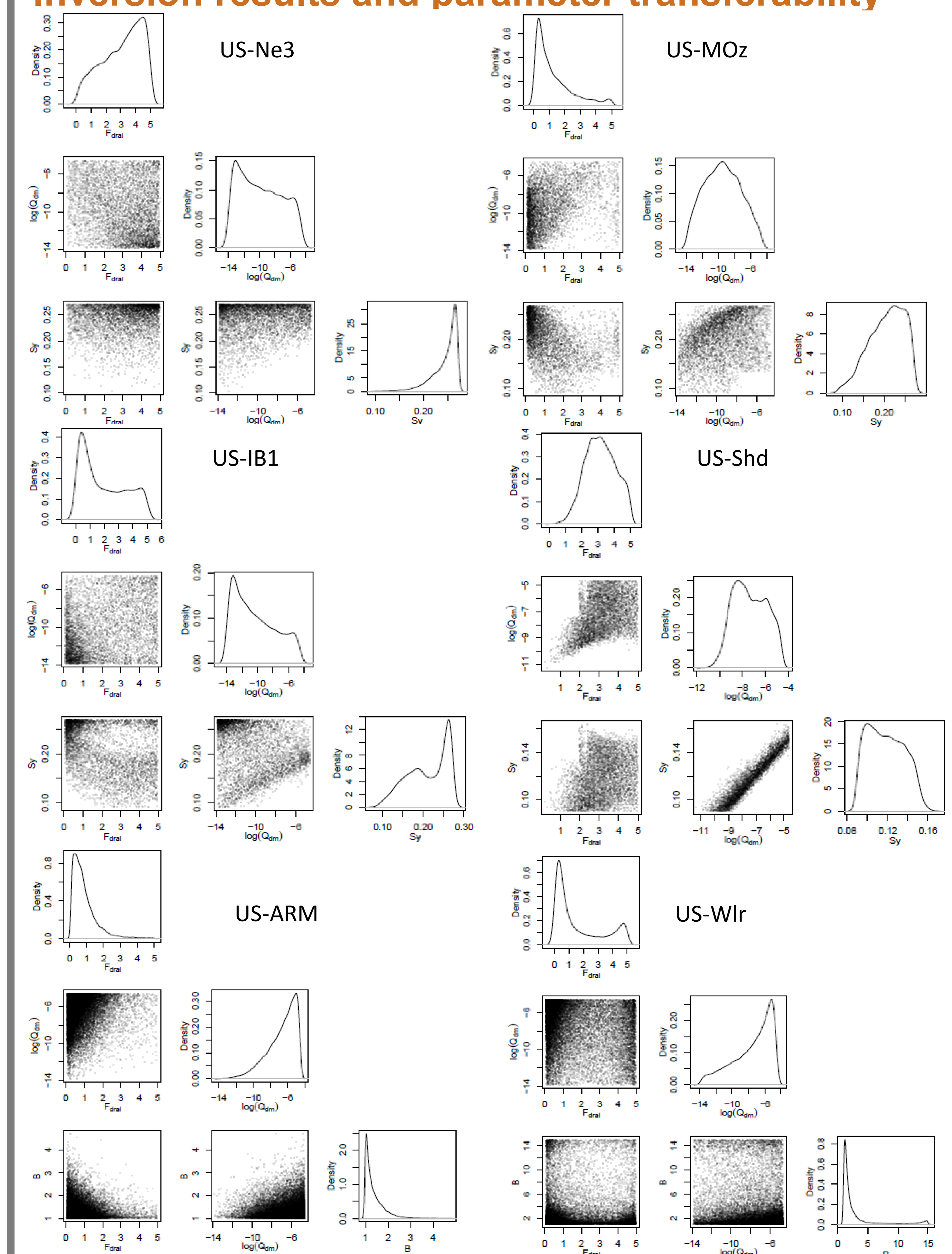


Figure 4. posterior distributions of parameters with subspace surrogates, and observation rank for validation.

Inversion results and parameter transferability



Conclusions

- Parameter screening is necessary to make the inverse problem less ill-posed.
- Surrogates, integrated with MCMC-Bayesian, enables efficient CLM model calibration
  - Task parallel computing enables simultaneous CLM simulations for parameter screening and then surrogate developments purposes .
- Feasibility of surrogate development needs to be checked before model calibration
- Screening of unrealistic combinations of parameter sets might be necessary where inconsistent or extremely nonlinear relationships exist between the response variables and unknown parameters to be estimated.
- Parameter transferability is to be summarized by studying more field sites that can be grouped into classes of similar climate and soil conditions.