



# Inverse modeling of hydrologic parameters using surface flux and runoff observations in the Community Land Model

Y. Sun<sup>1,2</sup>, Z. Hou<sup>1</sup>, M. Huang<sup>1</sup>, F. Tian<sup>2</sup>, and L. Ruby Leung<sup>1</sup>

<sup>1</sup>Pacific Northwest National Laboratory, P.O. Box 999, Richland, WA 99352, USA

<sup>2</sup>Department of Hydraulic Engineering, State Key Laboratory of Hydrosience and Engineering, Tsinghua University, Beijing 100084, China

Correspondence to: Z. Hou (zhangshuan.hou@pnnl.gov)

Received: 1 April 2013 – Published in Hydrol. Earth Syst. Sci. Discuss.: 23 April 2013

Revised: 26 August 2013 – Accepted: 10 November 2013 – Published: 10 December 2013

**Abstract.** This study demonstrates the possibility of inverting hydrologic parameters using surface flux and runoff observations in version 4 of the Community Land Model (CLM4). Previous studies showed that surface flux and runoff calculations are sensitive to major hydrologic parameters in CLM4 over different watersheds, and illustrated the necessity and possibility of parameter calibration. Both deterministic least-square fitting and stochastic Markov-chain Monte Carlo (MCMC)-Bayesian inversion approaches are evaluated by applying them to CLM4 at selected sites with different climate and soil conditions. The unknowns to be estimated include surface and subsurface runoff generation parameters and vadose zone soil water parameters. We find that using model parameters calibrated by the sampling-based stochastic inversion approaches provides significant improvements in the model simulations compared to using default CLM4 parameter values, and that as more information comes in, the predictive intervals (ranges of posterior distributions) of the calibrated parameters become narrower. In general, parameters that are identified to be significant through sensitivity analyses and statistical tests are better calibrated than those with weak or nonlinear impacts on flux or runoff observations. Temporal resolution of observations has larger impacts on the results of inverse modeling using heat flux data than runoff data. Soil and vegetation cover have important impacts on parameter sensitivities, leading to different patterns of posterior distributions of parameters at different sites. Overall, the MCMC-Bayesian inversion approach effectively and reliably improves the simulation of CLM under different climates and environmental conditions. Bayesian model averaging of the posterior estimates with

different reference acceptance probabilities can smooth the posterior distribution and provide more reliable parameter estimates, but at the expense of wider uncertainty bounds.

## 1 Introduction

Inverse problems (or parameter calibrations/optimizations) involve a general framework to derive information about a physical object or system from measurements directly or indirectly related to the physical object/system (Tarantola, 2005). During the past decades, numerous inversion strategies, deterministic or stochastic, have been developed and applied in earth systems science including atmospheric science, hydrology, geology, and geophysics. Three conditions (existence, uniqueness, stability of the solutions) are necessary for a well-posed inverse problem (Hadamard, 1902). However, as solution uniqueness and stability are usually violated in practice, some regularization is generally needed to introduce mild assumptions on the solution and prevent parametric over-fitting. It is also important for an inverse approach to be capable of quantifying and evaluating the prediction uncertainty, particularly for complex systems, where unknown parameters and observable variables have nonlinear and non-unique relationships, observations are limited, or the forward models are not perfect.

For a given inverse problem, one can choose different approaches depending on the requirements of parameter estimation accuracy, computational demand, and importance of prediction uncertainty (e.g., Hou et al., 2006; Chen et al., 2004; Hoversten et al., 2006), which requires understanding

of the advantages, disadvantages, and applicability of each method. Deterministic approaches have been used to obtain “optimal” parameter sets by evaluating the goodness of fit between observed and model simulated response variables (e.g., Sorooshian, 1981; Sorooshian and Dracup, 1980; Duan et al., 1992, 1993; Sorooshian et al., 1993; Hoversten et al., 2006). These approaches generally assume that an optimal parameter set exists and implicitly ignore the estimation of predictive uncertainties. However, a single optimal parameter set may not exist and the uncertainties associated with the optimal parameter sets could be large (e.g., Gupta et al., 1998; Klepper et al., 1991; Van Straten and Keesman, 1991; Beven and Binley, 1992; Yapo et al., 1996). Moreover, a model with the optimal parameter set may provide the best fit over the calibration period, but multiple parameter sets may result in comparable misfits and are likely to be the “true” values with certain probability, and therefore are considered to be acceptable or equally probable parameter sets (Van Straten and Keesman, 1991; Klepper et al., 1991). Stochastic approaches can address these limitations by describing the input parameter and output uncertainties in a statistical manner. Generally, the input parameter space is represented in a form of multivariate probability distributions of input parameters and must be sampled to generate multiple realizations of the model simulations so that the prediction range can be estimated based on the ensemble model simulations (Beven and Binley, 1992; Freer et al., 1996; Kuczera and Parent, 1998; Vrugt et al., 2003b). When multiple types of data are available, multi-objective calibration can be used to deal with parameter estimation uncertainty by combining several measures of performance for data fusion (Boyle et al., 2000; Kollat et al., 2012; Gupta et al., 1998; Vrugt et al., 2003a). In practice, different optimization methods can also be combined to improve the treatment of uncertainty in hydrologic modeling (Vrugt and Robinson, 2007; Feyen et al., 2007; Vrugt et al., 2005).

Uncertainty in model simulations can stem from model parameters, as well as from model inputs (e.g., atmospheric forcing and land cover data) and model structure (e.g., the physical parameterizations used to describe certain processes or the numerical solvers). Uncertainty in model input data or model structure can be reduced by improving the observation precision and taking advantage of multiple observational platforms with different error characteristics in space and time, or better understanding of the physical processes. Much progress has been made by the land surface modeling community in the last two decades to deal with uncertainty in model parameters, data, and model structure. The focus of this study is uncertainty in model parameters.

In previous studies (Hou et al., 2012; Huang et al., 2013), we investigated the sensitivity of surface fluxes and runoff simulations to major hydrologic parameters in version 4 of the Community Land Model (CLM4) by integrating CLM4 with a stochastic exploratory sensitivity analysis framework at 13 flux towers from the Ameriflux net-

work and 20 watersheds from the Model Parameter Estimation Experiment (MOPEX) (Huang et al., 2013) spanning a wide range of climate, landscape, and soil conditions. We found that the CLM4-simulated latent heat flux (LH), sensible heat flux (SH), and runoff show the largest sensitivity to subsurface runoff generation parameters. These studies demonstrated the necessity and possibility of parameter inversion/calibration using available measurements of surface fluxes and streamflow to invert the optimal parameter set, and provided guidance on reduction of parameter set dimensionality and parameter calibration framework design for CLM4.

This study aims to demonstrate the inversion methodology at selected sites based on the global sensitivity analyses detailed in Hou et al. (2012) and Huang et al. (2013). Among various inversion approaches, we adopt a stochastic Bayesian inversion approach integrated with Markov-chain Monte Carlo (MCMC) sampling. The unknowns to be estimated include model parameters for surface and subsurface runoff generation and vadose zone soil water movement. Different options for the inversion framework are evaluated at the selected sites. For example, it is important to evaluate the prior incompatibility issues in an inversion design (Hou and Rubin, 2005). As detailed in the remaining sections of this paper, we evaluated the impacts of prior information (e.g., initial guesses) on the inversion results. We also compared the consistency and reliability of inversion using both monthly and daily flux observations, and compared the performance of Bayesian model averaging to the individual inversion approaches for parameter estimation. We also discuss the issues related to data quality, data worth, and redundancy.

## 2 Site and data description

We perform parameter estimation at two flux tower sites and one MOPEX basin. The MOPEX basin is in close proximity to one of the flux tower sites to investigate model inversion using heat flux versus runoff data. These sites/basins are chosen based on previous sensitivity analyses over a larger set of flux tower sites and MOPEX basins (Hou et al., 2012; Huang et al., 2013) that demonstrated the feasibility and necessity of parameter calibration at those locations and to provide representative and contrasting climate and environmental conditions within the United States for more robust conclusions. US-ARM is located in Oklahoma and is covered by croplands (Allison et al., 2005; Baer et al., 2002; Fischer et al., 2007; Riley et al., 2009; Suyker and Verma, 2009). US-MOz is located in Missouri and is covered by deciduous broadleaf (Gu et al., 2012, 2006). Meteorological forcing, site information such as soil texture, vegetation cover, and satellite-derived phenology, as well as validation data sets, such as water and energy fluxes, are provided by the North American Carbon Program (NACP) site synthesis team. The site information is provided in Table 2 of Hou et al. (2012).

MOPEX basin (07147800) is the Walnut River basin in Kansas with a drainage area of 4869 m<sup>2</sup>, which is dominated by silty clay loam soil and covered by about 56 % C3 grass, 22 % C4 grass, and 20 % croplands according to the MODIS based land cover map in Ke et al. (2012). The meteorological forcing for the MOPEX basin was extracted from the phase two North America Land Data Assimilation System (NLDAS2) forcing at an hourly time step from 1979 to 2007 (Xia et al., 2012), including precipitation, shortwave and longwave radiation, air temperature, humidity and wind speed at a 1/8th degree resolution derived from the 32 km resolution 3 h North American Regional Reanalysis (NARR) following the algorithms detailed in Cosgrove et al. (2003). An area-average algorithm was then applied to the NLDAS2 forcing as inputs to CLM by treating the entire basin as a single computational unit. CLM was spun up by cycling the forcing at each site for at least five times until all the state variables reached equilibrium before any statistics were calibrated for inversion, based on the methodology described below.

### 3 Methodology

#### 3.1 Parameterization

Observational data used in parameter estimation are observed latent heat fluxes and runoff measurements, which are processed and gap-filled to obtain daily and monthly averaged data. For unknowns, we focus on the parameters that are the most identifiable from the response variables (i.e., they have significant, straightforward, and distinguishable influences on hydrologic processes, including soil hydrology and runoff generation processes). The 10 hydrologic parameters that we found to have impacts on the simulations of surface and subsurface runoff, latent and sensible heat fluxes, and soil moisture in CLM4 are  $f_{\max}$ ,  $C_s$ ,  $f_{\text{over}}$ ,  $f_{\text{drain}}$ ,  $Q_{\text{dm}}$ ,  $S_y$ ,  $b$ ,  $\Psi_s$ ,  $K_s$ , and  $\theta_s$ . Explanations of the 10 parameters and their prior information are shown in Table 1. Some vegetation parameters, such as leaf maximum carboxylation rate and the slope of the stomatal conductance, are important parameters that govern plant transpiration and modulate latent heat flux and soil moisture. In this study, we focus on the hydrological parameters that are more directly related to runoff and soil moisture, and the latter has important control over latent heat flux, and set the vegetation parameters using default values. Soil texture also affects soil evaporation and fluxes. In this study, the prior distributions for porosity, permeability, specific yield, Clapp and Hornberger exponent, and saturated soil matrix potential parameters are determined from soil texture information based on Cosby et al. (1984).

