1

2

3 Classification of Hydrological Parameter Sensitivity and

4 Evaluation of Parameter Transferability across 431 US MOPEX

5 Basins

6

7 Huiying Ren¹, Zhangshuan Hou^{1*}, Maoyi Huang¹,

8 Jie Bao¹, Yu Sun², Teklu Tesfa¹, L. Ruby Leung¹

9

10 ¹Pacific Northwest National Laboratory, Richland, WA, USA

11 ²Department of Hydraulic Engineering, Tsinghua University, Beijing, China

12

13

14

15

16

17

18 To be submitted to *Journal of Hydrology*

19

*Corresponding author address: Zhangshuan (Jason) Hou, Pacific Northwest National Laboratory, P.O. Box 999, Richland, WA 99352.
E-mail: zhangshuan.hou@pnnl.gov

Abstract

The Community Land Model (CLM) represents physical, chemical, and biological processes of the terrestrial ecosystems that interact with climate across a range of spatial and temporal scales. As CLM includes numerous sub-models and associated parameters, the high-dimensional parameter space presents a formidable challenge for quantifying uncertainty and improving Earth system predictions needed to assess environmental changes and risks. This study aims to evaluate the potential of transferring hydrologic model parameters in CLM through sensitivity analyses and classification across watersheds from the Model Parameter Estimation Experiment (MOPEX) in the United States. The sensitivity of CLM-simulated water and energy fluxes to hydrological parameters across 431 MOPEX basins are first examined using an efficient stochastic sampling-based sensitivity analysis approach. Linear, interaction, and high-order nonlinear impacts are all identified via statistical tests and stepwise backward removal parameter screening. The basins are then classified according to their parameter sensitivity patterns (internal attributes), as well as their hydrologic indices/attributes (external hydrologic factors) separately, using Principal component analysis (PCA) and expectation-maximization (EM) – based clustering approach. Similarities and differences among the parameter sensitivity-based classification system (S-Class), the hydrologic indices-based classification (H-Class), and the Koppen climate classification systems (K-Class) are discussed. Within each parameter sensitivity-based classification system (S-Class) with similar parameter sensitivity characteristics, similar inversion modeling setups can be used for parameter calibration, and the parameters and their contribution or significance to water and energy cycling may also be more transferrable. This classification study provides guidance on identifiable parameters, and on parameterization and inverse model design for CLM but the methodology is applicable to other models. A set of

experiments of model calibration were conducted to evaluate the transferability of model calibration strategies and parameter values within and between the classes. It was demonstrated that inverting parameters at representative sites belonging to the same class can significantly reduce parameter calibration efforts.

Keywords: Community Land Model, MOPEX, parameter identifiability, parameter transferability, basin classification.

1. Introduction

As land surface models in Earth system models grow in complexity and sophistication, they have become a more powerful tool for understanding biogeophysical and biogeochemical processes and their interactions and dynamic coupling with the climate system. Meanwhile, the potential for improving Earth system predictions still faces great challenges because of the high dimensionality of model parameters. In practice, many parameters in land surface models are expected to vary from site to site and subject to great uncertainties. However, with the high-dimensional parameter space, systematic calibration at numerous field sites faces great challenges because of the computational demand and the ill-posed nature of the inverse problems.

Previously, we adopted an uncertainty quantification (UQ) framework that integrates minimum-relative-entropy concept, efficient sampling, and statistical tests, to investigate, score, and rank the contributions of hydrological parameters in the Community Land Model (Hou et al., 2012; Huang et al., 2013). We evaluated the impacts of soil texture, plant functional types (PFTs), and climate conditions on the sensitivity of simulated water and energy fluxes to hydrological parameters for 13 flux tower sites and 20 Model Parameter Estimation Experiment

(MOPEX) basins. In general, a small subset of parameters were consistently identified that result in larger uncertainty in the heat flux and runoff simulations across the sites (Hou et al., 2012; Huang et al., 2013). However, some other parameters can play comparable roles at individual sites, and the contributions of the significant parameters vary from site to site. To generalize the findings, it is important to expand the parameter sensitivity study of Community Land Model (CLM) to all 431 MOPEX basins with climate and hydrologic conditions that vary across the continental United States. However, this requires tremendous computational resources so to acquire a general understanding of the importance of the hydrological parameters, we need to reduce the complex system to make model calibration work possible. A method that classifies basins based on their parameter sensitivity may offer a way to reduce parameter dimensionality. For example, basins within the same class may share the same design for transferrable parameter calibration and even transferrable parameter values. These could significantly reduce the computational requirement for improving model performance by calibrating fewer parameters and/or calibrating parameters only at selected representative sites within each class.

During the last two decades, considerable efforts have gone into analysis and development of methodological approaches to hydrologic classification that groups hydrologic systems that are similar and limits the variability within classes (Ali et al., 2012; Carrillo et al., 2011; McDonnell and Woods, 2004; Sawicz et al., 2011; Sivakumar, 2008; Sivakumar and Singh, 2012; Wagener et al., 2007). This helps guide modeling and measurement efforts, and provides constraints on predictions in ungauged basins, as well as on estimation of environmental impacts (Hou et al., 2012; Wagener et al., 2007). (Coopersmith et al., 2012) organized a large and diverse population of catchments in the continental United States into homogeneous groups based on climate seasonality and runoff regime. (Sivapalan et al., 2003) found that simple models with a

small number of parameters can adequately describe a majority of catchment responses. A number of studies encouraged the use of catchment classification using a small number of explanatory variables to improve understanding and predictions (Bloschl and Sivapalan, 1995; Dooge, 1986; MacDonnell et al., 2004; Olden et al., 2012). (Olden and Poff, 2003) examined the transferability of hydrologic indices across ‘stream types’ by identifying indices that consistently explain dominant patterns of variance across streams in varying climatic and geologic environments, such that the selected hydrologic indices are adequate to characterize flow regimes yet without redundancy information, that is, with the parsimonious selection of hydrologic indices for future hydrogeological studies.

Depending on the research objectives, different types of classification (or at least different descriptive terms) have already been in use, including those based on climate (humid versus arid, semi-arid, etc.), land cover (forested versus agriculture, urban, etc.), catchment response (fast versus slow flow), storage (groundwater dominated versus surface water dominated catchments), etc. The skill of the classification systems could depend on spatial and temporal scales. Olden and Poff (Olden and Poff, 2003) demonstrated that daily, monthly, and annual hydrologic indices are not always correlated, suggesting that different or independent information is present at these distinct temporal scales. Similarly, a number of studies (Atkinson et al., 2003; Farmer et al., 2003; Son and Sivapalan, 2007) showed that the dominant climatic and landscape controls on hydrologic behavior are time-scale dependent.

In this study, we aim to evaluate parameter transferability across the MOPEX basins and provide guidance on parameter screening and inversion design for CLM. Previous classification systems do not meet the specific goal to improve streamflow simulations across MOPEX basins

through model calibration, because they either do not cover the full range of time-scales, or do not account for model-specific behaviors. Instead of using existing classification schemes, we develop new classification indices based on scores of parameter sensitivity. The basis of the classification system is similarity of parameter significance across field sites with similar climate, soil, land cover/use, or geographic locations, noting that water and energy cycling processes are controlled predominantly by parameters associated with these properties. That is, sites within the same class share similar external features (e.g., hydrologic indices, forcings) and internal attributes (parameters/factors determining the system responses to external forcings) as characterized in the model. Determining these similar sites through cluster analysis, we expect that the parameter sensitivity, parameter calibration scheme, or even the parameter values, can be transferrable or shared between basins belonging to the same class. For example, sites in the same class are expected to share the same subset of parameters as the dominant ones, and which should be included in the unknown parameter list in the inversion modeling design.

The paper is organized as follows. In section 2, we describe the 431 MOPEX basins of study and the data used in the analysis; section 3 explains the task-parallel approach to run CLM4 simulations (431X128 runs) simultaneously for different parameter sets and field sites, and the clustering approaches. Sections 4 and 5 discuss the classification results, the applicability of the developed classification systems, and the transferability of model calibration design and calibrated parameter values.

