Uncertainty Analysis of Runoff Simulations and Parameter Identifiability in the Community Land Model: Evidence from MOPEX Basins

MAOYI HUANG, ZHANGSHUAN HOU, AND L. RUBY LEUNG
Pacific Northwest National Laboratory, Richland, Washington

YINGHAI KE
Pacific Northwest National Laboratory, Richland, Washington, and Base of the State Key Laboratory of Urban Environment Process and Digital Modelling, Department of Resource Environment and Tourism, Capital Normal University, Beijing, China

YING LIU AND ZHUFENG FANG
Pacific Northwest National Laboratory, Richland, Washington

YU SUN
Pacific Northwest National Laboratory, Richland, Washington, and Department of Hydraulic Engineering, State Key Laboratory of Hydrosience and Engineering, Tsinghua University, Beijing, China

(Manuscript received 20 September 2012, in final form 14 April 2013)

ABSTRACT

In this study, the authors applied version 4 of the Community Land Model (CLM4) integrated with an uncertainty quantification (UQ) framework to 20 selected watersheds from the Model Parameter Estimation Experiment (MOPEX) spanning a wide range of climate and site conditions to investigate the sensitivity of runoff simulations to major hydrologic parameters and to assess the fidelity of CLM4, as the land component of the Community Earth System Model (CESM), in capturing realistic hydrological responses. They found that for runoff simulations, the most significant parameters are those related to the subsurface runoff parameterizations. Soil texture–related parameters and surface runoff parameters are of secondary significance. Moreover, climate and soil conditions play important roles in the parameter sensitivity. In general, water-limited hydrologic regime and finer soil texture result in stronger sensitivity of output variables, such as runoff and its surface and subsurface components, to the input parameters in CLM4. This study evaluated the parameter identifiability of hydrological parameters from streamflow observations at selected MOPEX basins and demonstrated the feasibility of parameter inversion/calibration for CLM4 to improve runoff simulations. The results suggest that in order to calibrate CLM4 hydrologic parameters, model reduction is needed to include only the identifiable parameters in the unknowns. With the reduced parameter set dimensionality, the inverse problem is less ill posed.

1. Introduction

Observations and climate model predictions suggest that freshwater resources are vulnerable to climate and land use change, with wide-ranging consequences for the human societies and ecosystems. Climate models robustly project that annual average river runoff and water availability will increase as a result of climate change at high latitudes and in some wet tropical areas and will decrease over some dry regions at midlatitudes and in the dry tropics (Bates et al. 2008). Moreover, precipitation intensity and variability will increase and result in a higher risk of flooding and droughts in many regions. Global and regional earth system models are important tools for improving understanding of changes in the hydrological cycle under different climate change scenarios and the corresponding mitigation and adaptation strategies across multiple scales (Bates et al. 2008;
Janetos et al. 2009; Milly et al. 2008; Moss et al. 2010). In addition to effectively and efficiently representing interactions between the component models, earth system models should also be able to simulate the stochastic responses of the hydrologic variables to describe the temporal evolution of their probability density functions (pdfs) with estimates of uncertainty, which are crucial to guide water management and infrastructure design in a changing world (Milly et al. 2008). The latter relies on the ability of land surface models within earth system models to capture realistic hydrological processes and their response to the changing climate.

Compared to previous Community Land Model (CLM) versions, version 4 (CLM4) represents extensive modifications in its model structure and parameterizations over previous CLM versions, including enhancements in the representations of hydrological processes such as runoff generation, groundwater dynamics, soil hydrology, snow module, and surface albedo (Lawrence et al. 2011). Rooted from the climate modeling community, CLM4 has been designed and used for studies of interannual and interdecadal variability, paleoclimate regimes, and projections of future changes of the global climate system (Gent et al. 2010; Lawrence et al. 2012), but its capability for hydrologic simulations at watershed scales has not been adequately investigated (Li et al. 2011).

Uncertainty in hydrologic simulations could be attributed to different sources. One important source of uncertainty is associated with data availability and quality (e.g., forcing data, prescribed land surface properties, and initial and boundary conditions). Another source of uncertainty arises from model structure and parameterization based on our understanding of hydrological processes and how they should be parameterized. Reductions of such uncertainty rely on improved understanding of the physics and its effective representation in models. Such uncertainties can be reduced by optimizing model parameter sets using model calibration techniques with available historical data. Previous efforts include developments of goodness-of-fit measures for comparing model simulations and observations (e.g., Sorooshian 1981; Sorooshian and Dracup 1980) and global optimization schemes, such as the University of Arizona shuffled complex evolution (SCE-UA) method (e.g., Duan et al. 1992, 1993; Sorooshian et al. 1993), used to obtain “optimal” parameter sets. However, it has been acknowledged that such schemes suffer from limitations by assuming that a single optimal parameter set exists without accounting for uncertainties (e.g., Beven and Binley 1992; Gupta et al. 1998; Klepper et al. 1991; Van Straten and Keesman 1991; Yapo et al. 1996). One approach to address these limitations is to provide predictions over a range of parameter sets to represent possible uncertainty in model predictions. Thus, the parameter space needs to be sampled to generate realizations of the model simulations to provide uncertainty estimates (e.g., Beven and Binley 1992; Freer et al. 1996; Kuczera and Parent 1998; Vrugt et al. 2003).

The complex issues encountered in inversion efforts suggest the need for sensitivity analyses to identify a subset of parameters that could be optimized to make the inversion problems less ill posed (Hou et al. 2012; Rosero et al. 2010; van Werkhoven et al. 2009). A reliable sensitivity analysis framework can help increase understanding or quantification of the system behavior (e.g., understanding the relationships between input and output variables). The framework should be able to identify factors that contribute to the output variability, explore interactions between factors, quantify input uncertainty, investigate the uncertainty propagating through the system from inputs to outputs, and evaluate the robustness of the model predictions. Such a sensitivity analysis framework usually requires an efficient sampling technique to explore the multidimensional parameter space, particularly when the numerical simulations are computationally demanding and the parameter dimensionality is high. The success of the sensitivity analyses also relies on proper handling of prior information. Inappropriate assignment or handling of prior information to the input parameters may lead to dramatically different conclusions regarding parameter significances and input–output relationships when response surfaces are nonlinear and/or multimodal.