#### 3.2 Parameter calibration using least-square fitting approaches

Different approaches can be used to calibrate the selected hydrologic parameters using available observations such as runoff and surface fluxes data. The method of least squares is a standard approach to approximate the solution of over-determined systems, i.e., systems specified by more equations than unknowns. One well-known approach utilizing the least-square fitting concept is Parameter ESTimation (PEST) (Doherty, 2008), which is a general-purpose, model-independent, parameter estimation and model predictive uncertainty analysis package. Here we adopt PEST to perform single-objective least-square fitting of the observational data, with the loss function defined as the sum square of fitted residuals between calculated and observed runoff or latent heat fluxes.

There is a set of tuning parameters in the PEST inverse setup. We used the widely adopted default settings as given in the user guide. However, simulations of heat flux and runoff using the calibrated parameters show only small improvements compared to simulations using the default parameter values. PEST, using the least-square fitting approach, works well for more linear systems with monotonic relationships between unknown parameters and observable variables and weak interactions between the unknowns. For a complicated system, it is very likely that a single optimum set of parameter values does not exist and the solutions may converge to local minima. In this case, a probabilistic description of all possible solutions is more reasonable by assigning proper weights (i.e., likelihoods) to all possible solutions of parameter sets. In order to further explore the usefulness of the least-square fitting approach, the PEST-calibrated parameter estimates are used as one of the choices for initial guesses of the parameters for stochastic Bayesian updating as explained below to determine if such estimates may improve the convergence rate or robustness of the inversion results.

#### 3.3 Bayesian updating

Stochastic inversion/calibration approaches (e.g., Bayesian inference) can be used to describe the input/output uncertainties in a probabilistic manner. Bayesian inference derives the posterior probability as a consequence of two antecedents: a prior probability given prior information, and a “likelihood function” derived from a probability model for the data to be observed. Bayesian inference computes the posterior probability according to Bayes’ rule:

$$f(m|d, I) \propto f(d|m, I) \times f(m|I), \quad (1)$$

where  $f(m|d, I)$  represents the posterior pdf (probability density function) of parameter  $m$ ,  $f(d|m, I)$  denotes the likelihood of observing  $d$  given parameter  $m$ , and  $f(m|I)$  is the prior pdf of  $m$  given prior available information  $I$ .

**Table 1.** Selected hydrologic parameters in CLM4 and their prior information (from Hou et al., 2012).

Index	Symbol	Definition	Relevant process	Prior information
1	$f_{\max}$	Max fractional saturated area, from DEM	Surface runoff	Mean value taken from the default CLM4 input data set; standard deviation (SD) = 0.160; upper and lower bounds (0.01–0.907) determined from the default global data set for CLM4.
2	$C_s$	Shape parameter of the topographic index distribution	Surface runoff	Mean = 0.5 for flux towers, no SD information, upper and lower bound 0.01 and 0.9.
3	$f_{\text{over}}$	Decay factor ( $\text{m}^{-1}$ ) that represents the distribution of surface runoff with depth	Surface runoff	Hard coded to be 0.5 in CLM4. Mean = 0.5; upper and lower bounds: 0.1–5.
4	$f_{\text{drai}}$	Decay factor ( $\text{m}^{-1}$ ) that represents the distribution of subsurface runoff with depth	Subsurface runoff	Mean = 2.5; upper and lower bounds: 0.1–5.
5	$q_{\text{drai,max}}$ ( $Q_{\text{dm}}$ )	Max subsurface drainage ( $\text{kg m}^{-2} \text{s}^{-1}$ )	Subsurface runoff	Hard coded to be $5.5 \times 10^{-3} \text{ kg m}^{-2} \text{ s}^{-1}$ but typically should vary between $1 \times 10^{-6}$ and $1 \times 10^{-2}$ in hydrologic applications. Tuning range is $1 \times 10^{-6}$ to $1 \times 10^{-1}$ as suggested by NCAR.
6	$S_y$	Average specific yield	Groundwater dynamics	Hard coded to be 0.2. Based on the dominant soil type of the site. Converted to coarser soil texture classes using the USGS soil texture triangle. Mean = 0.02 for clay, 0.07 for sandy clay, 0.18 for silt, 0.27 for coarse sand; bounds are $\pm 50\%$ of the mean for the given soil texture.
7	$b$	Clapp and Hornberger exponent	Soil water	Based on the dominant soil type of the site. Used equations from Cosby et al. (1984). Mean values and SDs are from Table 5 in Cosby et al. (1984), except for SD of $\Psi_s$ , which is from Table 4 in Cosby et al. (1984).
8	$\Psi_s$	Saturated soil matrix potential (mm)	Soil water	
9	$K_s$	Hydraulic conductivity ( $\text{mm s}^{-1}$ )	Soil water	
10	$\theta_s$	Porosity	Soil water	

\* Reproduced with permission from American Geophysical Union.

We assume that the forward models (e.g., CLM4) are being characterized by  $d_{ij}^* = g_{ij}(m) + \varepsilon_{ij}$ , where  $m$  represents the vector of model parameters,  $d_{ij} = g_{ij}()$  is the forward model,  $\varepsilon_{ij}$  is the difference (i.e., residual) between the model  $d_{ij}$  and observation  $d_{ij}^*$ ,  $i = 1, \dots, K$ ;  $j = 1, \dots, N_i$ ,  $K$  being the number of data types and  $N_i$  being the number of observations for the  $i$ th data type. The residuals include both measurement errors in observations (e.g., heat fluxes) and modeling errors due to model parameters. As discussed above, model errors due to forcing data and model structural uncertainty are not addressed in this study. However, we used forcing data developed and adopted by the community (NACP and NLDAS2) that reduce some uncertainty by using multiple observational data sets with adjustments to account for local factors such as topography.

With the underlying assumptions that  $\varepsilon_{ij}$  is normally distributed with variance  $\sigma_{ij}^2$ , and the distributions are independent, the likelihood function can be represented as (Hou et al., 2006)

$$f_{D|M,\Sigma,I}(d^*|m, \sigma, I) = \prod_{i=1}^K \prod_{j=1}^{N_i} \left[ \frac{1}{\sqrt{2\pi}\sigma_{ij}} \exp \left\{ -\frac{1}{2\sigma_{ij}^2} [d_{ij}^* - g_{ij}(m)]^2 \right\} \right] \quad (2)$$

The posterior distributions of the input parameters are the products of the priors and the likelihood functions. As more information comes in, Bayes' rule can be applied iteratively, that is, after observing some evidence, the resulting posterior probability can be treated as a prior probability, and a new posterior probability is computed from the new evidence.

### 3.4 Sampling methods

An efficient sampling approach is important for the success of Bayesian inversion, especially when the forward modeling computational demand is high, the parameter dimensionality is high, or the parameters are weakly identifiable. In this study, the Metropolis–Hasting sampling method is used to draw samples from the joint posterior distribution functions (Hastings, 1970; Hou, 2008; Metropolis et al., 1953). The procedure is as follows.

1. Initialize a random vector  $m$  from the prior distributions  $m_i^{(0)}$ ,  $i = 1, \dots, p$ , where  $p$  is the number of parameters.
2. Generate a random variable  $m_i^*$ ,  $i = 1, \dots, p$  from the proposal distributions, and calculate the following ratio (note the probabilities in the formula are calculated using Eq. 2):

$$\alpha = \min \left( p_{\text{ra}}, \frac{\text{prob}(m_i^*|m_1^{(1)}, m_2^{(1)}, \dots, m_{i-1}^{(1)}, m_{i+1}^{(1)}, m_{i+2}^{(1)}, \dots, m_p^{(1)})}{\text{prob}(m_i^{(0)}|m_1^{(0)}, m_2^{(0)}, \dots, m_{i-1}^{(0)}, m_{i+1}^{(0)}, m_{i+2}^{(0)}, \dots, m_p^{(0)})} \right), \quad (3)$$

where  $p_{ra}$  is reference acceptance probability.

3. Generate a random value  $u$  uniformly from interval  $(0,1)$ .
4. If  $\alpha > u$ , let  $m_i^{(1)} = m_i^*$ ; otherwise, let  $m_i^{(1)} = m_i^{(0)}$ .

Repeating steps (b) to (d) by replacing index ( $k$ ) with index ( $k+1$ ), we can obtain many samples as follows:  $\{(m_i^{(k)}): i = 1, \dots, p, k = 0, 1, \dots, n\}$ , where  $n$  is the number of sample sets. From the procedure, we can see that the value  $m_i^{(k)}$  only depends on the current state of  $\mathbf{m}$ , but not the previous states; therefore, these samples form a Markov Chain. The summary statistics of the posterior samples during the burn-in period (e.g., the last 400 sample sets) can be used to check the convergence with a tolerance relative error (e.g., 0.001), and posterior distributions can be derived from the samples.

## 3.5 Inversion setup

### 3.5.1 Choices of initial values (default, mean, PEST estimates)

Initial values have little impact on precision and accuracy of a robust optimization algorithm, but affect the convergence speed. In order to increase the efficiency of inverse modeling, we compared choices of initial values including the default parameter values in CLM4, the mean values between the prior bounds, as well as the PEST estimates calibrated using observational data at each study site. The default and mean values are based on prior information in Table 1. Tests show that the convergence speed of the MCMC-Bayesian calibration is not affected by the choices of initial values, although the PEST estimates slightly reduce the discrepancies between calculated and observed responses compared to using the default parameter values for CLM4 model calculations.

### 3.5.2 Choices of proposal distribution widths (automatic versus multi-chain comparison)

The convergence speed (i.e., the number of steps needed to obtain consistent posterior statistics of the unknown parameters) and accuracy of parameter estimation is associated with several tuning parameters of the inversion setup. Proposal distribution width, an important tuning parameter of the MCMC algorithm, controls the searching precision and speed. We compared different widths, i.e., 1/2, 1/4, 1/8, and 1/16, relative to the prior bounds of the unknown parameters. The widths correspond to the 99 % confidence interval of Gaussian proposal distributions. Optimized convergence speed and stable results can be achieved with a proposal width of 1/8 relative to the prior bounds. To compare the performances of different tuning parameters, and to reduce

the time of convergence, we take advantage of high performance computing resources to conduct multi-chain calibrations simultaneously.