2. Site and data description

Community Land Model (CLM) is used for the simulation in this study. It is a land model within the Community Climate System Model (Oleson, 2010). By coupling with atmospheric

model, it also provides estimation of surface albedos, upward longwave radiation, sensible heat flux, latent heat flux (LH), water vapor flux and surface CO₂ exchanges required by atmospheric models (Oleson, 2010). CLM has been widely applied at continental and global scales to understand how land processes and anthropogenic impact on land states affect climate and spatiotemporal change of the climate (Bonan et al., 2002; Dickinson et al., 2006). The Model Parameter Estimation Experiment (MOPEX) is an international project aimed at developing enhanced techniques for the a priori estimation of parameters in hydrologic models and land surface parameterization schemes of weather and climate models (Duan et al., 2006). A comprehensive MOPEX database has been developed that contains historical hydrometeorological data and land surface characteristics data for many hydrologic basins in the United States (US) and in other countries. This database has been expanded to include more and more basins in all parts of the world. The distribution of the 431 MOPEX basins in the United States is shown in Figure 1. Climate and environmental conditions vary from basin to basin. At each site, 10 hydrologic parameters with important impacts on the simulations of hydrologic process were selected for evaluation. For surface runoff process, three parameters were selected: the maximum fractional saturated area ($f_{\max}$), the shape parameters of topographic index distribution (C_s), and the decay factor representing the distribution of surface runoff with depth (f_{over}). For subsurface runoff process, the parameters include the decay factor representing the distribution of subsurface runoff with depth (f_{drai}), and the maximum subsurface drainage (Q_{dm}). The groundwater dynamics process contained one parameter, the average specific yield (S_y). The rest four parameters: the Clapp and Hornberger exponent (b), the saturated soil matrix potential (Ψ_s), Hydraulic conductivity (K_s), and porosity (θ_s), are associated with soil water process. For

more information about the physical definitions of these parameters, please refer to (Hou et al., 2012; Huang et al., 2013).

To meet the data requirement for CLM4, we retrieved input data and parameters for CLM4 at the 431 basins using recent datasets, consistent with Huang et al. (Huang et al., 2013), as follows:

(1) The meteorological forcing in this study was extracted from the phase two North America Land Data Assimilation System (NLDAS2) forcing at an hourly time step from 1979-2007 (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-hourly North American Regional Reanalysis (NARR) following the algorithms detailed in (Cosgrove et al., 2003).

(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. (Ke et al., 2012), by overlaying the watershed boundaries of the basins provided by National Weather Service (NWS) with the data layers. Specifically, soil information was extracted from the Soil maps for the State Soil Geographic (STATSGO) dataset and the dominant soil type within each basin was used to derive the soil hydraulic parameters. 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 Tables 1 and 2 in Ke et al. (Ke et al., 2012).

(3) The default values of the selected hydrologic parameters in CLM4 (i.e., f_{max} and C_s) were obtained by processing the 90-m resolution DEMs from HydroSHEDS (Lehner et al., 2008). Default values of f_{drai} and Q_{dm} were from the National Center for Atmospheric Research (NCAR). Default values of soil texture based parameters (S_y , Ψ_s , b , K_s and θ_s) and the corresponding range

information were obtained based on (Cosby et al., 1984; Hou et al., 2012), where soil information was extracted from the soil maps for the State Soil Geographic (STATSGO) dataset (Miller and White, 1998).

(4) A number of hydrologic indices at annual time scales were estimated based on algorithms detailed in (Brooks et al., 2011; Voepel et al., 2011) to facilitate the characterization of hydrologic regimes. For each basin, let P represent the mean annual precipitation, Q represent the mean annual streamflow, Q_s and Q_b represent the annual mean surface flow and baseflow, and PET represent the mean annual potential evapotranspiration, we define the hydrologic indices as follows: (a) the Runoff Ratio ($R = Q/P$); (b) the Baseflow ratio ($B = Q_b/Q$); (c) the Horton index ($H = (P-Q)/(P-Q_s)$), and (d) the humidity index ($HU = P/PET$). Interested readers are referred to (Huang et al., 2013) for detailed derivations of these indices.

3. Methodology

In this study, the integrated entropy concept and efficient sampling uncertainty quantification (UQ) framework is adopted to generate multiple parameter sample sets at each of the 431 MOPEX basins. CLM simulations were conducted individually (and in parallel) at each of the 431 MOPEX basins. At each site, one soil type and one veg type is assumed. Sensitivity of the CLM model outputs (e.g., runoff) to the hydrologic parameters is then analyzed, and the resulted parameter sensitivity scores are used in basin classification using clustering approaches. The classifications system is based on model responses of runoff, which is the focus at MOPEX basins. The clustering can be directly used for other related studies (e.g., model calibration) as long as runoff is the concern at the study sites.

3.1. Sensitivity analysis with an uncertainty quantification (UQ) framework

Significant biases have been observed in simulating energy fluxes and runoff at the MOPEX sites using default parameter values currently hard-coded in CLM4, which makes it necessary to optimize the model parameters through inversion/calibration using heat fluxes and/or streamflow observations. Given the high-dimensionality of input parameter space, and the complexity in model behavior, an inversion setup can be developed based on 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). Such a sensitivity analysis framework usually requires an efficient sampling technique (e.g., quasi Monte Carlo (QMC), Latin Hypercube sampling (LHS)) to explore the multi-dimensional parameter space, particularly when the numerical simulations are computationally demanding and the parameter dimensionality is high.

We adopted the UQ framework in (Hou et al., 2012; Huang et al., 2013) and derived the prior probability density functions (pdf) of the input parameters using minimum-relative-entropy (MRE) theory at each of the 431 MOPEX basins. The parameter space was explored using QMC sampling of the pdfs, and 128 QMC samples were generated using the assumed prior distributions according to the MRE criteria and prior information such as bounds and statistical moments, for examples, means and variances, of the input parameters (Hou and Rubin, 2005). The multiple sets of input parameters and forcing were then arranged into pseudo-arrays as described in (Tesfa et al., 2014) to take advantage of the parallel computing infrastructure of CLM. That is, for the default parameter set and the i^{th} sample set, CLM is run once by distributing the 431 basins to multiple nodes on a Linux cluster. The multiple CLM runs were

conducted in parallel and the simulation results are extracted for sensitivity analysis. Generalized linear model (GLM) analyses with the Akaike's information criterion (AIC, see (Akaike, 1974))-based backward removal approach were performed to identify the significant parameters for each climatologically averaged monthly output (e.g., runoff) for each output variable at each study site (Hou et al., 2012; Huang et al., 2013).

In order to illustrate the significance of the parameters and their interactions, the generalized linear model (GLM) was used to conduct statistical tests of the linear (the first order) and quadratic (the second order) terms of these parameters followed by stepwise model selection by the Akaike information criterion (AIC). Significance scores were assigned to each of the input parameters for all months across the 431 MOPEX basins. A 5-level scoring system is defined based on the P-values of the statistical significance tests. The parameter significance score is set to be 4, 3, 2, 1, and 0 for $P\text{-value} \leq 0.001$; $0.001 < P\text{-value} \leq 0.01$; $0.01 < P\text{-value} \leq 0.05$; $0.05 < P\text{-value} \leq 0.1$, and $P\text{-value} > 0.1$ (insignificant), respectively. The summation and standard deviation of scores of all months were also calculated for all 431 MOPEX basins.

After the model outputs are extracted and sensitivity analysis is done using GLM analysis, sensitivity scores of the input parameters/factors are available, together with external factors such as hydrologic indices, at each of the 431 MOPEX basins. These information/data, representing the internal or external attributes/behaviors of the hydrological systems, are then used for clustering analysis to divide the basins into groups.

3.2. Principal Component Analysis

240 One approach that can be used for identifying similarities in the above sensitivity analysis results
241 is Principal Component Analysis (PCA). Its operation can be thought of as revealing the internal
242 structure of the data in a way that best explains the variance in the data. If a multivariate dataset
243 is visualized as a set of coordinates in a high-dimensional data space (one axis per variable),
244 PCA can supply the user with a lower-dimensional picture from its most informative viewpoint.
245 By using only the first few principal components, the dimensionality of the transformed data can
246 be reduced.

