Hydrol. Earth Syst. Sci. Discuss., 11, 6715–6751, 2014
www.hydrol-earth-syst-sci-discuss.net/11/6715/2014/
doi:10.5194/hessd-11-6715-2014

Hydrology and Earth System Sciences
Open Access
Discussions

Multi-objective parameter optimization of common land model using adaptive surrogate modelling

W. Gong[1,2], Q. Duan[1,2], J. Li[1,2], C. Wang[1,2], Z. Di[1,2], Y. Dai[1,2], A. Ye[1,2], and C. Miao[1,2]

[1]College of Global Change and Earth System Science (GCESS), Beijing Normal University, Beijing, 100875, China
[2]and Joint Center for Global Change Studies, Beijing, 100875, China

Received: 30 April 2014 – Accepted: 23 May 2014 – Published: 24 June 2014

Correspondence to: W. Gong (gongwei2012@bnu.edu.cn)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

Parameter specification usually has significant influence on the performance of land surface models (LSMs). However, estimating the parameters properly is a challenging task due to the following reasons: (1) LSMs usually have too many adjustable parameters (20–100 or even more), leading to the curse of dimensionality in the parameter input space; (2) LSMs usually have many output variables involving water/energy/carbon cycles, so that calibrating LSMs is actually a multi-objective optimization problem; (3) regional LSMs are expensive to run, while conventional multi-objective optimization methods needs a huge number of model runs (typically $10^5 \sim 10^6$). It makes parameter optimization computationally prohibitive. An uncertainty qualification framework was developed to meet the aforementioned challenges: (1) use parameter screening to reduce the number of adjustable parameters; (2) use surrogate models to emulate the response of dynamic models to the variation of adjustable parameters; (3) use an adaptive strategy to promote the efficiency of surrogate modeling based optimization; (4) use a weighting function to transfer multi-objective optimization to single objective optimization. In this study, we demonstrate the uncertainty quantification framework on a single column case study of a land surface model – Common Land Model (CoLM) and evaluate the effectiveness and efficiency of the proposed framework. The result indicated that this framework can achieve optimal parameter set using totally 411 model runs, and worth to be extended to other large complex dynamic models, such as regional land surface models, atmospheric models and climate models.

1 Introduction

Land surface models (LSMs), which offer land surface boundary condition for atmospheric models and climate models, are widely used in weather and climate forecasting. They are also a tool for studying the impacts of climate change and

HESSD
11, 6715–6751, 2014

human activities on our environment. Parameters of land surface models usually have significant influence on their simulation and forecasting capability. It has been shown that tuning even one or two parameters may significantly enhance the simulation ability of a land surface model (Henderson-Sellers et al., 1996; Liang et al., 1998; Lohmann et al., 1998; Wood et al., 1998). How to specify the parameters in a LSM model properly, however, remains a very challenging task because the LSM parameters are usually not directly measurable at the scale of model applications.

Automatic optimization approaches have been frequently used in calibrating the parameters of hydrological models. There is a plethora of optimization approaches available, including single-objective optimization methods such as SCE-UA (Duan et al., 1992, 1993, 1994), SCEM-UA (Vrugt et al., 2003), genetic algorithm (Wang, 1991), and multi-objective optimization methods such as MOCOM-UA (Boyle et al., 2000; Boyle, 2000; Gupta et al., 1998; Yapo et al., 1998) and MOSCEM-UA (Vrugt et al., 2003).

Compared to traditional hydrological models, the parameter calibration approach has not been practiced as much in LSM community, especially for large spatial scale applications. The major obstacles to calibrating land surface models over a large spatial scale are: (1) there are too many parameters to calibrate, (namely, the curse of dimensionality in parameters); (2) dimensionality of the output space is too high (i.e., many processes such as water/energy/carbon cycles are simulated simultaneously); (3) conventional optimization methods, especially multi-objective approach, need a large number ($\sim 10^5$–10^6) of model runs; and the large complex dynamic system models such LSMs are usually expensive to run (i.e., costing many CPU hours). There have been numerous attempts to use multi-objective optimization to calibrate the parameters of land surface models and significant improvement in LSM performance measures as a result of optimization have been reported (Bastidas et al., 1999; Gupta et al., 1999; Leplastrier et al., 2002; Xia et al., 2002). However, the optimization efforts in the past were usually limited to cases studies involving only point or limited spatial domain-scale applications of LSMs (Liu et al., 2003, 2004, 2005). To take a multi-

HESSD
11, 6715–6751, 2014

objective optimization to the calibration of LSM parameters for large scale applications, the key is to know how to reduce the number of model runs to an appropriate level that we can afford.