The concept of entropy (e.g., maximum entropy and minimum relative entropy) is often used to obtain prior probability distributions given by a set of conserved quantities (i.e., quantities that are constant with time, location, or any trajectory of a system), such as the average values of some moment functions and bounds (Jaynes 1968). The derived pdf is unique and is the least informative. A large amount of literature has been dedicated to the elicitation of entropy priors and links to the inversion problems less ill posed (Hou et al. 2012; Rosero et al. 2010; van Werkhoven et al. 2009). A reliable sensitivity analysis framework can help increase understanding or quantification of the system behavior (e.g., understanding the relationships between input and output variables). The framework should be able to identify factors that contribute to the output variability, explore interactions between factors, quantify input uncertainty, investigate the uncertainty propagating through the system from inputs to outputs, and evaluate the robustness of the model predictions. Such a sensitivity analysis framework usually requires an efficient sampling technique to explore the multidimensional parameter space, particularly when the numerical simulations are computationally demanding and the parameter dimensionality is high. The success of the sensitivity analyses also relies on proper handling of prior information. Inappropriate assignment or handling of prior information to the input parameters may lead to dramatically different conclusions regarding parameter significances and input–output relationships when response surfaces are nonlinear and/or multimodal.

In this paper, we aim to evaluate the capability of CLM4 for hydrologic simulations by applying an uncertainty quantification framework designed for CLM4 to 20 selected watersheds from the Model Parameter Estimation Experiment (MOPEX). In Hou et al. (2012), uncertainty in surface flux simulations from CLM4 was explored through a framework that integrates an entropy-based approach (minimum relative entropy), an exploratory sampling approach (quasi Monte Carlo), and generalized linear and additive model analysis for uncertainty quantification of land surface models. By
applying the framework with a focus on surface energy fluxes to flux tower sites, Hou et al. (2012) illustrated that uncertainty in input parameters related to hydrologic processes can affect how surface energy is partitioned between sensible and latent heat fluxes through changes in soil moisture, which has important implications to land–atmosphere interactions for climate and earth system models.

However, Hou et al. (2012) did not evaluate the hydrologic responses because of the lack of runoff observations at the flux tower sites and the scale mismatch between the hydrologic responses and the point measurements of the flux towers. That is, runoff generation parameterizations and hydrologic responses should be evaluated at a watershed scale, but the footprint of a typical flux tower only spans ~1 km. Therefore, we extended the analyses in Hou et al. (2012) in this study to 20 selected MOPEX basins by focusing on the total runoff and its surface and subsurface components, as well as major characteristics of runoff such as time of concentration (CT) and peak flow rate.

By identifying parameters that are more directly associated with the runoff generation parameterizations in CLM4, this study complements Hou et al. (2012) and provides direct guidance on model reduction and parameter calibration against runoff observations for future calibration efforts. More specifically, this study aims to understand the impacts of major hydrologic parameters and how the input uncertainty propagates through the forward CLM4 model to runoff and surface energy fluxes simulations. With these analyses, we can then look at the necessity and the potential of inverting the hydrologic parameters using streamflow records and how the inversion could be done effectively.

2. Site information

MOPEX is an international project aimed at developing enhanced techniques for a priori estimation of parameters in hydrologic models and in land surface parameterization schemes of atmospheric models (Duan et al. 2006). A high priority of MOPEX is to assemble historical hydrometeorological data and river basin characteristics for intermediate scale river basins (500–10000 km²) from a wide range of climate, soil, and vegetation characteristics throughout the world. The MOPEX basins are unregulated, so the dataset is appropriate for developing parameter estimation schemes for hydrologic models and land surface parameterization schemes that ignore the effects of water management.

A total of 431 MOPEX basins over the United States were selected by the Hydrologic Laboratory at the National Weather Service (NWS), with data freely available by the year 2004 (http://www.nws.noaa.gov/oh/mopex/mo_datasets.htm). Although meteorological forcing, soil, and vegetation characteristics are provided by the NWS for all the basins, we decided to update these datasets for consistency with recent developments in the community as well as data requirements for CLM4.

1) The meteorological forcing in this study was extracted from phase two of the North America Land Data Assimilation System (NLDAS-2) 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 1/8° resolution derived from the 32-km resolution, 3-hourly North American Regional Reanalysis (NARR) following the algorithms detailed in Cosgrove et al. (2003). The precipitation fields in NLDAS-2 were produced by combining observations from field stations, level 4 precipitation retrievals from NEXRAD system over the county, and satellites, and they are well suited for hydrologic studies.

2) Soil, vegetation, and land cover characteristics of the basins were derived from the 0.05° input dataset to CLM4 developed by Ke et al. (2012) by overlaying the watershed boundaries of the basins provided by the NWS with the data layers. Specifically, soil information was extracted from the State Soil Geographic (STATSGO) dataset, and the dominant soil type within each basin was used to derive the soil hydraulic parameters in Table A1. Fractional land and vegetation cover were calculated for each basin and assigned to CLM4 input data format for numerical experiments. The detailed site characteristics are listed in Table 1.

3) Default values of the selected hydrologic parameters (i.e., \( f_{\text{max}} \) and \( C_v \); see the appendix for details) were obtained by processing the 90-m resolution digital elevation models (DEMs) from Hydrological Data and Maps Based on Shuttle Elevation Derivatives at Multiple Scales (HydroSHEDS; Lehner et al. 2008). Although DEMs at resolutions of 30 m or higher exist over the basins, the goal of our study is to develop an uncertainty quantification (UQ) framework suitable for CLM4 applications worldwide. Therefore, a global database (i.e., HydroSHEDS), rather than high-resolution DEMs from the U.S. Geological Survey (USGS) database, were used in this study.

4) The 1-km-resolution evapotranspiration (ET) during the period of 2000–10 from Mu et al. (2011) based on the Moderate Resolution Imaging Spectroradiometer (MODIS) is aggregated for each selected basin for evaluating CLM4-simulated latent heat fluxes.