247 The runoff responses (and the patterns of their sensitivity to hydrologic parameters) at the
248 431 MOPEX basins are more or less cross-correlated, particularly for those basins that are
249 geographically close or have similar climate/soil conditions. The similarities in parameter
250 sensitivity patterns (and external forcing) can be analyzed and extracted using Principal
251 Component Analysis (PCA), which is to express basin similarity behavior and/or characteristics
252 in a much lower dimensional subspace. PCA uses orthogonal transformation to convert a set of
253 observations of possibly correlated variables into a set of values of linearly uncorrelated
254 variables called principal components. The first principal component has the largest possible
255 variance and accounts for the most variability in the data, and each succeeding component in turn
256 has the highest variance possible under the constraint that it is orthogonal to (i.e., uncorrelated
257 with) the preceding components.

258 The PCA analyses help identify similarities among parameter significance patterns across
259 basins, identify dominant factors as well as factors that have duplicate impacts, and therefore
260 provide information whether a grouping/clustering can be done and which factors should be
261 included. By combining with other quantitative clustering approaches (e.g., K-means, EM-

clustering), PCA helps understand which factors and how they are used for discriminating the individuals to be classified.

3.3. Expectation-Maximization clustering

Cluster analysis or clustering is the task of grouping a set of objects in such a way that objects in the same group (called cluster) are more similar (in one way or another) to each other than to those in other groups (clusters). It can be achieved by various algorithms that differ significantly in their notions of what constitute a cluster and how to efficiently find them. Popular notions of clusters include groups with smaller distances among the cluster members, dense areas of the data space, intervals, or particular statistical distributions(Jain et al., 1999). The appropriate clustering algorithm and parameter settings (including values such as the distance function to use, a density threshold, or the number of expected clusters) depend on the individual data set and intended use of the results.

Probability models have successfully formed a basis for cluster analysis. According to recent studies, clustering procedures based on probability models are increasingly favorable over heuristic methods (Fraley and Raftery, 1998; Fraley and Raftery, 2002; Lau and Green, 2007). A general model-based framework for clustering of multivariate data based on a mixture of Gaussian distributions was proposed by parameterizing the covariance matrices in terms of their eigenvalue decompositions (Banfield and Raftery, 1993). Each cluster corresponds clearly to a particular characteristic (orientation, volume, or shape) and it is allowed to vary for clusters. A maximum likelihood criterion of Gaussian clustering was developed by (Banfield and Raftery, 1993) to allow characteristics to vary for clusters, with the shapes of the clusters preserved but orientation and volumes varied.

Expectation-Maximization (EM) algorithm is a general approach for approximating maximum likelihood estimations in the presence of incomplete data (Dempster et al., 1977). Its use to carry out the maximization of mixture models has been fully explored in (Redner and Walker, 1984). Compared with other methods, The EM algorithm stands out as an objective clustering solution and offers uncertainty measures of resulting classification as well (Fraley and Raftery, 1998; Fraley and Raftery, 2007a). To determine the number of components/classes, Bayes factor, which is an approximation of the Bayesian information criterion (BIC), can be adopted (Dasgupta and Raftery, 1998) to compare models in terms of the number of components in the mixture and the underlying densities of various components.

In this paper, model-based clustering is exploited by EM-clustering ((Fraley and Raftery, 1998; Fraley and Raftery, 1999; Fraley and Raftery, 2007b). The clustering package is available from the Comprehensive R Archive Network (CRAN) ([http:// CRAN.R-project.org](http://CRAN.R-project.org)). In model-based clustering, it is assumed that data are generated by a mixture of underlying probability distributions in which each component represents a different cluster. The model for the composite of the clusters is usually formulated in two ways: the classification likelihood and the mixture likelihood approach. EM-clustering assumes a Gaussian mixture model, and the mixture likelihood approach maximizes

$$\prod_{i=1}^n \sum_{k=1}^G \tau_k f_k(\mathbf{x}_i | \boldsymbol{\theta}_k), \quad (1)$$

where $\mathbf{x} = (x_1, \dots, x_n)$ represents the data, G is the number of components in the mixture, τ_k is the probability that an observation belongs to the k^{th} component ($\tau_k \geq 0$; $\sum_{k=1}^G \tau_k = 1$). The parameters $\boldsymbol{\theta}_k$ consist of a mean vector μ_k and a covariance matrix Σ_k , and the density has the form

$$f_k(\mathbf{x}_i | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) = (2\pi)^{-p/2} |\boldsymbol{\Sigma}_k|^{-1/2} \exp \left\{ -\frac{1}{2} (\mathbf{x}_i - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1} (\mathbf{x}_i - \boldsymbol{\mu}_k) \right\}, \quad (2)$$

where the covariance matrix $\boldsymbol{\Sigma}_k = \lambda_k D_k A_k D_k^T$, D_k is the orthogonal matrix of eigenvectors that can determine the orientation of the principal components of $\boldsymbol{\Sigma}_k$, A_k is a diagonal matrix whose elements are proportional to the eigenvalues of $\boldsymbol{\Sigma}_k$, and λ_k is a scalar that controls the shape of the density contours and specifies the volume of the corresponding ellipsoid. These parameters enable the flexibility of using different orientations, shapes, and volumes for the classes. Allowing adjustments in the parameters λ_k , D_k , and A_k , the resulted classes from EM-clustering are ellipsoidal with orientation, shape, and volume that vary from class to class. Because of the many options of assigning ensemble members to classes, it is important to adopt a well-accepted criterion to choose among the many models by penalizing the number of classes while honoring the likelihood of a member belonging to its assigned class. The model selection can be based on BIC that has the form

$$BIC \equiv 2 \loglik_{\mathcal{M}}(\mathbf{x}, \theta_k^*) - m \log(n), \quad (3)$$

where the $\loglik_{\mathcal{M}}(\mathbf{x}, \theta_k^*)$ is the maximized mixture loglikelihood for the model, m is the number of independent parameters ($\theta_k^*, k = 1, \dots, m$) to be estimated in the model, and n is the number of observations in the data $\mathbf{x} = (x_1, \dots, x_n)$. The larger the value of BIC, the stronger the evidence (i.e., the more likely) supporting the selection of the model and number of clusters.

3.4. Parameter transferability evaluation through model calibration

To evaluate the effectiveness of the above classification system, two levels/options of parameter transferability can be studied: (1) the transferability of patterns of parameter

contribution and therefore model calibration strategies; and (2) the transferability of parameter values. Firstly model calibrations are conducted at representative sites from each of the sensitivity based classes, where each class has a unique subset of dominant parameters to be inverted, and the calibrations are validated to prove the first level of parameter transferability; next, the calibrated parameters are compared between and within classes, the parameter values would be transferrable if the calibrated values vary less within classes than between classes. The similarities and discrepancies between the calibrated parameter values are then summarized and explained based on knowledge about the climate, soil, and hydrologic attributes of the study sites.

At MOPEX basins, we are more interested in the runoff responses, but even with runoff as the focus, there are still many choices as to what quantity/attributes of runoff observations to be used for estimating the model parameters. Various metrics can be computed for runoff (e.g., time of concentration, flow duration curve (FDC) high/low flow rates, wavelet power coefficients (WPC), see (Hou et al., 2015) (Bao et al., 2015). All these metrics can be used for designing objective functions for either deterministic or stochastic inversion. Considering the computational demand, we've adopted surrogate-based model for inversion, and the idea is to first develop surrogates linking the input parameters and output responses/metrics (e.g., runoff and its WPC) (Gong et al., 2014; Ray et al., 2015; Ray et al., 2012). A surrogate is an explicit form of the relationship links model inputs and outputs and its applicability is dependent on the complexity and nonlinearity of the relationship, and therefore needs validation test before being applied for further analysis (e.g., used as efficient alternative to the expensive numerical CLM simulator). After reasonable surrogates (e.g., for runoff WPC) are developed and tested, they can be used for model calibration, and by fitting observed WPCs we expect to improve the simulated runoff series to match the temporal variations of the observed runoff.