In our recent works, we proposed a framework that can potentially reduce the number of model runs needed for parameter calibration of large complex system models (Wang et al., 2014). This framework involves the following steps: (1) a parameter screening step using global sensitivity analysis to identify only the most sensitive parameters to be included in the optimization; (2) surrogate modelling that can emulate the response surface of the dynamic system model to the change in parameter values; (3) an adaptive sampling strategy to improve the efficiency of the surrogate model construction; (4) a multi-objective optimization step to optimize the most sensitive parameters in the dynamic system models. In this paper, we will illustrate this parametric uncertainty quantification framework on the Common Land Model (CoLM), a widely used, physically based land surface model developed by Yongjiu Dai and colleagures (Dai et al., 2003, 2004; Ji and Dai, 2010). The work related to parameter screening and surrogate modeling based optimization (ASMO) method for single objective has already been published (Li et al., 2013; Wang et al., 2014). This paper will emphasize on the analysis of different surrogate model construction methods and multi-objective optimization method and results.

This paper contains the following parts: Sect. 2 introduces the basic information on CoLM, the study area and dataset, the adjustable parameters and the output variables to be analyzed; Sect. 3 presents an inter-comparison of 5 surrogate modeling methods, and discusses how many model runs will be sufficient to build a surrogate model for optimization; Sect. 4 carries out single and multiple objective optimization using an adaptive surrogate model based optimization strategy (ASMO); Sect. 5 provides the conclusion.

HESSD
11, 6715–6751, 2014

2 Experiment setup

2.1 Model and parameters

Common Land Model (CoLM) proposed by Yongjiu Dai and colleagues (Dai et al., 2003, 2004; Ji and Dai, 2010) is one of the most widely used land surface model in the world. It combines the advantages of Land Surface Model (LSM) (Bonan, 1996), Biosphere–atmosphere transfer scheme (BATS) (Dickinson et al., 1993) and Institute of Atmospheric Physics land-surface model (IAP94) (Dai and Zeng, 1997; Dai et al., 1998). CoLM considers physical processes of energy and water transmission in soil vegetation, snow cover and atmosphere. It also implements glacier, lake, wetland and dynamic vegetation processes. Similar to previous research in presented in Li et al. (2013), we select 40 adjustable parameters from CoLM. The parameter names, physical meanings and value ranges are shown in Table 1.

This study considers six output variables simulated by CoLM: sensible heat, latent heat, upward longwave radiation, net radiation, soil temperature and soil moisture. The Normalized Mean Squared Error is used as the objective function in our analysis:

$$\mathrm{NRMSE}_i = \frac{\sqrt{\sum_{j=1}^{N}\left(y_{i,j}^{\mathrm{sim}} - y_{i,j}^{\mathrm{obs}}\right)^2}}{\sum_{j=1}^{N} y_{i,j}^{\mathrm{obs}}} \quad (1)$$

where i is the index of output variables, j is the index of time step, N is the total number of observations, $y_{i,j}^{\mathrm{sim}}$ and $y_{i,j}^{\mathrm{obs}}$ are the simulated and observed values, respectively. Objective functions represent the performance of model simulation and a smaller objective function means better performance.

HESSD
11, 6715–6751, 2014

2.2 Study area and datasets

The study area and associated datasets are from the Heihe river basin, the same as in (Li et al., 2013). The Heihe river basin, which is located between 96°42′–102°00′ E and 37°41′–42°42′ N, is an inland river basin in the arid region of northwest China. The basin area is approximately 130 000 km^2 and its altitude varies from sea level to 5500 m. The Heihe river basin has a variety of land using types including forest, grassland, farmland, and glacier, among others, making it an ideal research region for LSM simulation. In this research we use the data from A'rou observation station located at the upstream region of the Heihe river basin. Its geographic coordinate is 100°28′ E, 30°08′ N, altitude is 3032.8 m above sea level and the land cover type is alpine steppe.