5) Daily streamflow measurements for each basin up to 1999 were included as part of the MOPEX database.
<table>
<thead>
<tr>
<th>Site ID</th>
<th>USGS ID</th>
<th>Site location</th>
<th>Longitude (°E)</th>
<th>Latitude (°N)</th>
<th>Soil texture</th>
<th>$F_{\text{max}}$</th>
<th>$C_s$</th>
<th>$R$</th>
<th>$B$</th>
<th>$H$</th>
<th>Area (km²)</th>
<th>Dominant PFTs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>01127000</td>
<td>Quinebaug River, Jewett City, CT</td>
<td>-71.0914</td>
<td>43.9908</td>
<td>Sandy loam</td>
<td>0.42</td>
<td>0.85</td>
<td>0.47</td>
<td>0.53</td>
<td>0.56</td>
<td>1.60</td>
<td>1847</td>
</tr>
<tr>
<td>2</td>
<td>01606500</td>
<td>South Branch Potomac River near Petersburg, WV</td>
<td>-79.1761</td>
<td>38.9911</td>
<td>Loam</td>
<td>0.40</td>
<td>0.79</td>
<td>0.50</td>
<td>0.40</td>
<td>0.68</td>
<td>1.74</td>
<td>1663</td>
</tr>
<tr>
<td>3</td>
<td>02102000</td>
<td>Deep River at Moncure, NC</td>
<td>-79.1161</td>
<td>35.6272</td>
<td>Silty loam</td>
<td>0.40</td>
<td>0.85</td>
<td>0.37</td>
<td>0.34</td>
<td>0.86</td>
<td>1.24</td>
<td>3714</td>
</tr>
<tr>
<td>4</td>
<td>02236000</td>
<td>St. Johns River near Deland, FL</td>
<td>-81.3828</td>
<td>29.0081</td>
<td>Sand</td>
<td>0.42</td>
<td>0.88</td>
<td>0.30</td>
<td>0.50</td>
<td>0.75</td>
<td>1.03</td>
<td>7951</td>
</tr>
<tr>
<td>5</td>
<td>02387500</td>
<td>Oostanaula River at Resaca, GA</td>
<td>-84.9414</td>
<td>34.5783</td>
<td>Silty loam</td>
<td>0.44</td>
<td>0.78</td>
<td>0.45</td>
<td>0.55</td>
<td>0.72</td>
<td>1.53</td>
<td>4149</td>
</tr>
<tr>
<td>6</td>
<td>03167000</td>
<td>Reed Creek at Grahams Forge, VA</td>
<td>-80.8869</td>
<td>36.9394</td>
<td>Silty loam</td>
<td>0.41</td>
<td>0.78</td>
<td>0.44</td>
<td>0.51</td>
<td>0.73</td>
<td>1.47</td>
<td>640</td>
</tr>
<tr>
<td>7</td>
<td>03364000</td>
<td>East Fork White River at Columbus, IN</td>
<td>-85.9256</td>
<td>39.2000</td>
<td>Silty loam</td>
<td>0.43</td>
<td>0.76</td>
<td>0.36</td>
<td>0.34</td>
<td>0.71</td>
<td>1.21</td>
<td>4421</td>
</tr>
<tr>
<td>8</td>
<td>04079000</td>
<td>Wolf River at New London, WI</td>
<td>-88.7403</td>
<td>44.3922</td>
<td>Sandy loam</td>
<td>0.43</td>
<td>0.84</td>
<td>0.31</td>
<td>0.66</td>
<td>0.72</td>
<td>1.05</td>
<td>5853</td>
</tr>
<tr>
<td>9</td>
<td>05525200</td>
<td>Fox River at Dayton, IL</td>
<td>-88.7906</td>
<td>41.3867</td>
<td>Silty loam</td>
<td>0.43</td>
<td>0.80</td>
<td>0.28</td>
<td>0.43</td>
<td>0.76</td>
<td>0.97</td>
<td>6843</td>
</tr>
<tr>
<td>10</td>
<td>05592500</td>
<td>Kaskaskia River at Vandalia, IL</td>
<td>-89.0899</td>
<td>38.9597</td>
<td>Silty loam</td>
<td>0.42</td>
<td>0.70</td>
<td>0.29</td>
<td>0.25</td>
<td>0.80</td>
<td>0.99</td>
<td>5025</td>
</tr>
<tr>
<td>11</td>
<td>06811500</td>
<td>Little Nemaha River at Auburn, NE</td>
<td>-95.8128</td>
<td>40.3925</td>
<td>Clay loam</td>
<td>0.38</td>
<td>0.76</td>
<td>0.18</td>
<td>0.29</td>
<td>0.93</td>
<td>0.69</td>
<td>2564</td>
</tr>
<tr>
<td>12</td>
<td>07147800</td>
<td>Walnut River at Winfield, KS</td>
<td>-96.9944</td>
<td>37.2242</td>
<td>Silty clay loam</td>
<td>0.41</td>
<td>0.82</td>
<td>0.14</td>
<td>0.24</td>
<td>0.86</td>
<td>0.60</td>
<td>4869</td>
</tr>
<tr>
<td>13</td>
<td>07172000</td>
<td>Caney River near Elgin, KS</td>
<td>-96.3150</td>
<td>37.0306</td>
<td>Silty clay loam</td>
<td>0.41</td>
<td>0.79</td>
<td>0.17</td>
<td>0.16</td>
<td>0.86</td>
<td>0.67</td>
<td>1153</td>
</tr>
<tr>
<td>14</td>
<td>08095000</td>
<td>North Bosque River near Clifton, TX</td>
<td>-97.5678</td>
<td>31.7858</td>
<td>Sandy loam</td>
<td>0.40</td>
<td>0.78</td>
<td>0.10</td>
<td>0.19</td>
<td>0.95</td>
<td>0.49</td>
<td>2507</td>
</tr>
<tr>
<td>15</td>
<td>09251000</td>
<td>Yampa River near Maybell, CO</td>
<td>-108.0292</td>
<td>40.5028</td>
<td>Sandy loam</td>
<td>0.41</td>
<td>0.75</td>
<td>0.11</td>
<td>0.60</td>
<td>0.84</td>
<td>0.51</td>
<td>832</td>
</tr>
<tr>
<td>16</td>
<td>09497500</td>
<td>Salt River near Chrysolite, AZ</td>
<td>-110.4992</td>
<td>33.3791</td>
<td>Clay</td>
<td>0.42</td>
<td>0.75</td>
<td>0.08</td>
<td>0.52</td>
<td>0.92</td>
<td>0.43</td>
<td>7379</td>
</tr>
<tr>
<td>17</td>
<td>11025500</td>
<td>Santa Ysabel Creek near Ramona, CA</td>
<td>-116.8653</td>
<td>33.1069</td>
<td>Sandy loam</td>
<td>0.40</td>
<td>0.78</td>
<td>0.04</td>
<td>0.32</td>
<td>0.97</td>
<td>0.28</td>
<td>290</td>
</tr>
<tr>
<td>18</td>
<td>11427000</td>
<td>American River at North Fork Dam, CA</td>
<td>-121.0228</td>
<td>38.9361</td>
<td>Loam</td>
<td>0.43</td>
<td>0.87</td>
<td>0.24</td>
<td>0.57</td>
<td>0.54</td>
<td>0.86</td>
<td>886</td>
</tr>
<tr>
<td>19</td>
<td>12098500</td>
<td>White River near Buckley, WA</td>
<td>-121.9486</td>
<td>47.1514</td>
<td>Loamy sand</td>
<td>0.41</td>
<td>0.80</td>
<td>0.62</td>
<td>0.46</td>
<td>0.34</td>
<td>2.52</td>
<td>1039</td>
</tr>
<tr>
<td>20</td>
<td>14359000</td>
<td>Rogue River at Raygold, OR</td>
<td>-122.9861</td>
<td>42.4375</td>
<td>Loam</td>
<td>0.42</td>
<td>0.68</td>
<td>0.33</td>
<td>0.70</td>
<td>0.58</td>
<td>1.15</td>
<td>5317</td>
</tr>
</tbody>
</table>
These records were expanded by the surface hydrology team at the University of Arizona (M. Durcik 2011, personal communication) by retrieving daily streamflow records up to 2008 from the USGS (available at http://waterdata.usgs.gov/nwis/dv/; Brooks et al. 2011; Voepel et al. 2011).