### 3.5.3 Reference acceptance probability

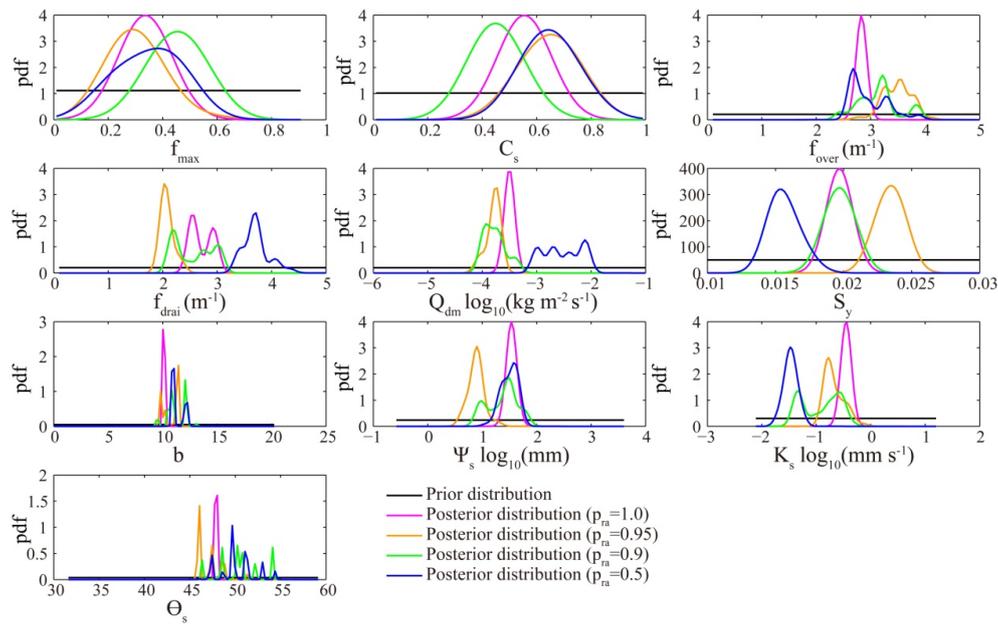
In this study, we also evaluated the effects of using different acceptance probabilities for newly generated parameter samples in the MCMC process. We named the criterion as reference acceptance probability, a trade-off between convergence speed and accuracy of parameter estimates. If the acceptance probability of new sample sets is greater than the product of reference acceptance probability and the acceptance probability of prior sample, we accept the new sample. Specifically, four reference acceptance probabilities, i.e., 1.0, 0.95, 0.9, 0.5, are adopted to set the multi-model for inverse modeling of CLM. Probabilistic model averaging can then be used to integrate the different sets of parameter predictions, where the weights are not determined arbitrarily, but rather based on the posterior probabilities of all the models sharing the same unknowns and observational data in the inverse setup. The integrated results might have larger spread but the estimates are generally more unbiased.

## 3.6 Sensitivity to data frequency, observation type, and site conditions

Daily and monthly data (runoff and heat flux) are both used in parameter calibration, and the performances are compared to study the impact of data temporal resolution on parameter inference. Compared to monthly data, daily data include higher-frequency characteristics of temporal variations in the observations; however, adjacent observations tend to have more information redundancy and data quality is an issue. A stochastic inversion approach usually provides parameter calibrations with lower uncertainty (more “precise”) as more and more data are used, but measurement errors could lead to over-fitting of errors and therefore biased estimates. Through comparison, we can identify the most appropriate observing timescale for calibration to improve CLM4 simulations.

Different model responses associated with different components of CLM have different sensitivities to the unknown parameters. We conduct MCMC-Bayesian inversion at the US-ARM site and at the nearby MOPEX basin (basin ID: 07147800). Although a basin is usually not homogeneous in climate and land cover, the two sites are the closest in geographical proximity among all the Ameriflux first priority and MOPEX sites, and they have similar climate and land surface conditions. By comparing the inversion results using observed heat flux and runoff, we can evaluate the impacts of using different types of observations on parameter inference in CLM.

Calibrations are also done at flux tower sites (US-ARM vs. US-MOz) with different soil and vegetation conditions to study the impacts of land surface conditions on parameter



**Fig. 1.** Posterior distribution of the parameters with four different reference acceptance probabilities ( $p_{ra}$ ) using monthly heat flux data at the US-ARM site.

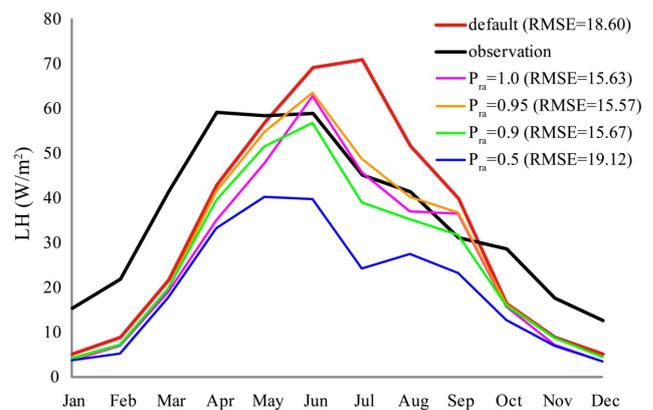
inference of CLM. Soil, climate, and vegetation conditions control water and energy fluxes, such as infiltration, evaporation, and surface radiation; therefore, they can affect the overall parameter identifiability and inversion results. Calibration results can also be significantly affected by the quality/accuracy of the observational and forcing data, as well as the model's structural design and parameterization.

## 4 Results of full-set parameter inversion

### 4.1 Parameter inversion at flux tower sites using heat flux observations

#### 4.1.1 Posterior distributions of input parameters and simulated heat flux from use of monthly data

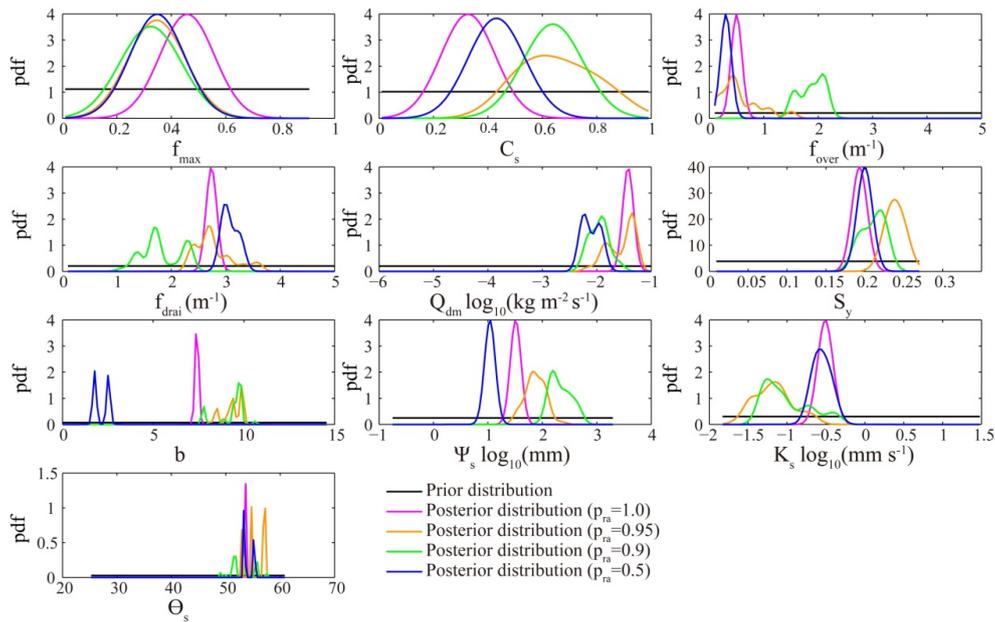
Using monthly heat flux data from the 2003–2006 period at the US-ARM site, we conducted inversion of ten CLM parameters with four reference acceptance probabilities. Figure 1 shows the prior and posterior distributions of the parameters, where the prior distributions are derived from prior information, and the posterior distributions are derived based on the last 200 samples of inverse modeling. Note that three parameters  $Q_{dm}$ ,  $\Psi_s$ , and  $K_s$  vary by several orders of magnitude, and are  $\log_{10}$ -transformed. Posterior distributions with different reference acceptance probabilities are generally consistent, except for those parameters that are less identifiable due to nonlinearity and non-uniqueness issues. Our previous study (Hou et al., 2012) showed that  $f_{over}$ ,  $f_{max}$ ,  $K_s$ , and  $\theta_s$  have weak impacts on model simulations at the US-ARM site. Therefore, the posterior distributions of these



**Fig. 2.** Simulated heat fluxes using the posterior estimates of parameters with monthly observations at the US-ARM site.

parameters are wide and non-informative as expected. This demonstrates the necessity of performing parameter screening such that the inverse problems can be less ill-posed. As shown in Eq. (3), a low reference acceptance probability  $p_{ra}$  means that the rejection standard and searching ranges are relaxed. As more potential estimates are identified and accepted, the bounds of posterior distributions increase, and multi-modal behaviors occur, especially for  $\theta_s$ . The posterior means/modes of the estimated parameters shifted farther from or closer to the prior means, particularly for  $f_{over}$  and  $\theta_s$ .

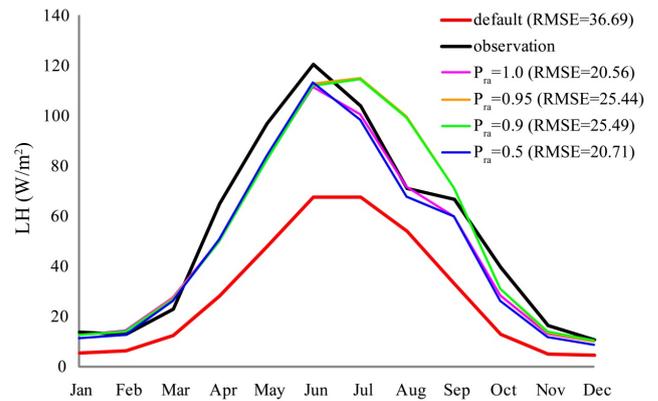
Figure 2 shows the simulated monthly mean heat fluxes using posterior estimates of parameters with four reference acceptance probabilities using monthly observations at the



**Fig. 3.** Posterior distribution of the parameters with four reference acceptance probabilities using monthly heat flux data at the US-MOz site.