The forcing data used include downward shortwave and longwave radiation, precipitation, air temperature, relative humidity, air pressure and wind speed (Hu et al., 2003); and the observation data used to validate the simulation of CoLM include: sensible heat, latent heat, upward longwave radiation, net radiation, soil temperature and soil moisture. The soil temperature and moisture were measured at depth 10, 20, 40, and 80 cm. In CoLM, the soil is divided into 10 layers and the linearly interpolated to the measured depth. Currently we have 2 years observation data. The data from year 2008 was used for spin up and that of 2009 was used for parameter screening, surrogate modeling and optimization. The simulation time step is set to 30 min and the simulation outputs are averaged to 3 h in order to compare with the observation data.

3 Comparison of surrogate models

After the sensitive parameters are identified using global sensitivity methods (see Li et al., 2013), the next step is to calibrate the sensitive parameters through multi-objective optimization. Since the calibration of CoLM in real world applications can be very expensive, we aim to establish a surrogate model to represent the response surface of the dynamic CoLM. Surrogate model, also called response surface, meta-

HESSD
11, 6715–6751, 2014

model, statistical emulator, is a statistical model that describes the response of output variable to the variation of input variables. Because the surrogate model only considers the statistical relationship between input and output, it is usually much cheaper to run than the original large complex dynamic model ("original model" for short). Parameter optimization usually needs thousands, or even millions times of model runs. It will be impossible to calibrate large complex dynamic models if running the original dynamic model is too time consuming. If we can do parameter optimization with surrogate model instead of original model, the time of running original model will be dramatically reduced, making it possible to calibrate the large complex dynamic models, such as land surface models, atmospheric models, or even global climate models. However, optimization based on surrogate models may be a challenging task because the response surface might be very bumpy and has many local optima. Razavi et al. (2012) gave a comprehensive review of the surrogate modeling methods and applications in water resources, and discussed the pitfall of surrogate modeling as well.

In this research, we first compared 5 different surrogate models: multivariate Adaptive Regression Spline (MARS), Gaussian Process Regression (GPR), Random Forest (RF), Support Vector Machine (SVM), and Artificial Neural Network (ANN). A brief introduction of these methods is provided in the Appendix. To build a surrogate, we need to choose a sampling method first. The sampling method used in this study is Latin Hypercube Sampling (LH) (Mckay et al., 1979). The sample sizes are set to 50, 100, 200, 400, 800, 1200, and 2000. The inter-comparison results are shown in Fig. 1, in which the x axis is the sample size, and y axis is the NRMSE (i.e., the ratio of the root mean square error (RMSE) of the simulation model and the surrogate model). For each output variable, we only construct surrogate models for the most sensitive parameters based on the screening results of several global sensitivity methods performed by Li et al. (2013). The list of sensitive parameters is shown in Table 2.

Figure 1 indicated that: (1) the error of surrogate model decreases as the sample size increases. The error becomes stable when the sample size is larger than 400. More samples can reduce the error but the benefit of additional samples is marginal.

HESSD
11, 6715–6751, 2014

(2) Among the 5 different surrogate models, GPR has the best performance, while ANN ranks the second. RF and MARS have lower accuracy. For some output variables (e.g., sensible and latent heat), the performance of SVM seems good, while for other variables (e.g., soil temperature), SVM's performance is not satisfactory. (3) The convergence speed for the 6 output variables is different. The elbow points (i.e., the point at which the objective function value changes from rapid decrease to a gradual one) of net radiation, soil temperature and soil moisture are significantly at 200 sample points, while for sensible heat, latent heat and upward long-wave radiation, the elbow points are not clear. Since among all methods, the GPR method can stably give the best performance for all the 6 output variables, we choose GPR in the following multi-objective optimization analysis.

4 Optimization

4.1 Single-objective optimization

Before we conduct multi-objective optimization, we first carried out single-objective optimization for each output variable using the GPR surrogate model. The Shuffled Complex Evolution (SCE) method (Duan et al., 1992, 1993, 1994) is used to find the optima of the surrogate model. In order to figure out how many sample points are sufficient to construct a surrogate model for optimization, different sample sizes (i.e., 50, 100, 200, 400, 800, 1200, and 2000) are experimented. To evaluate the optimization results based on the surrogate model, we also set up two control cases: (1) no optimization using the default parameters as specified in CoLM. (2) Optimization using the original CoLM (i.e., no surrogate model is used). The second case is referred as "direct optimization" in the following text. The control cases are used to confirm the following hypotheses: (1) parameter optimization can indeed enhance the performance of CoLM. (2) Optimization using the surrogate model can achieve similar optimization result as using the original model, but with fewer sample points.