3. Methodology

a. Hydrologic parameterizations in CLM4

The runoff generation scheme in CLM4 is a simplified TOPMODEL-based representation (Niu et al. 2005, 2007). Both surface and subsurface runoff are essentially parameterized as exponential functions of water table depth. The rate of surface runoff generation is a function of saturated fraction ($f_{sat}$) within a grid cell, the effective rainfall intensity estimated as the sum of throughfall (rainfall and dewfall after canopy interception) and snowmelt, and the soil infiltration capacity controlled by soil properties and soil moisture within the top soil layer. The $f_{sat}$ can be determined as a function of the depth to water table $z$ calculated as a state variable simulated by the model, as well as the maximum possible saturated area fraction $f_{max}$ and a coefficient $C_s$, which are TOPMODEL-based and could be derived from DEMs.

The rate of subsurface runoff generation is an exponential function of the depth to water table multiplied by the maximum subsurface runoff, defined as the subsurface runoff rate when the whole grid cell is saturated. The total soil column is divided into 10 layers, with the thickness of each layer increasing from top to bottom. The total soil depth for hydrologic simulation is assumed to be a uniform constant value of 3.802 m. Beneath the soil column, an aquifer with a 5-m depth was added as the low boundary condition, which could exchange water with the soil column. Specific yield $S_y$ has been assigned to the aquifer as an important parameter that controls the water exchanges. The parameters, their physical meanings, and prior information in terms of mean, range, and standard deviation (if it exists), are discussed in the appendix and Table A1. Details about the hydrologic parameterizations in CLM4 are described in Niu et al. (2005, 2007), Oleson et al. (2008, 2010), and Lawrence et al. (2011).

b. Experimental setup

In this study, each MOPEX basin is treated as an entity by area-averaging the NLDAS-2 meteorological forcing described in section 2, following the standard MOPEX protocol (Duan et al. 2006). The high-resolution CLM4 input dataset from Ke et al. (2012) was aggregated according to the basin boundaries in Fig. 1. For each parameter set at each basin, the model was spun up by cycling the forcing for at least 5 times (i.e., $28 \times 5$ yr) until all the state variables reached an equilibrium. Analyses, as will be detailed in section 3c, were performed on variables over the period of 1979–2007, with initial condition from the model spinup.

c. Uncertainty quantification framework

We applied the UQ framework used in Hou et al. (2012) to analyze the input parametric sensitivity and quantify uncertainty in model simulations from CLM4 at the selected MOPEX sites and flux tower sites. The readers are referred to Hou et al. (2012) for detailed descriptions and discussions of the UQ framework and its application to selected flux tower sites. For completeness, a description of the UQ framework, particularly on the entropy concept and efficient sampling approach, is included in the appendix. Briefly, our UQ framework features an entropy approach to quantify uncertainty in the input parameters. With this approach, the uncertainty associated with the calculated responses is most representative of our knowledge of the system. Our approach incorporates an efficient sampling method.
to explore the parameter space so that the output statistics are exploratory of most if not all the possible realities.

The approach uses multivariate generalized linear model analyses and significance statistical tests to rank the significance of input parameters and develop relationships between inputs and outputs using response surface plots and the finalized linear models. We performed the generalized linear model (GLM) analysis, which can be used to evaluate the statistical significance of the explanatory terms (e.g., linear, quadratic, and interaction) through statistical F tests. For each response variable, we start with a model that has all explanatory terms, and then by evaluating the fitted variance of each term, one parameter/factor/term that is least important is identified. We apply the Akaike information criterion (AIC; Akaike 1974) to reconcile the goodness of fit of a model and the model complexity to decide whether to keep a term in a fitted GLM based on its contribution. Terms that are kept in the GLM have enough significance to compensate for the penalty of increasing model complexity. The most removable terms are removed one by one using a procedure called AIC-based backward removal. Eventually, a model with the minimum AIC is achieved with only the necessary terms. Each time that an input parameter passes the significance test for a certain response variable is counted once. Summing over all the responses (e.g., runoff in different months and at different sites), we summarized the percentage of significance tests passed by each input parameter/factor as the significance score used in our analysis.

d. Derivation of hydrologic indices

In Hou et al. (2012), we focused on analyzing mean monthly averages of latent/sensible heat fluxes and runoff across the simulation periods. In this study, in addition to these outputs, we analyzed the responses of total, surface, and subsurface runoff and derived mean annual monthly peak flow and time of concentration (CT; Maurer et al. 2007; Stewart et al. 2005) over the simulation period from the observed and simulated hydrographs to represent the shapes of the hydrographs and important measures of hydrologic analyses. Specifically, CT for each year is calculated as

$$CT = \sum_{i}(t_i Q_i) / \sum Q_i,$$

where $t_i$ is time in months from the beginning of the water year (i.e., 1 October), and $Q_i$ is the corresponding streamflow for month $i$.