US-ARM site. The black line shows the monthly mean heat fluxes obtained from observations of the 2003–2006 period, and the red line shows the calculated monthly mean heat fluxes using default parameters based on prior information. Simulations using default parameter values overestimate the heat fluxes in summer (from May to September), and underestimate in winter at the US-ARM site. When the reference probabilities are 1.0, 0.95, and 0.9, the posterior estimates of parameters all significantly improve the heat flux simulation in summer, and the posterior model probabilities (i.e., the product of Gaussian probabilities of misfits between calculated and observed responses) are 0.730, 0.730, and 0.728 respectively, which are greater than 0.636 for the default parameter values. Note that the Gaussian probability of the default or a posterior parameter set is calculated using Eq. (2). The estimates with reference acceptance probability of 0.5 noticeably deviate from other inversion estimates, and tend to result in underestimates of simulated heat fluxes in summer. None of the parameter estimates is able to yield much better fits during winter, which might be due to errors in the observed heat fluxes, errors in the CLM forcing data, and/or under-representation of the complicated physical processes using the current parameterization schemes.

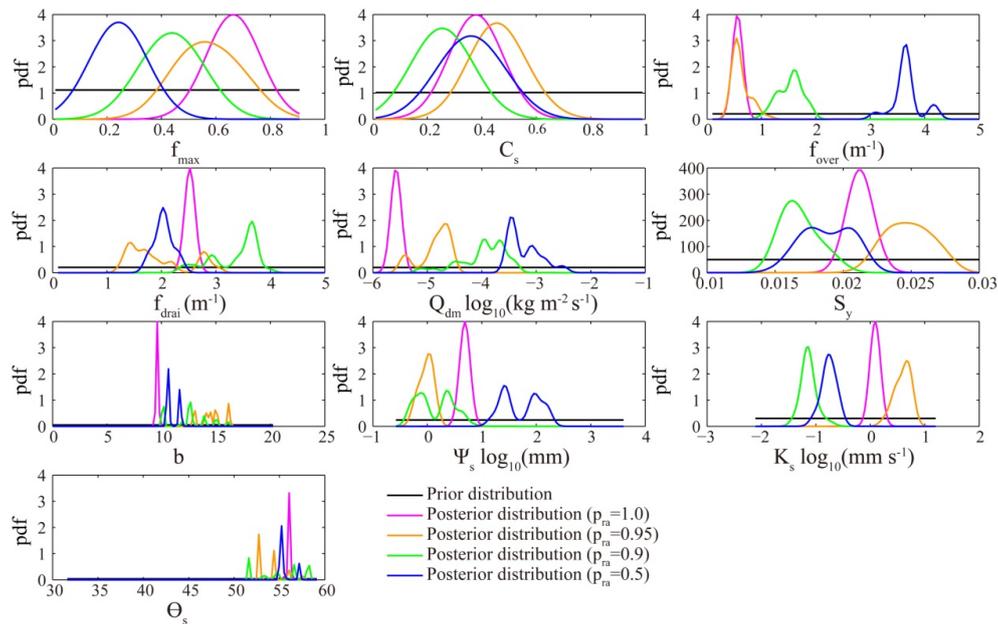
Similarly, inversion was also performed using monthly heat flux data from the 2004–2007 period at the US-MOz site with different soil and vegetation cover from US-ARM site. Figure 3 shows the posterior distributions of ten parameters with four reference acceptance probabilities. They show similar patterns for different reference acceptance probabilities, except for the parameter  $b$ . Even when the reference acceptance probability is 0.5, the inversion yields reasonable



**Fig. 4.** Simulated heat fluxes using the posterior estimates of parameters at the US-MOz site.

parameter estimates for most parameters except for  $b$  and  $\Psi_s$ . The relaxed rejection standard also leads to multi-modal, extended posterior distribution bounds, and more potential parameter estimates. It is noted that the posterior distribution of  $f_{max}$ ,  $C_s$ ,  $f_{drai}$ ,  $b$ , and  $\Psi_s$  are around the median values of the prior bounds.

Figure 4 shows the calculated monthly mean heat fluxes using the posterior parameter estimates from monthly observation at the US-MOz site. It is obvious that the default simulation underestimates the heat flux over all seasons. From January to June, all posterior estimates with different reference acceptance probabilities can significantly improve the simulation of heat flux, except for some small underestimations in April, May, and June. From July to



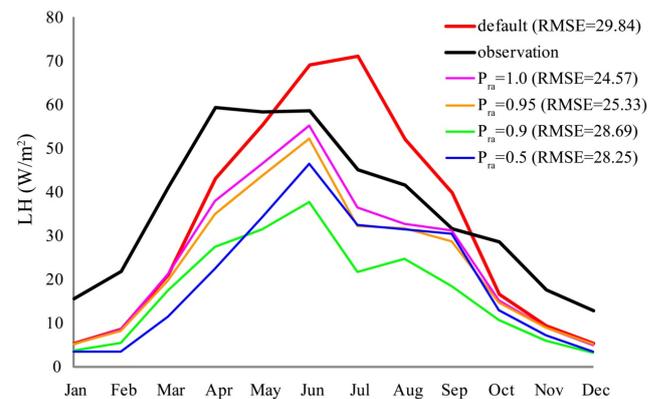
**Fig. 5.** Posterior distribution of the parameters with four reference acceptance probabilities using daily heat flux data at the US-ARM site.

December, the posterior estimates with reference acceptance probabilities of 1.0 and 0.5 are similar and close to observations, while the other two overestimate the heat flux a little in July and August. As a whole, all estimates of the inverse modeling can improve the simulation of heat flux over all seasons, and the posterior model probabilities are 0.893, 0.889, 0.886, and 0.892 respectively, which are greater than 0.876 of the default simulation. Differences in the posterior distributions with different reference acceptance probabilities are small.

#### 4.1.2 Posterior distributions of input parameters and simulated heat flux from use of daily data

Figure 5 shows the posterior distribution of model parameters with four reference acceptance probabilities using daily heat flux data from the 2003–2006 period at the US-ARM site. The posterior distributions disperse over the prior bounds for most parameters. Among the four sets of posterior distributions, reference acceptance probability of 1.0 and 0.95 identify similar bounds, while the other two sets yield different results, particularly for  $f_{\max}$ ,  $f_{\text{over}}$ ,  $Q_{\text{dm}}$ , and  $K_s$ . Moreover, multi-modal distributions occur for most parameters when the rejection standard is relaxed.

Figure 6 shows the calculated monthly mean heat fluxes using the posterior estimates of parameters from daily observations at the US-ARM site. The posterior estimates of parameters also improve the heat flux in summer. The acceptance probabilities of the simulations with four parameter sets are 0.636, 0.618, 0.584, and 0.593 respectively, which are all greater than 0.579 of the default simulation.



**Fig. 6.** Simulated heat fluxes using the posterior estimates of parameters with daily observations at the US-ARM site.

The posterior distributions of the parameters with the four reference acceptance probabilities using daily heat flux data from the 2004–2007 period at the US-MOz site show that the posterior distributions are more consistent for  $C_s$ ,  $f_{\text{over}}$ ,  $Q_{\text{dm}}$ ,  $S_y$ , and  $\Psi_s$ , but dispersed for  $b$ ,  $K_s$  and  $\theta_s$ . When the rejection standard is relaxed, the posterior bounds can become much wider, especially for  $f_{\text{drai}}$ ,  $Q_{\text{dm}}$ ,  $b$ ,  $K_s$  and  $\theta_s$ . In winter, all simulations of heat flux using the posterior estimates are close to observations. In summer, posterior estimates with reference acceptance probabilities of 1.0 and 0.95 significantly improve the heat flux simulation.

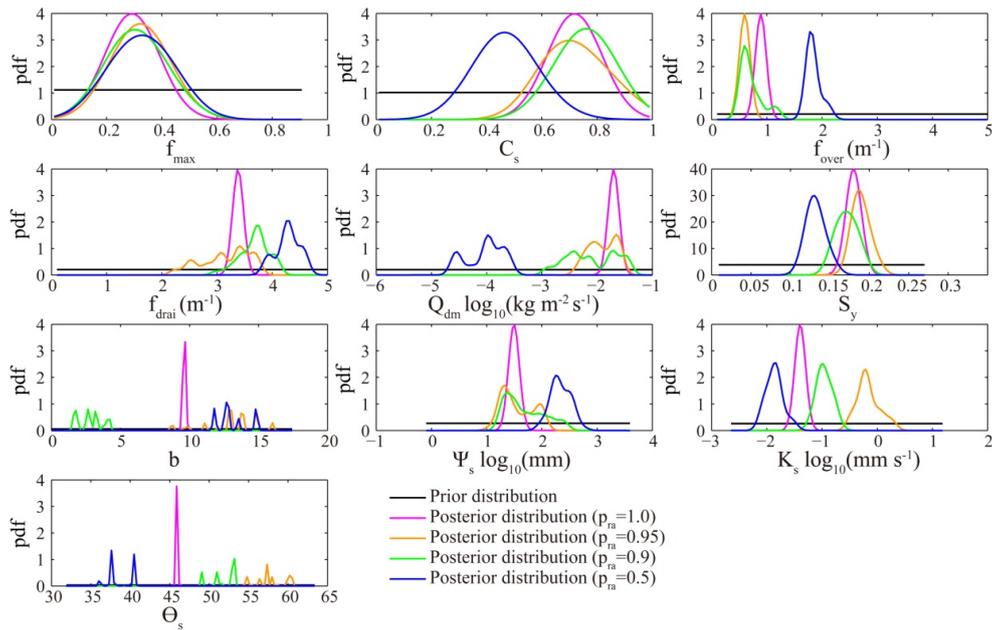


Fig. 7. Posterior distribution of the parameters with four reference acceptance probabilities using monthly runoff data at the MOPEX basin.

4.2 Parameter inversion at MOPEX sites using runoff observations

Runoff observations are also used in the inverse modeling of CLM. Figure 7 shows the posterior distribution of the parameters with four reference acceptance probabilities using monthly runoff data from the 2002–2005 period at the MOPEX basin close to US-ARM. Posterior distributions with strict reference acceptance probabilities (e.g., 1.0, 0.95 and 0.9) have consistent patterns for most parameters, except for  $b$ ,  $K_s$ , and  $\theta_s$ . It is interesting to see that  $f_{max}$  is identically estimated by inversions with different reference acceptance probabilities. When the rejection standards are relaxed, the bounds of posterior distributions of most parameters become wider, and multi-modal patterns occur.