Wavelet transform is used to analyze the simulation and observed runoff time series and compute the wavelet power coefficients. The computed WPCs for both the simulations and observational data are at 49 temporal scales, ranging from 1 day to 4096 days. With wavelet transform, the long time series of observations and simulation data, which are all mixture of signals of multiple temporal scales, are summarized as power energy and coefficients at 49 temporal scales. This reduces the number of surrogate models needed for calibrations. Model optimization/calibration usually requires thousands of sequential trials to find the optimal parameter values. Since CLM is a computationally expensive model, it is usually impractical to do model optimization/calibration directly using CLM. There are different ways developing surrogate models which are cost effective for calibration purpose. In this study, the multivariate adaptive regression splines (MARS) method (Friedman, 1988; Friedman, 1991; Friedman and Silverman, 1989) (Friedman and Fast, 1993) is applied to build the surrogate model at each temporal scale, using the R package ‘earth’ (Milborrow and Milborrow, 2007). The MARS model is a form of a non-parametric regression technique and can be seen as an extension of linear models that automatically model non-linearities and interactions between variables. The final MARS model is a summation of multiple basis functions. The basis function can be a constant, a hinge function, or a product of two or more hinge functiona. For building the surrogate model in this study, for each investigated site, 256 CLM simulations results were used for representing the relationship between the input parameters and the WPCs. Therefore, at each temporal scale, there are 256 realizations of wavelet power coefficients from the ensemble simulations, and the surrogate model was built to link the input parameters and the WPCs, one suggorate for runoff WPC at each temperal scale. The resulted 49 composite surrogate models can be used to predict the WPCs given arbitrary input parameter values.

The differential evolution (DE) method has been successfully used for multidimensional real-valued functions optimization (Rocca et al., 2011), and is adopted for global optimization to estimate the unknown input parameters with the finalized surrogates. The DE approach does not use the gradient of the problem being optimized and is appropriate for optimization with data noises. The parallel computational R package ‘DEoptim’ (Mullen et al., 2011) is used for implementation of the DE optimization. The optimization goal is to minimize discrepancies between the surrogate-predicted WPCs and the observed WPCs, which is defined as the residual sum of squares (RSS) over total sum of squares (TSS) (relative error=RSS/TSS).

4. Results

4.1. Parameter significance patterns

Sensitivity analysis are applied to runoff for each basin and then summed to get sensitivity scores. The significance test scores for the linear and polynomial terms of 431 basins are assigned separately, and are shown in Figure 2. Obviously, both the linear and nonlinear impacts of the input parameters vary from basin to basin. f_{drai} , Q_{dm} , and S_y are the three dominant parameters in general, but f_{drai} and Q_{dm} seem to have secondary impacts at a number of basins, while S_y becomes insignificant at some basins.

The overall sensitivity scores for the 10 hydrological parameters based on their impacts on runoff are summarized in Figure 3 and consistent with findings from previous work (Hou et al., 2012; Huang et al., 2013). At many sites, the most significant parameters for runoff are f_{drai} , Q_{dm} , and S_y , followed by the secondary parameters θ_s , f_{over} , C_s , K_s , b , Ψ_s , and f_{max} . Therefore, it could be meaningful to use the three most sensitive parameters for global inversion if the research interest is at the global scale or similar. If the interest lies in improving hydrologic

responses at the watershed level, classification is warranted. Specifically, the top three parameters actually vary across sites. For example, at some sites, there exists a 4th parameter which is comparable to the top three parameters in terms of its sensitivity. Under such circumstances, it is preferable to classify the MOPEX basins and consider the class-specific dominant parameter sets for model prediction improvement.

4.2. Principal components of parameter significance

PCA analyses are based on the parameter sensitivity scores of all input parameters across the 431 MOPEX basins based on monthly runoff values defined as the responses variable. As a comparison, the analysis is also conducted using hydrological indices at the 431 basins. In the biplot of PCA analysis, the horizontal and vertical axes show the projections onto the first and second principal components, respectively. The MOPEX basin index (Sivapalan et al., 2003) numbers ($MBI = 1, \dots, 431$) are used to indicate their projections on the two principal components. The red arrows illustrate the projections of the original features associated with each parameter onto the principal components.

As shown in Figure 4, Q_{dm} and f_{drai} have similar behaviors in influencing runoff responses, and they contribute to both the first and second principal components, while S_y contribute more to the second principal component than to the first component. Section 4.1 shows that these three parameters have the highest significance scores; while PCA study provides supplementary information that the scores of these parameters are also the major contributors to the overall variability in the parameter scoring system. In other words, they are the major factors differentiating different sites (and grouping similar sites) in terms of parameter identifiability/significance. A further look into the similarity of the parameters influencing the

simulated runoff is done using a dendrogram (see Figure 5), which is a tree diagram frequently used to illustrate the arrangement of the clusters produced by hierarchical clustering of samples (here the sensitivity scores). According to their impacts on runoff simulations, the input parameters can be roughly divided into three types: soil parameters (S_y , θ_s , K_s , Ψ_s), subsurface runoff parameters (f_{drai} , Q_{dm}), and surface runoff parameters (f_{max} , C_s , b , f_{over}). The above analyses help identify similarities among parameter significance patterns across basins, identify dominant factors as well as factors that have duplicate impacts, and therefore provide information whether a grouping/clustering can be done and which factors should be included.

4.3. EM-clustering of MOPEX basins based on parameter sensitivity patterns

The above analyses demonstrate the feasibility of clustering the MOPEX basins based on parameter significance patterns. However, various clustering approaches may yield different classes. Traditional clustering approaches (e.g., K-means) could be arbitrary with a predetermined number of classes, and usually do not have the flexibility to vary the orientation, shape, and volume of the classes. However, in reality the classes are expected to vary in these attributes. Here we adopt a stochastic expectation-maximization clustering approach that has options to automatically choose the orientation, shape, and volume parameters for the classes. In order to determine the number of classes in an objective way, Bayesian information criterion (BIC) is used to obtain a compromise between the number of classes and the likelihood of a member being assigned to a class.

In order to select the best clustering configuration and their parameters, we compare different clustering configurations using BIC. In this study we consider two clustering configurations: (1) the VVI configuration that assumes flexibility in the orientation and shape of

the classes; and (2) the VVV configuration that allows orientation, shape, as well as volume/size to be adjustable. As shown in Figure 6, the VVV model stands out for being capable of clearly identifying ‘good’ parameters (i.e., parameters resulting in larger sensitivity) against the bad ones, and is adopted in the basin classification system based on parameter sensitivity and hydrologic indices in this study, as will be discussed in section 4.4.

4.4. Map of classification

4.4.1. Sensitivity-based classes (S-Class)

The BIC-integrated EM-clustering analysis yields six classes based on the sensitivity of runoff simulations to the variations of the input parameters. Each MOPEX basin is then assigned a class ID accordingly (see Figure 7). The classification is based on sensitivity scores; therefore, the resulting classification system is called S-class hereafter.

The MOPEX basins tend to aggregate geographically in a bone-like structure, for example, those near the Great Lakes, along Appalachian Mountains, and in the Great Plains and Texas. One exception is class #5, with its members scattered across the US due to the complex impacts of soil and climate conditions. Such classification is based on parameter sensitivity, which is dependent on both internal and external attributes/forcing/processes. In the next subsections, we evaluate the impacts of hydrological indices and climate types on the resulted S-classes to explore the reason for certain basins being grouped together.

4.4.2. Hydrologic indices-based classes (H-Class)

Five hydrological indices are considered in the PCA study, and they are runoff ratio (the ratio of annual total runoff to annual total precipitation), humidity index, baseflow index (the

ratio of annual baseflow in a river to the annual total runoff), Horton index ($[\text{total precipitation} - \text{total runoff}] / [\text{total precipitation} - \text{surface/fast runoff}]$), and drainage density (the total length of all the streams and rivers in a drainage basin divided by the total area of the drainage basin). H-Classes obtained by EM-clustering of the basins based on their hydrological indices are shown in Figure 8.