Daily streamflow was partitioned into two components: surface flow ($Q_s$) and base flow ($Q_b$), which represent the response of streamflow to rainfall and discharge from the groundwater storage to the stream, respectively. The one-parameter recursive filter (Lyne and Hollick 1979) was used to estimate the base flow:

$$Q_{b,k} = aQ_{b,k-1} + \frac{1-a}{2}(Q_{b,k} + Q_{b,k-1}),$$

where $a$ is a filter parameter and was set to 0.925 (Arnold and Allen 1999; Eckhardt 2005; Lyne and Hollick 1979). This dataset was then further aggregated to the monthly time step in this study to evaluate simulations of total runoff and the partitioning between surface and subsurface runoff climatologically because, over long time scales, the effect of runoff routing along hydrologic pathways on the partitioning of streamflow components becomes negligible for the small- to intermediate-scale basins. Therefore, the partitioning of streamflow between surface and base flow at a monthly scale would reflect the partitioning of runoff between surface and subsurface runoff from a runoff generation perspective.

Based on the daily values of streamflow ($Q$), surface flow ($Q_s$), base flow ($Q_b$), meteorological data from the MOPEX dataset for 1948–99, and the Oregon State University Parameter-Elevation Regressions on Independent Slopes Model (PRISM) database for 2000–08, a number of hydrologic indices at annual time scales were estimated based on algorithms detailed in Brooks et al. (2011) and Voepel et al. (2011) to facilitate the characterization of hydrologic regimes.

The runoff ratio ($R$) is defined as

$$R = \frac{Q}{P},$$

where $P$ is precipitation, and the base flow ratio ($B$) is defined as

$$B = \frac{Q_b}{Q}.\ (4)$$

The Horton index ($H$) that represents the vaporation-to-wetting ratio (Troch et al. 2009) is defined as

$$H = \frac{V}{W} = \frac{P - Q}{P - Q_s}.$$

where vaporation ($V$) is the portion of precipitation that eventually vaporizes through soil and canopy evaporation and transpiration and is therefore excluded from streamflow and wetting ($W$) is the portion of precipitation retained by the catchment at storm time scales to sustain ET and base flow. At the catchment scale, $V$ can be estimated as $P - Q$, which represents a lower bound of the actual ET, and $W$ can be estimated as $P - Q_s$. Thus, $H$ can be treated as an estimate of the lower bound of vegetation water use during the growing
season. As discussed in Troch et al. (2009), $H \equiv 0$ in energy-limited regions because most precipitation will leave the catchment as streamflow, $H \equiv 1$ in water-limited regions (i.e., arid and semiarid) because total streamflow is mostly contributed by surface flow (i.e., $Q \approx Q_s$), and $H < 1$ in humid regions because total streamflow is greater than surface flow (i.e., $Q > Q_s$).

The humidity index (HU) is defined as

$$HU = \frac{P}{PET},$$

(6)

where the monthly potential evapotranspiration (PET) is calculated using Hamon’s equation (Hamon 1961; Voepel et al. 2011; Wolock and McCabe 1999) as

$$PET = 13.97(d)(D)^{\rho_w},$$

$$\rho_w = \frac{4.95e^{0.062T}}{100},$$

(7)

where $d$ is the number of days in a month, $D$ is the mean monthly hours of daylight in units of 12 h estimated using the algorithm described in Forsythe et al. (1995), $\rho_w$ is the saturated water vapor density, and $T$ is the monthly mean temperature in degrees Celsius estimated as the average of minimum and maximum temperatures (i.e., $T_{min}$ and $T_{max}$). The estimated values of the hydrologic indices are given in Table 1.

4. Results

Figure 2 shows the boxplots of surface heat fluxes and runoff for the ensemble simulations at two sites in the U.S. southern Great Plains. The boxplots summarize the output uncertainty using five numbers: the smallest observation (sample minimum), lower quartile (Q1), median (Q2), upper quartile (Q3), and the largest observation (sample maximum). The vertical length of the closed boxes denotes the interquartile range (Q3–Q1; Upton and Cook 1996). The simulated surface and hydrologic fluxes using the default parameter values (the values originally hard-coded in CLM4) are shown as red circles, while observations are denoted by the green symbols. In general, the default values of the parameters give higher estimates of latent heat flux for both sites. Therefore, the default parameter values obviously yield underestimates of runoff and its components, particularly for the early summer months, while the observed runoff is well bounded by the range of the ensemble predictions. These results indicate that we have assigned reasonable physical bounds for the input parameters as intended by the parameterizations and our sampling approach provided reasonable parameter values in the parameter space so that the observations fall within the output possibilities. They also demonstrate the necessity and possibility of calibrating the input parameters using runoff datasets at the MOPEX basins. We stress that it is crucial, before any...
actual inversion of model parameter, to check the feasibility of calibration in a systematic way to determine whether the actual observations can be captured by the possible output ranges, which depend on the skill of the model physics/parameterizations, and the appropriate assignment of prior bounds to the input parameters.

The ensemble predictions of latent heat fluxes during the summer months span a wide range from 0 to 200 W m\(^{-2}\), but have very narrow ranges in December and January. Similar behaviors are found for runoff and sensible heat fluxes, which reflect the dominant control of radiation on ET during winter and the dominant control of soil moisture on ET during summer. That is, during winter, the potential ET and therefore actual ET is low because of the limited energy inputs (i.e., incoming shortwave and longwave radiation), but during summer when the energy input is sufficient, soil moisture availability, which depends on the hydrologic parameters, becomes more important in determining the partitioning of the incoming energy into latent and sensible heat. These results agree with what we demonstrated for the flux tower sites (Hou et al. 2012), that data collected in the warmer seasons are more informative for parameter optimization; in other words, misfits during such time periods could be improved the most by using observations.

We assigned significance scores to each of the input parameters for all 12 months across the 20 MOPEX sites (Fig. 3). The scores are defined as the percentage of occurrences that an input parameter passes the significance test for all months and all sites. For runoff, the most significant parameter is identified as \(f_{\text{drain}}\), followed by \(Q_{\text{dom}}\) and \(S_y\). It is not surprising that these three subsurface runoff generation parameters are of the first-order significance when we focus on runoff. Soil texture–related parameters and surface runoff parameters are all of secondary significance, although soil texture parameters show slightly higher significance in general.

To look at the effects of those significant parameters on runoff, we show the individual impacts of the parameters using boxplots as shown in Fig. 4. There are several major observations: 1) larger values of \(f_{\text{drain}}\) and \(S_y\) correspond to lower total runoff, but \(Q_{\text{dom}}\) has an opposite influence; 2) the impacts of the parameters decrease from warmer to colder months, as discussed earlier; and 3) the uncertainty ranges of runoff predictions increase with \(Q_{\text{dom}}\), decrease with \(S_y\), and vary uniformly with \(f_{\text{drain}}\). The physical implications of the parameter sensitivity will be discussed in section 5.