Figure 8 shows the calculated monthly mean runoff using the posterior estimates of parameters from monthly observations at the MOPEX basin. The default simulation barely shows any variability of runoff. The posterior estimates significantly improve the runoff simulations in all seasons, albeit larger variability than observations is noted from July to October. Further analysis (Fig. 15) shows large peaks in the observed runoff in June and July, which the CLM4 model simulations cannot match well by adjusting model parameter values. This suggests some systematic biases in the model parameterizations that cannot be fully addressed by parameter calibration. However, we cannot exclude the possibility of errors in either the external forcing or observational heat fluxes. For example, the atmospheric forcing data are calculated over the 1/8th degree grid, which may underestimate rainfall intensity for heavy precipitation events, leading to

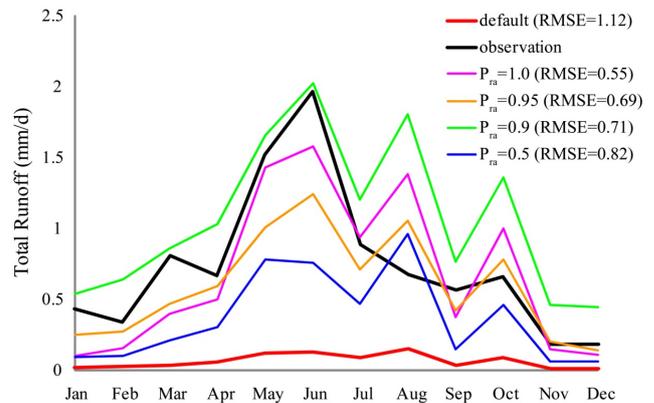


Fig. 8. Simulated runoff using the posterior estimates of parameters at the MOPEX basin.

underestimation of runoff peaks in the summer and accumulation of soil moisture for runoff generation in subsequent periods.

The acceptance probabilities are 0.846, 0.767, 0.758, and 0.737, all greater than 0.707 of the default runoff simulation. Among the four sets of simulations based on inversion, more stringent sample rejection criterion results in a better match between the simulated responses with observations. Because hydrological parameters are more sensitive to hydrological processes, this conclusion is not surprising for inversion with runoff. In cases using heat fluxes for calibration, we also found that the RMSEs of flux simulations using the posterior estimates with the reference acceptance probability of 1.0 are better than others. It is worth mentioning that for

observational data that are more directly related to the processes influenced by the model parameters, a more stringent rejection criterion can help to improve calibration accuracy and precision. For indirect information, a relaxed criterion is helpful to avoid biased estimates.

For comparison with inversion using monthly runoff data, we also performed inversion using daily runoff data from the 2002–2005 period at the MOPEX basin. The posterior distributions with different reference acceptance probabilities disperse over the prior bounds, except for  $S_y$ . The calculated monthly mean runoff values using the posterior estimates of parameters from daily observations are reasonable. Compared to using monthly runoff data, the big fluctuations disappear when using daily runoff. Posterior estimates with reference acceptance probabilities of 1.0 and 0.95 can significantly improve the runoff simulation over all seasons, but underestimate the runoff in summer, and overestimate the runoff in winter.

## 5 Results of subset parameter inversion

Our global sensitivity analyses across 13 flux towers and 20 MOPEX basins (Hou et al., 2012; Huang et al., 2013) suggest that simulated LH and runoff are most sensitive to three subsurface parameters:  $f_{\text{drai}}$ ,  $Q_{\text{dm}}$ , and  $S_y$ . Our full-set parameter inversion shows that the posteriors for the most sensitive parameters are more consistent than the unidentifiable parameters, which is expected. Because the other parameters are less identifiable, the inverse problem will be less ill-posed by fixing the trivial parameters. In this section, we test the feasibility of only inverting a subset of identifiable parameters to determine if similar or improved model skill may be achieved compared to using the full-set parameter inversion results. We first conducted the inversion with a reduced set of parameters using monthly observation at the two flux tower sites and the MOPEX basin. Only the most identifiable parameters are calibrated, the other parameters are fixed as the default values in CLM4.

Figure 9 shows the posterior distribution of the reduced set of parameters at the US-MOz site. Compared with the results of ten parameters (see Fig. 3), the medians of the posterior distribution of  $f_{\text{drai}}$  and  $Q_{\text{dm}}$  are smaller, the width of the posterior bounds of  $f_{\text{drai}}$  and  $S_y$  is unchanged, while that of  $Q_{\text{dm}}$  expands.

Figure 10 shows the calculated monthly mean heat flux using posterior estimates of parameters at the US-MOz site. The simulations using posterior estimates of the reduced parameter set significantly improve the heat flux simulation over all seasons, and are similar to the results of inversion with ten parameters.

Inversion at the US-ARM site also shows that in general, the posterior bounds start to narrow, and the multi-modal patterns disappear, compared to the inversion results with the full-set of parameters. Using posterior estimates of the re-

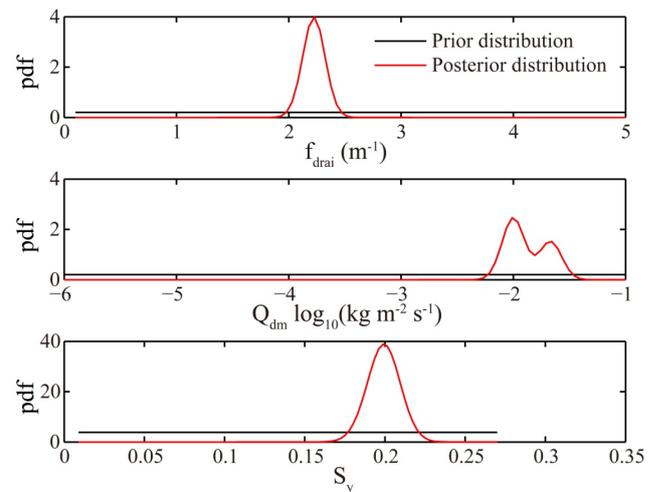


Fig. 9. Posterior distribution of the reduced parameter set from previous sensitivity analysis at the US-MOz site.

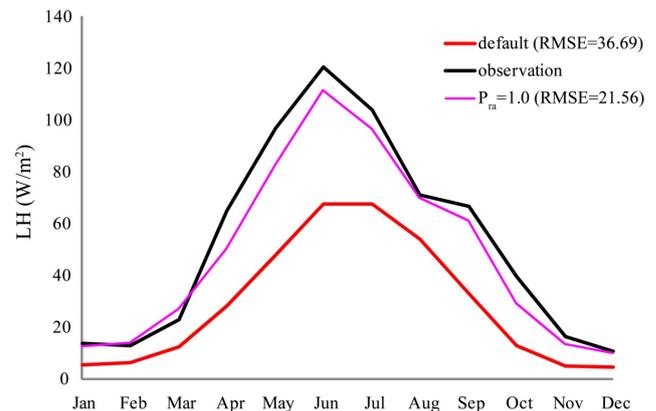
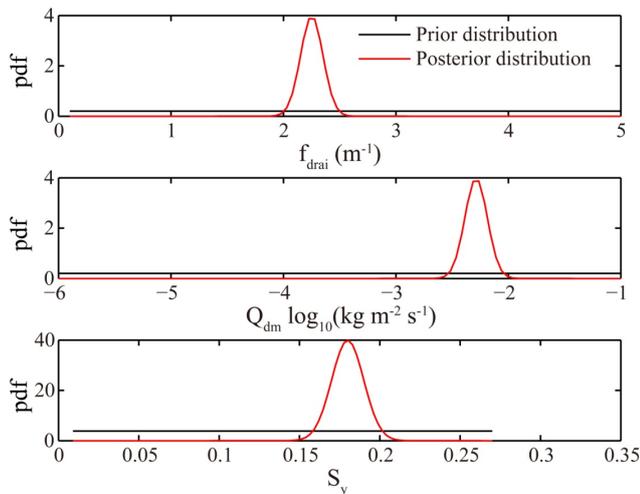


Fig. 10. Simulated heat fluxes using the posterior estimates of parameters at the US-MOz site.

duced parameter set can significantly improve the latent heat flux simulations compared to the results using the full-set of parameters, especially from October to December, and from January to May. Results from US-ARM and US-MOz are consistent, so only one set of figures is shown.

Figure 11 shows the posterior distribution of the reduced parameters at the MOPEX basin. Compared with the results of ten parameters (see Fig. 7), the medians of posterior distribution of  $f_{\text{drai}}$  and  $Q_{\text{dm}}$  are smaller, while that of  $S_y$  does not change, and the widths of posterior distribution of all parameters stay the same.

Figure 12 shows the calculated monthly mean runoff using posterior estimates of parameters at the MOPEX basin. Inversion with reduced parameters improves the runoff simulation, which is similar to the posterior simulation with ten parameters, while the simulation with the default parameter is far away from observation.



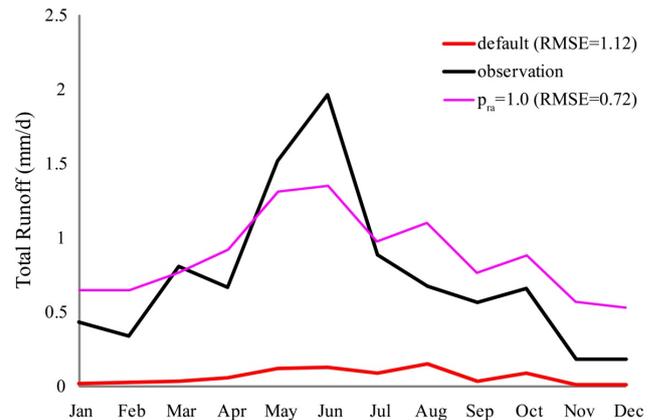
**Fig. 11.** Posterior distribution of the reduced parameter set from previous sensitivity analysis at the MOPEX basin.

Overall, inverse modeling with a reduced set of parameters identified from previous sensitivity analysis yield comparable predictions obtained from the full-set parameter inversion, and since the inverse problems become less ill-posed with fewer unknowns, the convergence of inverse modeling is faster and the resulting posteriors are more consistent without multi-modal patterns. However, the simulations of heat flux at the US-MOz and runoff at the MOPEX basin are comparable between the inversion for the reduced and full set of parameters. Theoretically, one may expect improvement using the reduced parameter set because the inversion becomes less ill-posed, but in practice, getting a faster convergence of the solution may be the main advantage, which is important especially when calibrating parameters for computationally intensive models.