As mentioned above, different from the S-classes that are associated with hydrological processes expressed as parameterizations in CLM influenced by different parameter values, the H-classes represent attributes of the watersheds derived based on observations. Still we found that S-class #3 (the Great Lakes region) corresponds well to H-class #5, and that S-class #2 (roughly the Appalachian area) matches H-class #1. In addition, although many basins are grouped together in Southeast US with similar climate (H-class #4) given hydrological indices, they are considered to be within several separate S-classes with quite different parameter sensitivity patterns. The differences between the H- and S-classes show that although observationally based watershed attributes might have important impacts on the hydrological responses of these watersheds, the representation of hydrologic processes in CLM4, in the forms of parameterizations could be different from those manifested through the hydrologic indices because model sensitivity to parameters depends on how hydrological processes are parameterized and how uncertain the input parameters are. In addition, uncertainty in both model parameters and forcing could drive the hydrologic behaviors at the watersheds to a different regime (e.g., in this context, a different cluster). On the other hand, the forcing as well as the observations used to derive the hydrologic indices are also subject to uncertainties from instrument errors and derivation procedures, and various time scales of interest (e.g., the indices used are at an annual time step). All these factors could contribute to the mismatches between S-

classes and H-classes, and further demonstrate the necessity of a new classification based on model sensitivity to guide inversion setup and transferability studies.

4.4.3. Koppen climate classes (K-Class)

For the purpose of comparison, we retrieve the Koppen climate classes (Peel et al., 2007) of the MOPEX basins, as shown in Figure 9, and denoted as K-class. More than half of the basins are grouped into K-classes Cfa (a humid subtropical climate characterized by hot, humid summers and generally mild winters) and Dfa (a humid continental climate typified by large seasonal temperature differences, with warm to hot (and often humid) summers and cold (sometimes severely cold) winters), although they belong to several different smaller classes in the H- and S-classification systems. In fact, comparing the H- and K-classes, it is obvious that by taking into account of hydrologic information, the H class is more refined in reflecting characteristics that are intrinsic to the hydrologic regimes, while the K-classes are purely based on climate mainly using annual and monthly averages of temperature and precipitation, without considering the topographical and soil conditions. However, based on our sensitivity analyses, some soil parameters are among dominant factors determining water and energy fluxes. Despite the differences, the three clustering systems seems to agree the most near the Great Lakes area, which has a hot summer subtype in humid continental climate, a relatively high Horton index, with drainage and storage parameters dominating the model runoff responses. For other regions, especially for the southeast US with humid subtropical climate and high humidity index, the single K-class Cfa and H-class #4 seem to be too coarse given the variability in soil, weather, and terrain and basin attributes.

A new classification system is therefore warranted as the classification system should vary depending on the research objectives. Here in the study of the MOPEX basins, one of the major purposes is to improve streamflow prediction. Our S-class is more about parameter transferability and similarity in parameter sensitivity patterns, and is focused on runoff responses. The K/H-classes provide information to explain why the parameters are grouped together at certain regions (e.g., the Great Lake region and the Appalachian area), but climate/hydrologic conditions are inadequate to determine whether the hydrological parameters/factors are transferrable and/or which parameters/factors are invertible.

5. Discussion

Taking advantage of the USGS streamflow observations at the U.S. MOPEX basins, we focus on the runoff-based classification system for model calibration or parameter inversion design in following studies with the goal to improve runoff simulations.

5.1. Effects of parameter sensitivity scoring

The scoring system for the parameter significance has been explained in section 4.1. Since there is no standard way of defining parameter sensitivity quantitatively, we evaluated the effect of a different scoring system, in order to check whether the classification system is consistent. Here we adopted a 2-level scoring system by using p-value 0.05 as the threshold. The summation and standard deviation of scores of all months are then calculated for the 431 MOPEX basins.

Figure 10 shows the updated scores of the parameters/factors using a 2-level scoring system. The score distributions are very close to those obtained from the 5-level scoring system

(Figure 2). Therefore, the resulted classification system changed very little, and is insensitive to the ways of defining sensitivity scores of the parameters.

5.2. Guidance on parameter inversion at all MOPEX basins

The resulted clusters of the MOPEX basins have the potential to reduce the parameter calibration effort in that the parameter values, or at least parameter sensitivity patterns, are similar within each class. Given a MOPEX basin, instead of inverting a full set of unknown parameters, one would identify its class first, and then determine the dominant subset of unknowns to be included in the calibration, as shown in Figure 11 and Table 1. For example, for MOPEX basins of S-classes #1, the three parameters f_{drai} , Q_{dm} , and S_y , followed by the parameter B , are potentially the most identifiable parameters with runoff responses since they contribute the most to the runoff total variability and therefore should be included in the unknowns; for S-classes #2 and #3, f_{drai} , Q_{dm} , and S_y could be the most identifiable; f_{over} is also an important parameter for class #4; while Ψ_s is critical for classes 5 and #6. This information provides guidance on which parameters are to be included in the inversion design, such as the one demonstrated in (Sun et al., 2013), in which the hydrologic parameters in CLM were inverted using latent heat and recorded streamflow using a Markov Chain Monte Carlo approach. Such parameter sensitivity behaviors can be partially explained by the hydrologic conditions (see Figure 12). For example, as mentioned above, H-class #5 corresponds to S-class #3, where f_{drai} , Q_{dm} , and S_y are dominant, which is reasonable since the Horton index is high and therefore drainage and storage processes are important. Similarly, H-class #1 corresponds to S-class #2, where specific yield is one of the dominant parameters since humidity index is high while runoff ratio is low at the corresponding sites.

5.3. Evaluation of parameter transferability within and between classes

The above analyses enable further evaluation of transferability of parameter values, or at least provide guidance on calibration strategies. To evaluate the transferability of model calibration strategies and parameter values, the wavelet power coefficients (WPCs) of the simulated and observed runoff time series are computed. Figure 13 shows an example of wavelet power distribution of runoff simulations and observations at one MOPEX site, at 49 different temporal scales ranging from 1 to 4096 days. Each thin line corresponds to the 49 wavelet power coefficients of runoff simulations from one of the ensemble parameter sets. It is obvious that simulations using the default parameter values systematically underestimate the wavelet power (i.e., variability) at intermediate time scales (e.g., weekly, monthly, quarterly, yearly), which confirms the necessity of model calibration. Surrogates are developed at each of the temporal scales using MARS, and validated before being used for calibration. The surrogate validation tests involve calculation of the training/test errors and correlation coefficients between observations and surrogate predictions. At each site, 85% of the parameter samples and the corresponding simulations are used for surrogate training, and the remaining 15% sample/simulation sets are for testing. The training and testing errors are computed as the averages across a 100-fold test. Generally, training and testing correlation coefficients are above 0.8, and the errors are under 35% (residual sum square to total sum square of WPCs) at all temporal scales, which indicates that the surrogates are satisfactory. Also, the surrogate models are developed individually for each site where ensemble simulations are available for individual surrogate development.

567 The differential evolution (DE) optimization method is then applied for model
568 calibration using the surrogates (instead of the direct CLM simulations), by minimizing the
569 discrepancies between the surrogate-predicted WPCs and the observed WPCs. A set of
570 experiments of model calibration were conducted to evaluate the feasibility of calibration for
571 each class with reduced number of model parameters after parameter sensitivity analysis and
572 basin classification. The inversion for full suite of parameters is very ill-posed and under-
573 determined, and therefore the calibration results would be questionable. By reducing the
574 parameter dimensionality to a reasonably low number, the classification makes the inverse
575 modeling possible and less ill-posed. Figure 14 shows the WPCs of runoff predicted using the
576 optimized values of the subset parameters, at randomly selected sites. Three sites are randomly
577 selected from each of six classes. Models were calibrated based on simulations from year 2002 to
578 2008, and validated at different time periods up to 2010. Compared with the runoff WPCs
579 simulated using default parameter values, the WPCs of simulated runoff from the calibration
580 match the observations better at each site in each class. The predictions match the observed
581 WPCs very well, which means the variability at all temporal scales are well captured using the
582 calibrated parameter values. The overall improvement using the calibrated parameter values
583 demonstrate the effectiveness of the classification and the reduced parameter dimensionality in
584 the inverse modeling design. Note that the optimization enables improvement in the parameter
585 estimates, but doesn't address the topic of narrowing the bounds of parameters. A complete and
586 interesting study could be to explore the possibility of reducing uncertainty and evaluate
587 transferability of uncertainty reduction. Exceptions occur at a few sites, where model calibrations
588 have little impacts on matching the observed WPCs. At these sites, we found that the observed
589 WPCs barely overlap with the ensemble simulations, which indicates the dominance of model

structural errors over the parametric uncertainties, and that model improvement cannot be achieved by simply adjusting parameter values.