For each of the 20 MOPEX sites, we calculated 12 monthly averages of latent heat flux and total runoff throughout the corresponding simulation period. We also calculated annual peak flow rate and CT from 1979 to 2009. Therefore, overall we evaluated 20 \((12 \times 3) + 31 \times 2 = 782\) response variables. For each response variable, we conducted statistical tests and generalized linear model analyses and recorded the number of occurrences of each input parameter passing the significance test. We summarized the results for different sites and different seasons, as shown in Figs. 5–9.

Figure 5 summarizes the significance of input parameters for all 12 months for each of the 20 study sites. Most sites have the same three parameters identified above as the most significant, with several exceptions. For example, sites 4, 14, and 17 have important parameters \(b\) (the Clapp and Hornberger exponent) and air-entry pressure \(\Psi_s\) and sites 2, 3, and 5 have porosity \(\theta_s\) identified to be significant, while sites 1, 6, and 19 are relatively insensitive to the input parameters. As shown in Table 1, sites 4, 14, and 17 are moderately wet to arid basins dominated by shallow-rooted vegetations with loamy to sandy soil texture. Ecosystems under such site conditions are water limited and are sustained by rainfall pulse infiltrated into the root zone. Therefore, soil hydraulic parameters such as \(b\) and \(\Psi_s\) become controlling factors as they moderate infiltration as well as runoff generation through the infiltration excess runoff mechanism. Sites 1, 6, and 19 are wet basins (characterized by high values of the humidity index), so surface fluxes are more energy limited than water limited, and runoff generation is determined by saturation excess runoff, a characteristic of the wet climate condition. Therefore, simulated surface fluxes and runoff at these sites are dominated by ambient climate condition, rather than by site characteristics. Sites 2, 3, and 5 are also wet sites, but with much finer soil textures than the group of 1, 6, and 19. At these finer-soil sites, the output variability is large and most of the input parameters contribute to such variability. Finer soil texture generally results in stronger sensitivity of output variables to the inputs.
To evaluate the behavior of CLM4 outputs in different hydrologic regimes, we plotted in Fig. 6 the overall parameter sensitivity (defined as the frequency or number of occurrences that an input parameter passes the significance test for all months for a particular site, using both monthly latent heat flux and runoff simulations as the response variables) against the various hydrologic indices described in section 3d, given that the sensitivity of latent heat flux and runoff simulations are similar at the selected sites. As can be seen from Fig. 6,
the simulated response variables tend to be sensitive to more and more input parameters when the site condition is more water limited, indicated by higher values of the Horton index, or lower values of the runoff ratio, the base flow ratio, and the humidity index.

Moreover, although a high base flow ratio generally indicates wetter conditions, similar to the runoff ratio and the humidity index, it also corresponds to larger subsurface flow storage or better hydrologic connectivity in the subsurface system and larger contributions of base flow to the total measured streamflow. Therefore, by introducing information on the partitioning of total streamflow into its surface and subsurface components, the base flow ratio $B$ could better explain the sensitivity of simulated variables to input parameters than the runoff ratio $R$ and the humidity index $HU$, with more consistent relationships with the overall sensitivity of response variables to the input parameters. This is supported by both visualization and the larger $R^2$ (the fitted variance of overall sensitivity by the selected index), as shown in Fig. 6b as compared to those in Figs. 6a and 6d.

Similarly, the Horton index also considers information on the partitioning of precipitation into surface and subsurface flow in addition to climate and therefore shows better explanatory capability than the runoff ratio and the humidity index $HU$. To summarize, Fig. 6 suggests that 1) climate condition alone might be a poor indicator of the hydrologic regimes of the selected sites and 2) the Horton index and base flow ratio could be used as matrices for better classifications of selected sites into different hydrologic regimes and guide parameter inversion in our follow-up studies.

Figure 7 shows the seasonal variations in the parameter sensitivity patterns. The same three parameters identified above are significant in all months. During summer, there are slightly more pixels with warmer colors, indicating that heat fluxes are sensitive to more parameters. Such behavior is less obvious for runoff, so the significant patterns are more uniform for runoff.

There are different ways to describe runoff, including hourly time series, monthly average, peak flow rate, and arrival time (i.e., time of concentration). Although we
focused on monthly averages of total runoff in the above analyses, the other characteristics of runoff (e.g., peak flow rate and CT) have practical importance for managing water resources. Therefore, we performed statistical tests and assigned significance scores to each of the input parameters for each year across the 20 MOPEX sites (Fig. 8).

Compared to Fig. 4, we found that the ranks of the top three parameters change as the response variable is changed from monthly averaged runoff to peak flow rate and CT. Specifically, for monthly averages, the top three parameters are $f_{drai}$, $Q_{dm}$, and $S_y$, in that order, but for peak flow rates they are $S_y$, $Q_{dm}$, and $f_{drai}$. For CT, only two parameters can be clearly identified to be significant, $f_{drai}$ and $S_y$, although other parameters such as $Q_{dm}$ have comparable significance, and none of the other parameters are negligible. For peak flow rate, it is not surprising to see the impacts of $Q_{dm}$ (the maximum subrunoff rate) and $S_y$ (drainable porosity) since they can control extreme runoff rates directly, while the parameter $f_{drai}$ plays a role indirectly by affecting the storage. The peak flow rate usually occurs during the warmer season, when the important parameters stand out among all the input parameters. However, CT is controlled by the shape of the hydrograph, which depends on the flow rate throughout the year. As we discussed earlier, the other secondary parameters are also important during the colder and dryer seasons, so we can see that the CT scores for these parameters, especially the surface runoff parameters, are higher.

The significance patterns for peak flow rate and CT from site to site (Fig. 9) are not as systematic as those for monthly averages of runoff because of the mixed parameter effects across the different seasons, particularly for CT because it is influenced by the accumulated flow rate. In contrast, the significance patterns across the years from 1979 to 2009 are very consistent (Fig. 10). Small variations from year to year represent the impact of the overall climate conditions. During wetter years and wetter seasons, the output variability is larger and more sensitive to the variations of input parameters.