## 6 Discussion

### 6.1 Impacts of temporal resolution of heat flux observation on inverse modeling

Both monthly and daily heat flux data have been used in the inverse modeling at selected flux tower sites. Using monthly observation data, the inversion with reference acceptance probabilities (except for  $p_{ra}$  of 0.5 at US-ARM) is able to identify proper parameter estimates to improve heat flux simulation. Using daily observation data, inversion improves the heat flux simulation only with reference acceptance probabilities of 1.0 and 0.95, indicating that using data of higher temporal resolution might need a relatively more stringent acceptance criterion (i.e., higher  $p_{ra}$ ). Comparing Figs. 1 and 5, inversion using daily instead of monthly observations favors values of  $f_{over}$ ,  $Q_{dm}$ ,  $\Psi_s$  towards the lower bounds at the US-ARM site. At the US-MOz site, inversion using daily



**Fig. 12.** Simulated runoff using the posterior estimates of parameters at the MOPEX basin.

observation also favors values of  $f_{over}$  and  $Q_{dm}$  towards the lower bounds. Hence, finer temporal resolution of observation data favors smaller  $f_{over}$  and  $Q_{dm}$ . For specific sites, it may also lead to changes of other parameters.

Regarding inversion using runoff data, we also found that finer temporal resolution of runoff observations leads to more dispersion of the posterior distributions of most parameters, except for  $f_{over}$ ,  $S_y$  and  $\theta_s$ . Using monthly observation data, inversions with all reference acceptance probabilities are able to improve monthly mean runoff simulations over all seasons. However using daily observation data, inversion improves runoff simulation only with reference acceptance probabilities of 1.0 and 0.95.

Overall, finer temporal resolution of observation data leads to more dispersion of the posterior distributions and increases the risk of using a relaxed rejection standard. These are likely related to increased measurement errors, data redundancy, and over-fitting with higher temporal frequency observations.

It is worth pointing out that temporal patterns in observational data may affect the performance of inversion. For example, the smaller seasonal variations of LH at US-ARM present unfavorable conditions for calibrations compared to US-MOz. At US-ARM, precipitation has a larger seasonal cycle with relatively dry winters and wet summers. Because annual rainfall is lower, LH is more limited by soil moisture availability. In contrast, US-MOz receives more rainfall throughout the year so LH is more controlled by solar radiation and hence peaks in July. Because US-ARM is more moisture limited, LH is more sensitive to model parameterization of soil hydrology. The inability of the model to capture the correct timing of the peak LH despite parameter calibration suggests that there may be structural limitations of the parameterizations used in CLM4 that cannot be adequately addressed by inverse modeling of uncertain parameters.

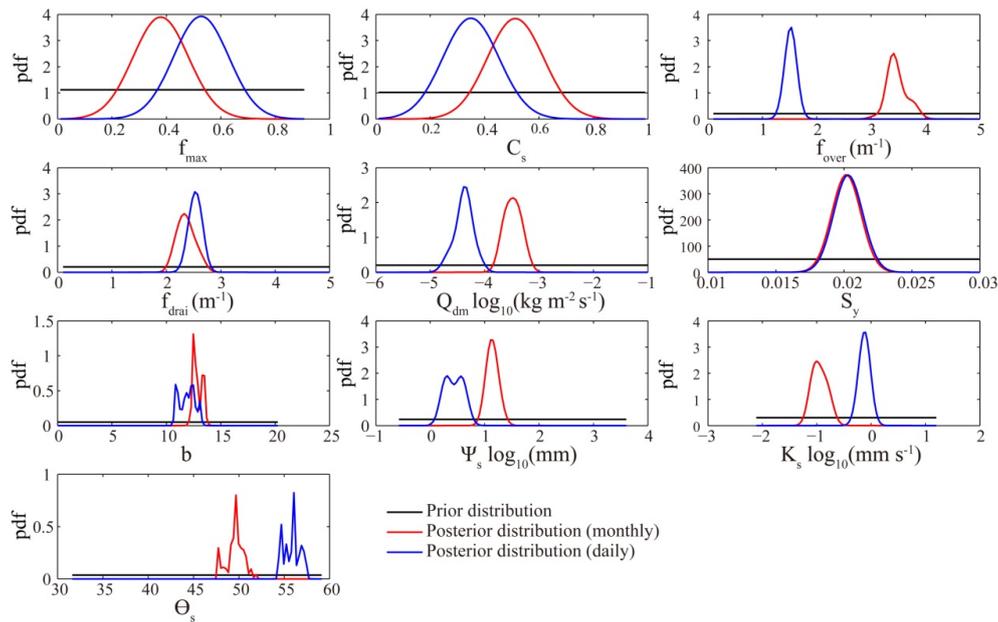


Fig. 13. Posterior distribution of the parameters through Bayesian model averaging at the US-ARM site.

## 6.2 Impacts of soil and vegetation cover on inverse modeling

Compared to US-MOz, inverse modeling at the US-ARM site identifies smaller  $Q_{dm}$ , and greater  $f_{over}$  and  $\theta_s$ . In addition, the bounds of posterior distribution identified by the inversion show more consistency across different reference acceptance probabilities for  $f_{over}$ ,  $f_{drai}$ ,  $Q_{dm}$ ,  $b$ , and  $\Psi_s$  at US-ARM than US-MOz, especially with monthly heat flux data. At US-MOz, the bounds of posterior distribution are mainly consistent for  $f_{max}$ ,  $f_{over}$ ,  $Q_{dm}$ ,  $S_y$ , and  $\theta_s$ . These inversion results are consistent with the sensitivity analysis performed by Hou et al. (2012), which shows larger sensitivity to the respective parameters at the two sites related to the soil and vegetation properties. The best estimated parameters are different at sites with different climate, land use, and soil conditions; hence soil and vegetation cover may inform the selection of sensitive parameters that can be used in reduced parameter sets for inverse modeling. It is therefore necessary to analyze parameter sensitivity and identifiability across the flux tower sites and MOPEX basins and classify them into different groups/classes with similar climate and soil conditions, and then evaluate parameter transferability within each class or between classes through inverse modeling studies.

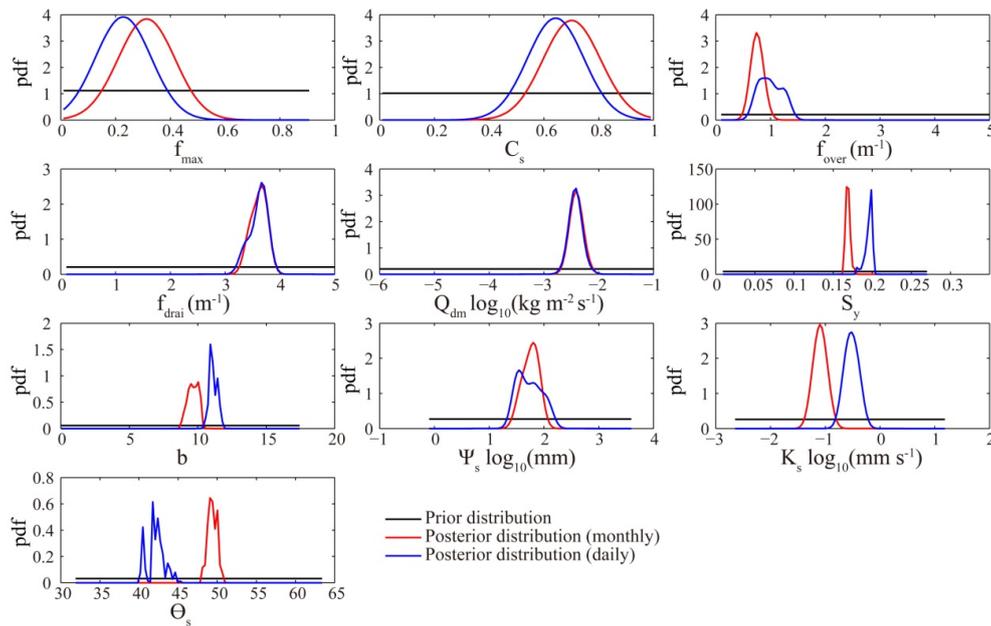
## 6.3 Impacts of reference acceptance probability

In this study, we set the reference acceptance probability in the inverse modeling to relax rejection standard to allow more freedom in searching for optimal parameter estimates. However, relaxing the rejection standard leads to broadening of the bounds of posterior distribution and multi-modal be-

haviors. That is, the posterior estimates tend to be more “accurate” but less “precise”, and the corresponding inversion process usually takes longer to converge.

## 6.4 Impacts of different types of observations on inverse modeling

Inverse modeling using heat flux at US-ARM and runoff at the MOPEX basin, which is located close to US-ARM, provides an opportunity to assess the impacts of data type on inverse modeling. Note that a basin is not homogeneous in climate and land cover, and there are differences in spatial scales and heterogeneity between processes that influence the flux tower site and the MOPEX basin. Nevertheless, they are geographically close and have comparable climate and soil conditions, so it is possible to resolve the impacts of using different data types of observations for calibration. Comparing Figs. 1 and 7, the posterior distributions that optimize the simulations of heat flux can differ from those that optimize the simulations of runoff. Since the calibrated model parameters are directly related to the soil’s hydrological processes including surface and subsurface runoff, it is not surprising that model inversion leads to more significant improvements in runoff (Fig. 8) than heat flux (Fig. 2) compared to simulations that use the default parameter values. It is also possible that the default parameters that control hydrological processes may be more poorly defined in CLM in general or for the particular sites being evaluated so that there is more room for improvements. The simulations of heat flux can nevertheless be improved by inverting hydrologic parameters because surface heat flux is influenced by soil moisture, which is closely related to runoff processes. The improvement in



**Fig. 14.** Posterior distribution of the parameters through Bayesian model averaging at the MOPEX basin.

simulations of heat flux is particularly more noticeable at US-MOz than at US-ARM, where structural uncertainty may have larger effects due to its stronger dependence of heat fluxes on soil moisture availability. In addition, being a cropland site, LH at US-ARM, may be influenced by other factors such as crop management that are not well represented in the model.

Although inverse modeling leads to larger improvements in the runoff simulations compared to the simulations using the default parameter values, the runoff simulations with the posterior estimates still deviate quite significantly from the observed runoff in late summer and fall. This indicates possible model biases in runoff due to model structural errors in the hydrologic parameterizations or the quality and spatial resolution of the forcing data discussed earlier.