The transferability of model calibration strategies has been demonstrated with the results shown in Figure 14. Next we evaluate the transferability of optimized parameter values. Because of the significant variations in soil and climate conditions, we expect the parameter values to vary within and between classes. The three most common unknown parameters are f_{drai} , Q_{dm} , and S_y . We calculate the relative differences in calibrated parameter values within each class (see Table 2) and between classes (see Table 3-5). It indicates larger between-class variability than within-class if the values off the diagonal lines are bigger than those on the diagonal lines in Table 3-5. Note that each class has a different combinations of critical parameters, and trivial parameters have been excluded in the calibrations. As shown in Table 2-5, we can see that the three parameters values for classes #2, #3, and #4 are closer within each class than between classes. For example, as shown in Table 2, the relative differences within classes for f_{drai} are below 1.9; while the relative differences between the classes can reach up to 3.1. It means that these parameter values are more transferrable for sites in the same classes. But given the variations of the climate and soil conditions within the same class, we do not expect all the parameters to have exactly the same parameter values. We conducted further analysis to evaluate the relationships between hydrologic indices and optimized parameter values and found that such relationships are generally weak. Some positive correspondences exist between f_{drai} and the baseflow index, and between Q_{dm} and the Horton index. These findings are reasonable as baseflow index is a measure of the proportion of the river runoff derived from stored sources, and that the more permeable the rock, superficial deposits and soils in a catchment, the higher the baseflow and the quicker water is drained (i.e., higher f_{drai}); at sites with a low Horton index,

most precipitation leaves the catchment as stream flow, the maximum subsurface drainage (Q_{dm}) is generally limited. We also found that soil parameters such as Ψ_s are positively related to clay percentage, because clayey soils, with adhesive and osmotic binding, release water at lower (more negative) potentials. These above statements will be more solid after we do calibration at larger numbers of basins in the future.

6. Conclusions

In this study, we analyzed the sensitivity of CLM-simulated streamflow/runoff and heat flux responses to major hydrological parameters at 431 MOPEX basins across the US. A UQ framework consisting of importance sampling, exploratory data analyses, and HPC-enabled numerical simulations, was employed for the sensitivity analysis. Parameter significance tests were conducted and sensitivity scores were assigned to each input parameter/factor. The sensitivity analysis results are consistent with findings from previous work regarding parameter importance and potential identifiability. The resulted parameter significance scores are used to classify the MOPEX basins into different classes by grouping basins with similar parameter significance patterns. Such a parameter-sensitivity-based classification system helps reduce the complex climate/hydrologic system into subsets of more homogeneous and smaller systems, and provides necessary information to setup the parameter estimation or model calibration problem, using either deterministic or stochastic approaches. Classes based on parameter sensitivity, hydrologic indices, and climate conditions are compared, and their differences are related to the facts that both model parameters and forcing may drive the hydrologic behaviors at the watersheds to a different regime, and that the forcing as well as the observations used to derive the hydrologic indices are also subject to uncertainties from instrument errors and derivation

procedures, and various time scales of interest. These all necessitate a new classification system based on model sensitivity to guide inversion setup and transferability studies for specific models for efficiency and accuracy. The transferability of model calibration strategies is validated by model calibration results at representative sites from each class. The optimized values for the major parameters are somewhat more transferrable within each class than between classes. That is, their intra-class variability is generally smaller than inter-class variability. However, using the same parameter values for sites belonging to the same class may not yield the best model performance if the soil and climate conditions vary substantially within the class. Given the initial findings and somewhat mixed results, a more complete evaluation of parameter transferability would involve calibration over a larger number of basins building on the classification approach, using both deterministic and stochastic approaches to evaluate uncertainty reduction issues as well.

Acknowledgements

This work is funded by the U.S. Department of Energy Office of Science Biological and Environmental Research (BER) as part of the Earth System Modeling Program and also supported by the Office of Science Advanced Scientific Computing Research (ASCR). PNNL is operated for the US DOE by Battelle Memorial Institute under Contract DE-AC05-76RLO1830.

References:

- Akaike, H., 1974. New Look at Statistical-Model Identification. *Ieee T Automat Contr*, Ac19(6): 716-723.
- Ali, G., Tetzlaff, D., Soulsby, C., McDonnell, J.J., Capell, R., 2012. A comparison of similarity indices for catchment classification using a cross-regional dataset. *Advances in Water Resources*, 40: 11-22.
- Atkinson, S.E., Sivapalan, M., Woods, R.A., Viney, N.R., 2003. Dominant physical controls on hourly flow predictions and the role of spatial variability: Mahurangi catchment, New Zealand. *Advances in Water Resources*, 26(3): 219-235. DOI:[http://dx.doi.org/10.1016/S0309-1708\(02\)00183-5](http://dx.doi.org/10.1016/S0309-1708(02)00183-5)
- Banfield, J.D., Raftery, A.E., 1993. Model-Based Gaussian and Non-Gaussian Clustering. *Biometrics*, 49(3): 803-821. DOI:Doi 10.2307/2532201
- Bao, J., Huang, M., Hou, Z., Y., L., Ren, H., 2015. Assessing Approaches for Analyzing the Sensitivity of Simulated Land Surface Fluxes to Hydrologic Parameters in the Community Land Model. *Stochastic Environmental Research and Risk Assessment*.
- Bloschl, G., Sivapalan, M., 1995. Scale Issues in Hydrological Modeling - a Review. *Hydrol Process*, 9(3-4): 251-290. DOI:DOI 10.1002/hyp.3360090305
- Bonan, G.B. et al., 2002. The land surface climatology of the community land model coupled to the NCAR community climate model*. *Journal of Climate*, 15(22): 3123-3149.
- Brooks, P.D., Troch, P.A., Durcik, M., Gallo, E., Schlegel, M., 2011. Quantifying regional scale ecosystem response to changes in precipitation: Not all rain is created equal. *Water Resour. Res.*, 47(7): W00J08. DOI:10.1029/2010wr009762
- Carrillo, G. et al., 2011. Catchment classification: hydrological analysis of catchment behavior through process-based modeling along a climate gradient. *Hydrology and Earth System Sciences Discussions*, 8(3): 4583-4640.
- Coopersmith, E., Yaeger, M.A., Ye, S., Cheng, L., Sivapalan, M., 2012. Exploring the physical controls of regional patterns of flow duration curves - Part 3: A catchment classification system based on regime curve indicators. *Hydrology and Earth System Sciences*, 16(11): 4467-4482. DOI:DOI 10.5194/hess-16-4467-2012
- Cosby, B.J., Hornberger, G.M., Clapp, R.B., Ginn, T.R., 1984. A Statistical Exploration of the Relationships of Soil Moisture Characteristics to the Physical Properties of Soils. *Water Resour. Res.*, 20(6): 682-690. DOI:10.1029/WR020i006p00682
- Cosgrove, B.A. et al., 2003. Real - time and retrospective forcing in the North American Land Data Assimilation System (NLDAS) project. *Journal of Geophysical Research: Atmospheres* (1984 - 2012), 108(D22).
- Dasgupta, A., Raftery, A.E., 1998. Detecting features in spatial point processes with clutter via model-based clustering. *Journal of the American Statistical Association*, 93(441): 294-302. DOI:Doi 10.2307/2669625
- Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum Likelihood from Incomplete Data Via Em Algorithm. *J Roy Stat Soc B Met*, 39(1): 1-38.
- Dickinson, R.E. et al., 2006. The Community Land Model and its climate statistics as a component of the Community Climate System Model. *Journal of Climate*, 19(11): 2302-2324.
- Dooge, J.C.I., 1986. Looking for Hydrologic Laws. *Water Resour Res*, 22(9): S46-S58. DOI:Doi 10.1029/Wr022i09sp0046s
- Duan, Q. et al., 2006. Model Parameter Estimation Experiment (MOPEX): An overview of science strategy and major results from the second and third workshops. *J Hydrol*, 320(1-2): 3-17. DOI:10.1016/j.jhydrol.2005.07.031
- Farmer, D., Sivapalan, M., Jothityangkoon, C., 2003. Climate, soil, and vegetation controls upon the variability of water balance in temperate and semiarid landscapes: Downward approach to water balance analysis. *Water Resour Res*, 39(2): 1035. DOI:10.1029/2001wr000328