In Fig. 11, the variability of response variables (i.e., runoff and heat fluxes) is calculated as the standard deviation of the response variable for each month. It shows
that finer soil texture yields stronger sensitivity of runoff as well as heat fluxes to the input parameters. A field site with coarse soil texture facilitates the distribution and discharge of precipitation and tends to have more uniformly distributed water in the rooting zones, which reduces variability of the runoff responses. This finding agrees with the results from sensitivity analyses at the flux tower sites (Hou et al. 2012).

5. Summary and future work

In this study, we adopt an uncertainty analyses framework to study the sensitivity of various CLM4 simulated fluxes and their derived variables (heat fluxes, runoff, peak flow rate, and time of concentration) to selected hydrological parameters. For total runoff and its subsurface and surface components, the most significant parameter is identified as $f_{\text{drain}}$, followed by $Q_{\text{dm}}$ and $S_y$. Soil texture-related parameters and surface runoff parameters are all of secondary significance, although soil texture parameters show slightly higher significance in general.

Different output variables have different sensitivity patterns. When peak flow rate and CT are used, we found that the significance patterns are not as systematic as those for monthly averages of runoff because of the
mixed parameter effects across the different seasons, particularly for CT, which is influenced by the accumulated flow rate, and the intraseasonal runoff variations are smoothed out through integration over time.

Climate and soil conditions play important roles on the parameter sensitivity. At wetter sites, surface fluxes are rather energy limited than water limited, and runoff generation is determined by saturation excess runoff, which is characteristic of the wet climate condition. Therefore, simulated surface fluxes and runoff at those sites are dominated by ambient climate condition, rather than by site characteristics. In general, the

![Fig. 8. Overall significance of the parameters on peak flow, CT, and coefficients for runoff (COR).](image)

![Fig. 9. Sensitivities of (a) peak flow rate and (b) CT to the parameters at the 20 MOPEX sites.](image)
water-limited hydrologic regime and finer soil texture result in stronger sensitivity of output variables to the input parameters in CLM4. However, it is necessary to extend the analysis to more MOPEX basins to confirm the finding in this study and to provide guidance for parameter inversion/calibration for continental- and global-scale applications of CLM4 for hydrologic predictions.

The model sensitivity to the significant parameters identified in this study can be interpreted from their physical meanings. As discussed in Li et al. (2011), we note that the physical meanings of $f_{\text{drai}}$ in CLM4 are
different from those defined in the original TOPMODEL and implemented by Niu et al. (2005, 2007). Interested readers are referred to Li et al. (2011) for more details. Specifically, $f_{\text{drai}}$ is the reciprocal of the effective subsurface storage capacity scaled by the water table depth for the subsurface runoff generation calculation. In Fig. 4, when $f_{\text{drai}}$ increases, the storage capacity decreases, which means slower depletion of the deep-layer soil moisture and, therefore, a lower flux from the surface layers to recharge the deep layers (Beven 1997; Iorgulescu and Musy 1997; Kirkby 1997; Li et al. 2011; Niu et al. 2005). When $f_{\text{drai}}$ is larger than 2, less water will leave the soil column as subsurface runoff so that $f_{\text{drai}}$ is no longer a limiting factor for soil moisture and runoff generation. Hence, in general, the uncertainty range of runoff predictions is large when the runoff is large (i.e., when $f_{\text{drai}}$ is relatively small).

It is expected that increasing $Q_{\text{dm}}$, the maximum subsurface runoff rate, increases the total runoff generation at a given water table depth with a given $f_{\text{drai}}$ value [i.e., Eq. (A5)]. Such an impact is more obvious during summer with higher precipitation for most basins; for colder and dryer seasons, however, $Q_{\text{dm}}$ is not a limiting factor on runoff generation.

The specific yield $S_y$ would affect the recharge to aquifer and hence the subsurface runoff generation by modifying the interactions between water within the soil column and the groundwater aquifer. Figure 4 suggests that $S_y$ is important except for the winter. A smaller $S_y$ means that the water table will increase more per unit recharge, which could affect runoff generation more significantly and is thus associated with a wider uncertainty bound, while the opposite is true for a large $S_y$.

Accurate simulations of runoff and surface fluxes using CLM4 rely on information about the model parameters, many of which, however, are not measureable or subject to great uncertainty. This necessitates parameter calibration using available data such as streamflow observations. In this study, we demonstrated that parameters play different roles in runoff simulations in CLM4. Parameters with the highest significance ranks are the most identifiable through streamflow inversion. We also demonstrated that in order to better characterize the parameter sensitivity of runoff simulations, as well as latent heat flux simulations, it is helpful to use metrics such as the Horton index and base flow ratio, which consider the partitioning of runoff components, to guide the design of the inversion framework.

Considering the availability, resolution, and redundancy of streamflow data, the inverse problem of calibrating all input parameters is ill posed and has nonunique solutions. It is reasonable to adopt a stochastic inversion framework to address the corresponding nonuniqueness and uncertainty of parameter estimation. No matter whether a deterministic (e.g., least squares minimization) or stochastic (e.g., Bayesian with Markov chain Monte Carlo sampling) inversion approach is used, it is desirable to make the inverse problem less ill posed by reducing the parameter set dimensionality. The significance tests provided a systematic approach to identify insignificant parameters for each response variable (e.g., runoff) that should be excluded in model calibration in follow-up studies. Guided by this study, we will apply the UQ framework with inverse modeling capability to all 431 MOPEX sites over the United States to provide calibrated model parameters for improving CLM4 hydrologic simulations. We will extend our analysis on parameter sensitivity pattern and evaluate the possibility of parameter calibration within each class of watersheds. This way, we will identify which types of watershed favor parameter calibration and which types of watersheds share common parameter significance patterns or even parameter values so that model parameter calibration can be performed to improve the accuracy of regional- and global-scale simulations.

**Acknowledgments.** This work is supported by the project “Climate Science for a Sustainable Energy Future” funded by the DOE 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 datasets used in the numerical experiments. The authors thank Drs. Matej Durcik, Peter Troch, and Murugesu Sivapalan, as well as the National Weather Service for providing datasets for the MOPEX sites, and Drs. D. Lawrence, S. Swenson, and G.-Y. Niu and the three anonymous reviewers for their constructive comments and suggestions on this study. PNNL is operated for the U.S. DOE by Battelle Memorial Institute under Contract DE-AC05-76RL01830.