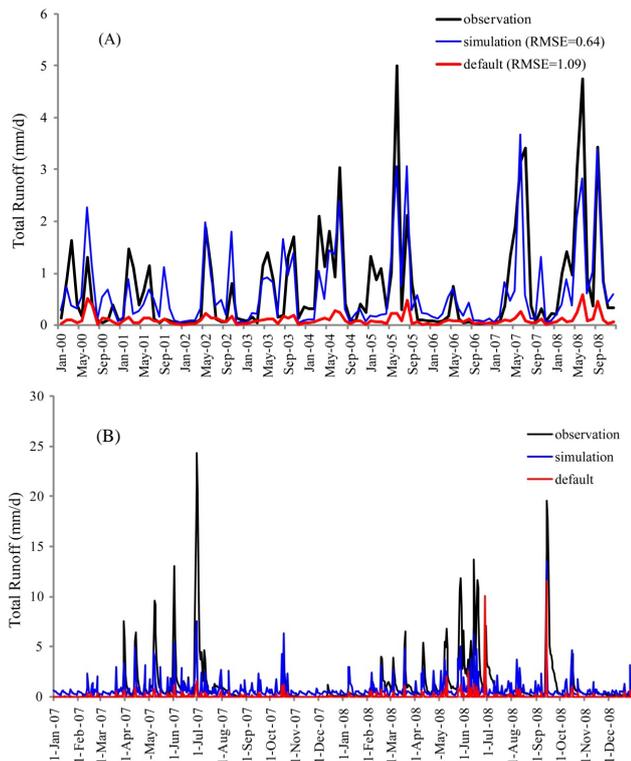
The LH flux can be measured by flux tower, which is representative of more local conditions, while runoff is a composite response of a drainage basin of a large area. These provide complementary information for model calibration. We did not perform an integrated inversion because energy flux observations are available only at the flux tower sites (US-ARM and US-MOz), and runoff data are only available at the MOPEX basins, each representing the best available observations at each scale, respectively. Therefore, in this study, we focused on evaluating the potential of improving CLM simulations using the best available observations at appropriate scales. However, our study clearly demonstrated the potential of multi-objective calibration, which will be attempted in future studies.

### 6.5 Improvements through Bayesian model averaging

For each reference acceptance probability for the MCMC-Bayesian inversion, one can obtain a set of posterior distributions of the unknowns. Bayesian model averaging is used to integrate the different sets of predictions by weighting the posteriors according to their posterior model probability.

By integrating inversion results of different reference acceptance probabilities, Bayesian model averaging produces smoother posterior distributions. Figure 13 shows the posterior distributions of the parameters through Bayesian model averaging at the US-ARM site. The black lines represent the prior distributions based on prior information. The red and blue curves represent the posterior distributions of the parameters using monthly and daily heat flux observations, respectively. These two sets of posterior distributions are similar to each other for most parameters, except for  $f_{over}$ . Daily heat flux favors smaller  $C_s$ ,  $f_{over}$ ,  $Q_{dm}$ ,  $\Psi_s$ , and greater  $f_{max}$ ,  $K_s$ ,  $\theta_s$ . The posterior distributions using monthly and daily observations at the US-MOz site are also similar, but daily heat flux favors smaller  $C_s$ ,  $f_{over}$ ,  $f_{drai}$ ,  $Q_{dm}$ ,  $S_y$ ,  $\Psi_s$  and greater  $f_{max}$  and  $K_s$ .

Figure 14 shows the posterior distribution of the parameters through Bayesian model averaging at the MOPEX basin. The posterior distributions using monthly and daily runoff observations are also similar. Daily observation favors smaller  $\theta_s$  and greater  $S_y$ ,  $b$  and  $K_s$ . It is noted that the differences between the posterior distributions from monthly and daily data are even smaller from inversions using runoff compared to inversions using heat flux, especially for  $f_{drai}$  and  $Q_{dm}$ . This may be related to the characteristic timescales of



**Fig. 15.** Comparison among the observations, default and optimal simulations of (A) monthly and (B) daily runoff, during the inversion (2002–2005) and validation (2000–2001 and 2006–2008) periods at the MOPEX basin. The RMSEs are calculated for the validation periods.

the physical processes. Surface heat flux may have less day-to-day variability (hence larger data redundancy) compared to runoff, which responds more directly to precipitation that has larger temporal variability than temperature during the wet season. These differences could be site and season dependent so analyses over a larger number of sites can provide further insights on the sensitivity of model inversion to data temporal frequency.

Model integration represents a compromise of all possibilities of the inversion setup. In general, it results in “safer” (i.e., more likely to be unbiased) estimates but lower resolution (i.e., wider posterior distributions).

## 6.6 Model validation

In the above analyses, we compared observed and model calibrated responses to check whether smaller misfits can be achieved through the calibration process, and to evaluate the different calibration power of runoff versus heat flux observations, monthly versus daily data, and different tuning parameters. In an inverse study, it is important to validate the inversion approach. It is straightforward to validate the results when true values of the unknown input parameters are available. Otherwise, people may design “synthetic” models

by assuming the “true” parameter values are available, then “generate” the corresponding “true” responses, which are then used for testing the inversion approach. An alternative way of validation is to separate the data set to training (for calibrating the parameters) and testing periods, assuming the parameters are intrinsic to the system and not time-varying. Fig. 15a and b show the observations as well as model simulated monthly and daily runoff calculated using default and optimal parameter values. The inversion (training) time period is 2002–2005, and validation periods are 2000–2001 and 2006–2008. The RMSEs are calculated for the validation periods only. We found that RMSEs are reduced more for monthly data than for daily data. Sub-monthly variations involve more internal/external factors and processes that are more complicated. It is harder to capture the higher frequency components of the output variability. In general, runoff calculations using optimal parameters from the training period can significantly improve the model misfits during the testing periods, and the major patterns of inter-annual and seasonal variability are well captured.

## 7 Conclusions

In this study, we demonstrated the possibility of inverting hydrologic parameters using surface flux and runoff observations in CLM4. Calibrating model parameters using the deterministic least-square fitting method provides little improvement in simulating heat flux and runoff, but using the calibrated values as initial guesses in the MCMC-Bayesian calibration reduces the discrepancies between simulated and observed responses, however the convergence rate is unaffected by the choice of initial guesses.

Focusing on the MCMC-Bayesian inversion method, we conducted inverse modeling at two flux tower sites and one MOPEX basin. We also discussed the impacts of relaxing the rejection standard, data temporal resolution, data types, and soil and vegetation on parameter inference. Informed by our previous sensitivity analysis, we also performed inversion with reduced parameter dimensionality. Moreover, Bayesian model averaging is adopted to integrate the posterior estimates with different reference acceptance probabilities. The major conclusions are as follows.

1. Inversion results at the flux tower and MOPEX sites using monthly and daily surface flux and runoff observations show that the MCMC-Bayesian inversion approach in general can improve the simulation of CLM under different climates and environmental conditions.
2. Temporal resolution of observations has clear impacts on the results of inverse modeling using heat flux data, but the impacts are smaller using runoff data. Due to data redundancy and quality, finer temporal resolution of observations may yield biased estimates and multimodal posterior distributions.

3. Significant improvements can be achieved to better match the observed heat flux and runoff by using the estimated parameters compared to using the default parameter values. The improvement is more obvious for runoff because the calibrated parameters are directly related to runoff processes, and also because default parameter values yield runoff simulations that deviate far from observations. However, improvements in heat flux can also be quite significant especially in areas (e.g., forest) with strong energy and water constraints. Soil and vegetation cover have important impacts on parameter sensitivities, leading to different posterior distributions of parameters at different sites.
4. Reducing the parameter set can make the inverse problem less ill-posed. Numerically, it also speeds up the convergence. In this study, inverse modeling with the reduced parameter set favors parameter estimates closer to the lower bounds than using the full set of parameters.
5. Bayesian model averaging that integrates the posterior estimates with different reference acceptance probabilities can smooth the posterior distribution and provide more reliable parameter estimates, but at the expense of wider uncertainty bounds.

Overall, the MCMC-Bayesian inversion approach is found to provide effective and reliable estimates of model parameters at the site and watershed level to improve CLM simulations of surface flux and runoff. To apply the method for inversion over a region or globally, there are a number of challenges, including computational requirements and availability and quality of observation data. The analyses presented in this study should be extended to a larger number of sites with a wider range of climate, hydrologic, and vegetation/soil conditions to determine if and how model parameters may be transferrable based on site conditions to larger areas or river basins. Exploring model inversion at the river basin level rather than site level using combinations of local flux measurements, area averaged flux data (e.g., derived from satellite), and basin total runoff, each with their own uncertainty estimates, may provide an alternative strategy for calibrating model parameter values for each river basin. To reduce the computational demand, we will also test the performance of the MCMC-Bayesian inversion approach using surrogates (i.e., approximated relationships between inputs and output responses) as alternatives to the CLM4 numerical simulator.

*Acknowledgements.* This work is supported by the project Climate Science for a Sustainable Energy Future, funded by the Office of Science of the US DOE through the Earth System Modeling Program. The Pacific Northwest National Laboratory (PNNL) Platform for Regional Integrated Modeling and Analysis (PRIMA) Initiative provided support for the model configuration and data sets used in the numerical experiments. PNNL is operated for the US DOE by Battelle Memorial Institute under Contract DE-AC05-76RLO1830. Yu Sun and Fuqiang Tian would like to acknowledge the support from the National Science Foundation of China (NSFC51190092, 51222901) and the foundation of the State Key Laboratory of Hydrosience and Engineering of Tsinghua University (2012-KY-03).