700 Fraley, C., Raftery, A.E., 1998. How many clusters? Which clustering method? Answers via model-based
 701 cluster analysis. *Comput J*, 41(8): 578-588. DOI:DOI 10.1093/comjnl/41.8.578
 702 Fraley, C., Raftery, A.E., 1999. MCLUST: Software for model-based cluster analysis. *J Classif*, 16(2): 297-
 703 306. DOI:DOI 10.1007/s003579900058
 704 Fraley, C., Raftery, A.E., 2002. Model-based clustering, discriminant analysis, and density estimation. *J*
 705 *Am Stat Assoc*, 97(458): 611-631.
 706 Fraley, C., Raftery, A.E., 2007a. Bayesian regularization for normal mixture estimation and model-based
 707 clustering. *J Classif*, 24(2): 155-181. DOI:DOI 10.1007/s00357-007-0004-5
 708 Fraley, C., Raftery, A.E., 2007b. Model-based methods of classification: Using the mclust software in
 709 chemometrics. *J Stat Softw*, 18(6).
 710 Friedman, J., Fast, M., 1993. Technical Report LCS110. Stanford University, Stanford, CA.
 711 Friedman, J.H., 1988. Fitting functions to noisy data in high dimensions. Department of Statistics-
 712 Stanford University Technical Report(101): 1-36.
 713 Friedman, J.H., 1991. Multivariate adaptive regression splines. *The annals of statistics*: 1-67.
 714 Friedman, J.H., Silverman, B.W., 1989. Flexible parsimonious smoothing and additive modeling.
 715 *Technometrics*, 31(1): 3-21.
 716 Gong, W. et al., 2014. Multi-objective parameter optimization of common land model using adaptive
 717 surrogate modelling. *Hydrol. Earth Syst. Sci. Discuss.*, 11(6): 6715-6751. DOI:10.5194/hessd-11-
 718 6715-2014
 719 Hou, Z., Huang, M., Leung, L.R., Lin, G., Ricciuto, D.M., 2012. Sensitivity of surface flux simulations to
 720 hydrologic parameters based on an uncertainty quantification framework applied to the
 721 Community Land Model. *J. Geophys. Res.*, 117(D15): D15108. DOI:10.1029/2012jd017521
 722 Hou, Z. et al., 2015. Quantification of Structural Uncertainty in a Land Surface Model. Society for
 723 Industrial and Applied Mathematics (SIAM), CSE2015.
 724 Hou, Z.S., Rubin, Y., 2005. On minimum relative entropy concepts and prior compatibility issues in
 725 vadose zone inverse and forward modeling. *Water Resour Res*, 41(12). DOI:Artn W12425
 726 Doi 10.1029/2005wr004082
 727 Huang, M. et al., 2013. Uncertainty Analysis of Runoff Simulations and Parameter Identifiability in the
 728 Community Land Model: Evidence from MOPEX Basins. *Journal of Hydrometeorology*, 14(6):
 729 1754-1772. DOI:10.1175/jhm-d-12-0138.1
 730 Jain, A.K., Murty, M.N., Flynn, P.J., 1999. Data clustering: a review. *ACM computing surveys (CSUR)*,
 731 31(3): 264-323.
 732 Ke, Y. et al., 2012. Development of high resolution land surface parameters for the Community Land
 733 Model. *Geosci Model Dev*, 5(6): 1341-1362. DOI:DOI 10.5194/gmd-5-1341-2012
 734 Lau, J.W., Green, P.J., 2007. Bayesian model-based clustering procedures. *Journal of Computational and*
 735 *Graphical Statistics*, 16(3): 526-558.
 736 Lehner, B., Verdin, K., Jarvis, A., 2008. New Global Hydrography Derived From Spaceborne Elevation
 737 Data. *Eos, Transactions American Geophysical Union*, 89(10): 93-94.
 738 DOI:10.1029/2008eo100001
 739 MacDonnell, J.E., Lunn, F.A., Bearne, S.L., 2004. Inhibition of E. coli CTP synthase by the "positive"
 740 allosteric effector GTP. *Bba-Proteins Proteom*, 1699(1-2): 213-220. DOI:DOI
 741 10.1016/j.bbapap.2004.03.002
 742 McDonnell, J.J., Woods, R., 2004. On the need for catchment classification. *Journal of Hydrology*, 299(1):
 743 2-3.
 744 Milborrow, S., Milborrow, M.S., 2007. The earth Package. The R Project for Statistical Computing.

- Miller, D.A., White, R.A., 1998. A conterminous United States multi-layer soil characteristics data set for regional climate and hydrology modeling, *Earth Interactions*, 2. Web-based publication, Res. Citeseer. (available at <http://EarthInteractions.org>).
- Mullen, K., Ardia, D., Gil, D.L., Windover, D., Cline, J., 2011. DEoptim: An R package for global optimization by differential evolution. *J Stat Softw*, 40(6): 1-26.
- Olden, J.D., Kennard, M.J., Pusey, B.J., 2012. A framework for hydrologic classification with a review of methodologies and applications in ecohydrology. *Ecohydrology*, 5(4): 503-518. DOI:10.1002/Eco.251
- Olden, J.D., Poff, N.L., 2003. Redundancy and the choice of hydrologic indices for characterizing streamflow regimes. *River Research and Applications*, 19(2): 101-121. DOI:10.1002/rra.700
- Oleson, K., 2010. Technical description of version 4.0 of the community land model version (CLM), NCAR/TN-478+ STR. National Center for Atmospheric Research, Boulder, CO, USA. http://www.cesm.ucar.edu/models/cesm1.0/clm/CLM4_Tech_Note.pdf.
- Peel, M.C., Finlayson, B.L., McMahon, T.A., 2007. Updated world map of the Köppen-Geiger climate classification. *Hydrol. Earth Syst. Sci.*, 11(5): 1633-1644. DOI:10.5194/hess-11-1633-2007
- Ray, J., Hou, Z., Huang, M., Sargsyan, K., Swiler, L., 2015. Bayesian Calibration of the Community Land Model Using Surrogates. *SIAM/ASA Journal on Uncertainty Quantification*, 3(1): 199-233. DOI:doi:10.1137/140957998
- Ray, J., Sargsyan, K., Hou, Z., Huang, M., 2012. Bayesian calibration of the Community Land Model using surrogates, Sandia National Laboratories.
- Redner, R.A., Walker, H.F., 1984. Mixture Densities, Maximum likelihood and the EM Algorithm. *SIAM Review*, 26(2). DOI:10.1137/1026034
- Rocca, P., Oliveri, G., Massa, A., 2011. Differential evolution as applied to electromagnetics. *Antennas and Propagation Magazine, IEEE*, 53(1): 38-49.
- Rosero, E. et al., 2010. Quantifying parameter sensitivity, interaction, and transferability in hydrologically enhanced versions of the Noah land surface model over transition zones during the warm season. *J. Geophys. Res.*, 115(D3): D03106. DOI:10.1029/2009jd012035
- Sawicz, K., Wagener, T., Sivapalan, M., Troch, P., Carrillo, G., 2011. Catchment classification: empirical analysis of hydrologic similarity based on catchment function in the eastern USA. *Hydrology and Earth System Sciences*, 15(9): 2895-2911.
- Sivakumar, B., 2008. Dominant processes concept, model simplification and classification framework in catchment hydrology. *Stochastic Environmental Research and Risk Assessment*, 22(6): 737-748.
- Sivakumar, B., Singh, V., 2012. Hydrologic system complexity and nonlinear dynamic concepts for a catchment classification framework. *Hydrology and Earth System Sciences*, 16(11): 4119-4131.
- Sivapalan, M. et al., 2003. IAHS decade on Predictions in Ungauged Basins (PUB), 2003-2012: Shaping an exciting future for the hydrological sciences. *Hydrolog Sci J*, 48(6): 857-880. DOI:10.1623/hysj.48.6.857.51421
- Son, K., Sivapalan, M., 2007. Improving model structure and reducing parameter uncertainty in conceptual water balance models through the use of auxiliary data. *Water Resour Res*, 43(1): W01415. DOI:10.1029/2006wr005032
- Sun, Y., Hou, Z., Huang, M., Tian, F., Leung, L.R., 2013. Inverse modeling of hydrologic parameters using surface flux and runoff observations in the Community Land Model. *Hydrology and Earth System Sciences* 17: 4995-5011. DOI:doi:10.5194/hess-17-4995-2013
- Tesfa, T.K. et al., 2014. A subbasin-based framework to represent land surface processes in an Earth system model. *Geosci Model Dev*, 7(3): 947-963.
- van Werkhoven, K., Wagener, T., Reed, P., Tang, Y., 2009. Sensitivity-guided reduction of parametric dimensionality for multi-objective calibration of watershed models. *Advances in Water Resources*, 32(8): 1154-1169.