**APPENDIX**

**An Uncertainty Quantification Framework Designed for CLM**

**a. Prior information and sampling**

Prior information about input parameters can be obtained from literature, databases, or site-specific prior data collection. The information is either hard (direct measurement) or soft (indirect obtained from inference) and in the form of ranges or default values. Strictly derived prior probability density distributions can be obtained using the entropy principle (Kesavan and Kapur 1989;
also superior to the Latin hypercube sampling method terms of accuracy, confidence level, and speed, and it is approach beats regular Monte Carlo approaches in Carlo (QMC), to generate samples from the pdfs. The use a low-discrepancy sampling technique, quasi Monte (2011).

Given the prior pdfs derived from MRE principle, we use a low-discrepancy sampling technique, quasi Monte Carlo (QMC), to generate samples from the pdfs. The approach beats regular Monte Carlo approaches in terms of accuracy, confidence level, and speed, and it is also superior to the Latin hypercube sampling method for high-dimensionality problems.

b. Output responses and selected input hydrologic parameters

For each MRE–QMC sample set of input parameters, we calculate the time series of runoff during the site-specific simulated time period with runoff observations. The simulated runoffs are then postprocessed to obtain monthly means of runoff and its components, time of concentration (arrive time of center mass), and peak flow rates. The first set of output responses are the monthly means of runoff and its components, that is, base flow (subsurface runoff) and overland flow (surface runoff). Analyzing these datasets would tell us the seasonal variability and parameter sensitivity of runoff. The other sets of output responses are peak flow rate of runoff and time of concentration, which would help understand the sensitivity of major features of the runoff processes to the input parameters.

To obtain the components of runoff, we apply base flow separation techniques, which are often used to determine what portion of a streamflow hydrograph occurs from base flow and what portion occurs from overland flow. Common methods include using isotope tracing, among others.

In CLM4, soil water is estimated in a 10-layer equation as

$$\frac{\partial \theta}{\partial t} = -\frac{\partial q}{\partial z} - s,$$

(A1)

where $\theta$ is the volumetric soil moisture content, $z$ is the height above some datum in the soil column, $t$ is time, and $s$ is a soil moisture sink term (i.e., root extraction). The upper boundary condition of this equation is the infiltration flux ($q_{\text{infl}}$) into the top soil layer given by

$$q_{\text{infl}} = q_{\text{liq,grnd}} - q_{\text{over}} - q_{\text{evp}},$$

(A2)

where $q_{\text{evp}}$ is the evaporation of liquid water from the top soil layer and $q_{\text{liq,grnd}}$ is the liquid precipitation reaching the ground plus any meltwater from snow. The surface runoff $q_{\text{over}}$ is parameterized as

$$q_{\text{over}} = f_{\text{sat}} q_{\text{liq,grnd}} + (1 - f_{\text{sat}}) \max[0, (q_{\text{liq,0}} - q_{\text{infl,max}})],$$

(A3)

where $q_{\text{infl,max}}$ is the maximum soil infiltration capacity; $f_{\text{sat}} = f_{\text{max}} \exp(-C_f z_{\text{v}})$ is the saturated fraction of the grid cell; $z_{\text{v}}$ is the water table depth; and $f_{\text{max}}$, $C_f$, and $f_{\text{over}}$ are model parameters.

In Eq. (A1), the lower boundary condition and the sink term for the soil layers from the bottom of the soil column to the water table depth are parameterized through the recharge to the aquifer ($q_{\text{recharge}}$) and the subsurface runoff ($q_{\text{drai}}$):

$$q_{\text{recharge}} = \frac{\Delta \theta_{\text{liq,N}_{\text{levsoi}}} \Delta z_{\text{N}_{\text{levsoi}}}^1}{\Delta t}$$

and

$$q_{\text{drai}} = q_{\text{drai,max}} \exp(-f_{\text{drai}} z_{\text{v}}),$$

(A5)

where $\Delta \theta_{\text{liq,N}_{\text{levsoi}}}^1$ and $\Delta z_{\text{N}_{\text{levsoi}}}^1$ are the change in liquid water content solved numerically based on Eq. (A1) and the thickness of the bottom soil layer, respectively; $q_{\text{drai,max}}$ and $f_{\text{drai}}$ are model parameters; and $z_{\text{v}}$ is the groundwater table depth, which is calculated from the aquifer water storage scaled by the average specific yield $S$, as described in Niu et al. (2007).

Therefore, we select 10 parameters that have direct influences on hydrologic processes, including soil hydrology and runoff generation processes. The selected parameters are $f_{\text{max}}$, $C_f$, $f_{\text{over}}$, $f_{\text{drai}}$, $q_{\text{drai,max}}$ (denoted as $Q_{\text{drai}}$, hereafter), $S$, $b$, $\Psi_s$, $K_s$, and $\theta_s$. Explanations of the 10 parameters and their prior information are shown in Table A1 (i.e., same as Table 2 of Hou et al. 2012). Uncertainty ranges and prior information of the parameters were determined based on literature (Niu et al. 2005; Niu et al. 2007; Oleson et al. 2008, 2010) and discussions with the CLM developers (G. Niu, S. Swenson, and D. Lawrence 2011, personal communication).

c. Uncertainty analysis and statistical tests

Given information about the 10 parameters in Table A1, we derive the closed-form MRE prior pdfs, as shown in Hou et al. (2012, their Fig. 3), which take various forms including exponential, Gaussian, truncated exponential, and truncated Gaussian distribution. We then generate quasi Monte Carlo samples of the input parameters from these MRE prior pdfs. The marginal histograms
of these samples and the scatters are shown in Figs. 4 and 5 in Hou et al. (2012), respectively. The MRE-QMC approach enables us to honor all the information, including mean, variance, and bounds for each parameter.

We analyze the output responses and their dependency on input parameters using various exploratory data analysis tools. We use nonparametric boxplot analyses to look at the output variability with respect to soil types, plant functional types (PFTs), and input model parameters and visualize whether an input parameter (or independent variable) has a nonlinear relationship with the output response variable and how the input uncertainty propagates through CLM4 simulations.

We then use generalized linear model analyses and AIC-based (Akaikes 1974) backward removal approach to identify the significant parameters for each monthly output data for each output variable for each field site. The GLM analyses yield quantitative measures of how these input parameters/factors (including soil texture, plant functional types, and climate conditions) control the parameter sensitivities. We can then develop predictive relationships between input parameters and output responses, as well as reduce the dimensionality of input parameter space in future sensitivity or inversion studies, by removing insignificant parameters.

REFERENCES