Edited by: H.-J. Hendricks Franssen

## References

- Allison, V. J., Miller, R. M., Jastrow, J. D., Matamala, R., and Zak, D. R.: Changes in Soil Microbial Community Structure in a Tallgrass Prairie Chronosequence, *Soil Sci. Soc. Am. J.*, 69, 1412–1421, doi:10.2136/sssaj2004.0252, 2005.
- Baer, S. G., Kitchen, D. J., Blair, J. M., and Rice, C. W.: CHANGES IN ECOSYSTEM STRUCTURE AND FUNCTION ALONG A CHRONOSEQUENCE OF RESTORED GRASSLANDS, *Ecol. Appl.*, 12, 1688–1701, doi:10.1890/1051-0761(2002)012[1688:ciesaf]2.0.co;2, 2002.
- Beven, K. and Binley, A.: The future of distributed models: Model calibration and uncertainty prediction, *Hydrol. Process.*, 6, 279–298, doi:10.1002/hyp.3360060305, 1992.
- Boyle, D. P., Gupta, H. V., and Sorooshian, S.: Toward improved calibration of hydrologic models: Combining the strengths of manual and automatic methods, *Water Resour. Res.*, 36, 3663–3674, doi:10.1029/2000wr900207, 2000.
- Chen, J., Hoversten, G. M., Vasco, D., Rubin, Y., and Hou, Z.: Joint inversion of seismic AVO and EM data for gas saturation estimation using a sampling-based stochastic model, 2004 SEG Annual Meeting, 2004,
- Cosby, B. J., Hornberger, G. M., Clapp, R. B., and Ginn, T. R.: A Statistical Exploration of the Relationships of Soil Moisture Characteristics to the Physical Properties of Soils, *Water Resour. Res.*, 20, 682–690, doi:10.1029/WR020i006p00682, 1984.
- Cosgrove, B. A., Lohmann, D., Mitchell, K. E., Houser, P. R., Wood, E. F., Schaake, J. C., Robock, A., Marshall, C., Sheffield, J., Duan, Q., Luo, L., Higgins, R. W., Pinker, R. T., Tarpley, J. D., and Meng, J.: Real-time and retrospective forcing in the North American Land Data Assimilation System (NLDAS) project, *J. Geophys. Res.*, 108, 8842, doi:10.1029/2002jd003118, 2003.
- Doherty, J.: PEST, Model Independent Parameter Estimation, 2008.
- Duan, Q., Sorooshian, S., and Gupta, V.: Effective and efficient global optimization for conceptual rainfall-runoff models, *Water Resour. Res.*, 28, 1015–1031, doi:10.1029/91wr02985, 1992.
- Duan, Q. Y., Gupta, V. K., and Sorooshian, S.: Shuffled Complex Evolution Approach for Effective and Efficient Global Minimization, *J. Optimiz. Theory App.*, 76, 501–521, 1993.
- Feyen, L., Vrugt, J. A., Nualláin, B. Ó., van der Knijff, J., and De Roo, A.: Parameter optimisation and uncertainty assessment for large-scale streamflow simulation with the LISFLOOD model,

- J. Hydrol., 332, 276–289, doi:10.1016/j.jhydrol.2006.07.004, 2007.
- Fischer, M. L., Billesbach, D. P., Berry, J. A., Riley, W. J., and Torn, M. S.: Spatiotemporal variations in growing season exchanges of CO<sub>2</sub>, H<sub>2</sub>O, and sensible heat in agricultural fields of the Southern Great Plains, *Earth Interactions*, 11, 1–21, 2007.
- Freer, J., Beven, K., and Ambrose, B.: Bayesian Estimation of Uncertainty in Runoff Prediction and the Value of Data: An Application of the GLUE Approach, *Water Resour. Res.*, 32, 2161–2173, doi:10.1029/95wr03723, 1996.
- Gu, L., Meyers, T., Pallardy, S. G., Hanson, P. J., Yang, B., Heuer, M., Hosman, K. P., Riggs, J. S., Sluss, D., and Wullschlegel, S. D.: Direct and indirect effects of atmospheric conditions and soil moisture on surface energy partitioning revealed by a prolonged drought at a temperate forest site, *J. Geophys. Res.*, 111, D16102, doi:10.1029/2006jd007161, 2006.
- Gu, L., Massman, W. J., Leuning, R., Pallardy, S. G., Meyers, T., Hanson, P. J., Riggs, J. S., Hosman, K. P., and Yang, B.: The fundamental equation of eddy covariance and its application in flux measurements, *Agr. Forest Meteorol.*, 152, 135–148, 2012.
- Gupta, H. V., Sorooshian, S., and Yapo, P. O.: Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information, *Water Resour. Res.*, 34, 751–763, doi:10.1029/97wr03495, 1998.
- Hadamard, J.: Sur les problèmes aux dérivées partielles et leur signification physique, *Princeton University Bulletin*, 49–52, 1902.
- Hastings, W. K.: Monte Carlo sampling methods using Markov chains and their applications, *Biometrika*, 57, 97–109, doi:10.1093/biomet/57.1.97, 1970.
- Hou, Z.: Hydraulic Conductivity Estimation Using Tomographic Ground Penetrating Radar Data within a Sampling-Based Bayesian Inversion Approach, *SAGEEP*, 21, 477–489, doi:10.4133/1.2963288, 2008.
- Hou, Z. and Rubin, Y.: On minimum relative entropy concepts and prior compatibility issues in vadose zone inverse and forward modeling, *Water Resour. Res.*, 41, W12425, doi:10.1029/2005wr004082, 2005.
- Hou, Z., Rubin, Y., Hoversten, G. M., Vasco, D., and Chen, J. S.: Reservoir-parameter identification using minimum relative entropy-based Bayesian inversion of seismic AVA and marine CSEM data, *Geophysics*, 71, O77–O88, doi:10.1190/1.2348770, 2006.
- Hou, Z., Huang, M., Leung, L. R., Lin, G., and Ricciuto, D. M.: Sensitivity of surface flux simulations to hydrologic parameters based on an uncertainty quantification framework applied to the Community Land Model, *J. Geophys. Res.*, 117, D15108, doi:10.1029/2012jd017521, 2012.
- Hoversten, G. M., Cassassuce, F., Gasperikova, E., Newman, G. A., Chen, J., Rubin, Y., Hou, Z., and Vasco, D.: Direct reservoir parameter estimation using joint inversion of marine seismic AVA and CSEM data, *Geophysics*, 71, C1–C13, 2006.
- Huang, M., Hou, Z., Leung, L. R., Ke, Y., Liu, Y., Fang, Z., and Sun, Y.: Uncertainty Analysis of Runoff Simulations and Parameter Detectability in the Community Land Model – Evidence from MOPEX Basins and Flux Tower Sites, *J. Hydrometeorol.*, 14, 1754–1772, doi:10.1175/JHM-D-12-0138.1, 2013.
- Ke, Y., Leung, R. L., Huang, M., Coleman, A. M., Li, H., and Wigmosta, M. S.: Developing high resolution land surface parameters for Community Land Model, *Geosci. Model Dev.*, 5, 1341–1362, doi:10.5194/gmd-5-1341-2012, 2012.
- Klepper, O., Scholten, H., and Van Kamer, J. P. G. D.: Prediction uncertainty in an ecological model of the oosterschelde estuary, *J. Forecast.*, 10, 191–209, doi:10.1002/for.3980100111, 1991.
- Kollat, J. B., Reed, P. M., and Wagener, T.: When are multiobjective calibration trade-offs in hydrologic models meaningful? *Water Resour. Res.*, 48, W03520, doi:10.1029/2011wr011534, 2012.
- Kuczera, G. and Parent, E.: Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the Metropolis algorithm, *J. Hydrol.*, 211, 69–85, 1998.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E.: Equation of State Calculations by Fast Computing Machines, *The J. Chem. Phys.*, 21, 1087–1092, 1953.
- Riley, W. J., Biraud, S. C., Torn, M. S., Fischer, M. L., Billesbach, D. P., and Berry, J. A.: Regional CO<sub>2</sub> and latent heat surface fluxes in the Southern Great Plains: Measurements, modeling, and scaling, *J. Geophys. Res.-Biogeo.*, 114, G04009, doi:10.1029/2009jg001003, 2009.
- Sorooshian, S.: Parameter-Estimation of Rainfall – Runoff Models with Heteroscedastic Streamflow Errors – the Non-Informative Data Case, *J. Hydrol.*, 52, 127–138, 1981.
- Sorooshian, S. and Dracup, J. A.: Stochastic Parameter-Estimation Procedures for Hydrologic Rainfall-Runoff Models – Correlated and Heteroscedastic Error Cases, *Water Resour. Res.*, 16, 430–442, 1980.
- Sorooshian, S., Duan, Q. Y., and Gupta, V. K.: Calibration of Rainfall-Runoff Models - Application of Global Optimization to the Sacramento Soil-Moisture Accounting Model, *Water Resour. Res.*, 29, 1185–1194, 1993.
- Suyker, A. E. and Verma, S. B.: Evapotranspiration of irrigated and rainfed maize-soybean cropping systems, *Agr. Forest Meteorol.*, 149, 443–452, 2009.
- Tarantola, A.: Inverse Problem Theory and Methods for Model Parameter Estimation, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2005.
- Van Straten, G. T. and Keesman, K. J.: Uncertainty propagation and speculation in projective forecasts of environmental change: A lake-eutrophication example, *J. Forecast.*, 10, 163–190, doi:10.1002/for.3980100110, 1991.
- Vrugt, J. A. and Robinson, B. A.: Improved evolutionary optimization from genetically adaptive multimethod search, *Proc. Natl. Acad. Sci. USA*, 104, 708–711, doi:10.1073/pnas.0610471104, 2007.
- Vrugt, J. A., Gupta, H. V., Bastidas, L. A., Bouten, W., and Sorooshian, S.: Effective and efficient algorithm for multiobjective optimization of hydrologic models, *Water Resour. Res.*, 39, 1214, doi:10.1029/2002wr001746, 2003a.
- Vrugt, J. A., Gupta, H. V., Bouten, W., and Sorooshian, S.: A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters, *Water Resour. Res.*, 39, 1201, doi:10.1029/2002wr001642, 2003b.
- Vrugt, J. A., Diks, C. G. H., Gupta, H. V., Bouten, W., and Verstraten, J. M.: Improved treatment of uncertainty in hydrologic modeling: Combining the strengths of global optimization and data assimilation, *Water Resour. Res.*, 41, W01017, doi:10.1029/2004wr003059, 2005.

Xia, Y., Mitchell, K., Ek, M., Cosgrove, B., Sheffield, J., Luo, L., Alonge, C., Wei, H., Meng, J., Livneh, B., Duan, Q., and Lohmann, D.: Continental-scale water and energy flux analysis and validation for North American Land Data Assimilation System project phase 2 (NLDAS-2): 2. Validation of model-simulated streamflow, *J. Geophys. Res.*, 117, D03110, doi:10.1029/2011jd016051, 2012.

Yapo, P. O., Gupta, H. V., and Sorooshian, S.: Automatic calibration of conceptual rainfall-runoff models: Sensitivity to calibration data, *J. Hydrol.*, 181, 23–48, 1996.