793 Voepel, H. et al., 2011. Quantifying the role of climate and landscape characteristics on hydrologic
 794 partitioning and vegetation response. *Water Resour. Res.*, 47(1): W00J09.
 795 DOI:10.1029/2010wr009944
 796 Wagener, T., Sivapalan, M., Troch, P., Woods, R., 2007. Catchment classification and hydrologic similarity.
 797 *Geography Compass*, 1(4): 901-931.
 798 Xia, Y., Ek, M., Wei, H., Meng, J., 2012. Comparative analysis of relationships between NLDAS - 2
 799 forcings and model outputs. *Hydrol Process*, 26(3): 467-474.

800

801

Figure captions:

Figure 1. The locations of 431 US MOPEX basins with adequate precipitation gauges. The border of each basin was marked in red color.

Figure 2. Summary sensitivity scores of the input parameters at the 431 MOPEX basins, with logarithmic runoff as the response variable. The scores are for parameters of a) linear main effects; b) linear and quadratic effects. A higher score means stronger sensitivity of model responses to the input parameter.

Figure 3. Sensitivity scores of the 10 hydrological parameters based on runoff.

Figure 4. PCA Biplot of the parameter sensitivity scores based on runoff responses at the 431 MOPEX sites.

Figure 5. Dendrogram of the similarity of parameter significance based on the parameter sensitivity scores for the runoff responses.

Figure 6. BIC for the clustering models against the number of classes for two model options “VVI” and “VVV”. The sensitivity scores for clustering are based on GLM analysis with log runoff as the response variable.

Figure 7. S-Class map of the 431 MOPEX basins classified using parameter sensitivity scores with runoff as the response variable in the GLM sensitivity analysis.

Figure 8. H-Class map of the 431 MOPEX basins classified based on hydrologic indices.

Figure 9. K-Class for 431 MOPEX sites based on Koppen climate classification. (BSH – hot semi-arid climates; BSk – cold semi-arid climates; BWk – cold desert climate; Cfa – humid subtropical climate; Cfb – oceanic climate; Csa – hot-summer Mediterranean climate; Csb – warm-summer Mediterranean climate; Dfa – hot summer subtype in humid continental climate; Dfb – mild to warm summer subtype in humid continental climate; Dfc – continental subarctic climate; Dsb – mild to warm summer subtype in humid continental climate with a dry summer; Dsc – perhumid Mediterranean snow climate; Dwa – hot summer subtype in humid continental climate with a dry winter.)

Figure 10. Summary score plots for 12 months by 2-level score system for 431 MOPEX sites. The input parameters are GLM fitted model with linear main effects about log runoff. A higher score means stronger sensitivity of model responses to the input parameter.

Figure 11. Summary of sensitivity scores based on runoff within each of the six S-classes.

Figure 12. Hydrology indices within each of the six H-classes. (HI1 – runoff ratio; HI2 – humidity index; HI3 – baseflow index; HI4 – Horton index; HI5 – drainage density).

Figure 13. Wavelet power coefficients of observed and simulations runoff time series.

Figure 14. Comparison of runoff WPCs among observations, model simulations with calibrated parameters, and simulations with default parameters.

Table 1. Dominant parameters for each class.

Class number	Parameter1	Parameter2	Parameter3	Parameter4
Class 1	F_{drai}	Q_{dm}	Sy	b
Class 2	F_{drai}	Q_{dm}	Sy	
Class 3	F_{drai}	Q_{dm}	Sy	
Class 4	F_{drai}	Q_{dm}	F_{over}	b
Class 5	F_{drai}	Q_{dm}	Sy	Ψ_s
Class 6	F_{drai}	Sy	Ψ_s	

Table 2. Relative differences in calibrated parameters within each class. The difference is defined as $\left| \frac{max-min}{mean} \right|$. Note that each class has a different combinations of critical parameters, and trivial parameters have been excluded in the calibrations.

	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6
F_{drai}	0.781	0.895	1.004	1.714	0.304	1.885
F_{over}				1.947		
$\log_{10}(Q_{dm})$	0.491	0.266	0.047	0.206		0.057
Sy	0.101	0.333	0.186		0.093	0.128
b	2.188			2.093		
$\log_{10}(\Psi_s)$					0.822	1.104

Table 3. Relative differences in calibrated F_{drai} within and between classes.

F_{drai}	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6
Class 1	0.781	1.046	1.075	1.981	0.825	1.729
Class 2		0.895	0.988	2.070	0.887	1.419
Class 3			1.004	2.277	0.896	1.598
Class 4				1.714	1.750	3.128
Class 5					0.304	1.403
Class 6						1.885

Table 4. Relative differences in calibrated $\log_{10}(Q_{dm})$ within and between classes.

$\log_{10}(Q_{dm})$	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6
Class 1	0.491	0.490	0.539	0.532	0.548	
Class 2		0.266	0.366	0.401	0.408	
Class 3			0.047	0.209	0.088	
Class 4				0.206	0.213	
Class 5					0.057	
Class 6						

857 **Table 5. Relative differences in calibrated S_y within and between classes.**

Sy	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6
Class 1	0.101	0.340	0.224		0.129	0.156
Class 2		0.333	0.463		0.321	0.395
Class 3			0.186		0.253	0.195
Class 4						
Class 5					0.093	0.185
Class 6						0.128

858

Figure 1. The locations of 431 US MOPEX basins with adequate precipitation gauges. The border of each basin was marked in red color.

Figure 2. Summary sensitivity scores of the input parameters at the 431 MOPEX basins, with logarithmic runoff as the response variable. The scores are for parameters of a) linear main effects; b) linear and quadratic effects. A higher score means stronger sensitivity of model responses to the input parameter.

Figure 3. Sensitivity scores of the 10 hydrological parameters based on runoff.

Figure 4. PCA Biplot of the parameter sensitivity scores based on runoff responses at the 431 MOPEX sites.

Figure 5. Dendrogram of the similarity of parameter significance based on the parameter sensitivity scores for the runoff responses.

Figure 6. BIC for the clustering models against the number of classes for two model options “VVI” and “VVV”. The sensitivity scores for clustering are based on GLM analysis with log runoff as the response variable.

Figure 7. S-Class map of the 431 MOPEX basins classified using parameter sensitivity scores with runoff as the response variable in the GLM sensitivity analysis.

Figure 8. H-Class map of the 431 MOPEX basins classified based on hydrologic indices.

Figure 9. K-Class for 431 MOPEX sites based on Koppen climate classification. (BSh – hot semi-arid climates; BSk – cold semi-arid climates; BWk – cold desert climate; Cfa – humid subtropical climate; Cfb – oceanic climate; Csa – hot-summer Mediterranean climate; Csb – warm-summer Mediterranean climate; Dfa – hot summer subtype in humid continental climate; Dfb – mild to warm summer subtype in humid continental climate; Dfc – continental subarctic climate; Dsb – mild to warm summer subtype in humid continental climate with a dry summer; Dsc – perhumid Mediterranean snow climate; Dwa - hot summer subtype in humid continental climate with a dry winter.)

Figure 10. Summary score plots for 12 months by 2-level score system for 431 MOPEX sites. The input parameters are GLM fitted model with linear main effects about log runoff. A higher score means stronger sensitivity of model responses to the input parameter.

Figure 11. Summary of sensitivity scores based on runoff within each of the six S-classes.

Figure 12. Hydrology indices within each of the six H-classes. (HI1 – runoff ratio; HI2 – humidity index; HI3 – baseflow index; HI4 – Horton index; HI5 – drainage density).

Figure 13. Wavelet power coefficients of observed and simulations runoff time series.

Figure 14. Comparison of runoff WPCs among observations, model simulations with calibrated parameters, and simulations with default parameters.