HESSD
11, 6715–6751, 2014

The converged parameters from the single-objective optimization runs are shown in Fig. 2. In each subfigure the optimal parameter values are normalized to [0, 1]. The bold black line is the optimal parameter set obtained by direct optimization using the original CoLM, and other lines are optimal parameters given by surrogate models created with different sample sizes. Table 3 summarizes the optimized NRMSE values of all surrogate model based optimization runs with different sample sizes, as well as the control cases. The numbers of original model runs that SCE takes are also listed in the brackets.

The optimization results reveal that: (1) parameter optimization can significantly improve the simulation ability of CoLM significantly for all output variables. (2) For sensible heat, upward longwave radiation, net radiation, soil moisture, the optimal parameters given by the surrogate model optimization runs are very similar to those obtained by direct optimization. The optimal parameters obtained for different sample sizes are also close to each other. For latent heat and soil temperature, however, the optimal parameters given by surrogate model optimization and direct optimization are significantly different. The discrepancy between the results with different sample size is also significant, comparing to the previous 4 outputs. (3) Surprisingly, for some variables (sensible heat, upward longwave radiation, net radiation, soil moisture), sample size does not have significant influence on the optimization results. As shown in Table 3, even a surrogate model constructed with 50 samples is similar to the one constructed with 2000 samples and with the direct optimization. For soil temperature, 200 samples are sufficient, and for latent heat, more than 400 samples are enough. The result confirms the findings in Sect. 3 that approximately 200 sample points might be sufficient to construct a surrogate model for parameter optimization. This finding is very meaningful because it implies that we can dramatically reduce the number of model runs required by optimization down to hundreds of times. (4) The number of original model runs that SCE takes before converge is also listed in Table 3. The result indicated that SCE can get better, or similar optimal NRMSE, but the cost of original model runs is larger than using surrogate model. If the original dynamic model costs

HESSD
11, 6715–6751, 2014

too much CPU time to run, surrogate based optimization can be more efficiency than SCE. (5) Different output variables require different optimal parameters, indicating the necessity of multi-objective optimization. For example, P6, the Clapp and Hornberger "b" parameter, is sensitive to many outputs. For sensible heat, latent heat and soil moisture, the optimal value for P6 is high, while for upward longwave radiation, net radiation and soil temperature, the optimal value for P6 is low. In order to balance the performance measures of all output variables, it is necessary to choose a compromised value for P6. Multi-objective optimization is an objective approach that can provide such a compromised optimal parameter that balances the requirements of many output variables.

4.2 Multi-objective optimization

The results of single-objective optimization from previous section have highlighted the necessity for multi-objective optimization. Many multi-objective optimization methods have been proposed and validated in numerous studies (Boyle et al., 2000; Boyle, 2000; Gupta et al., 1998, 1999; Yapo et al., 1998; Vrugt et al., 2003; Bastidas et al., 1999; Leplastrier et al., 2002; Pollacco et al., 2013; Xia et al., 2002), but in the context of this research, we need a method that can satisfy the following conditions: (1) the method should be compatible with surrogate model optimization; (2) for practical reasons, it should provide a single best parameter set instead of a full Pareto optimum set with many non-inferior parameter sets. The Pareto set with hundreds, or more parameter sets have appeals in that it can provide an ensemble of model outputs, which in turn can be used to assess parametric uncertainty. For large complex dynamic models such as CoLM, it is generally impractical to run the model in an ensemble mode with hundreds of model runs. In this study, we use a weighting function method to convert the multi-objective optimization into a single objective optimization. The weight assigned to each objective function is based on Liu et al. (2005). The general idea is that we assign more weight to the objective function of an output, if that output is simulated more poorly as compared to the other outputs. Table 4 shows the RMSE

HESSD
11, 6715–6751, 2014

and NRMSE of CoLM using default parameterization scheme, and the weight of each output is proportional to the NRMSE.

After the weights are determined, the weighted objective function is as follows:

$$F = \sum_{i=1}^{n} w_i \mathrm{RMSE}_i \tag{2}$$

in which the RMSE_i is the Root Mean Squared Error of each output variable that defined as $\mathrm{RMSE}_i = \frac{1}{N}\sqrt{\sum_{j=1}^{N}\left(y_{i,j}^{\mathrm{sim}} - y_{i,j}^{\mathrm{obs}}\right)^2}$, w_i is the weight of each output, and $\sum_{i=1}^{n} w_i = 1$.

In order to use the information offered by surrogate model more effectively, we developed an adaptive surrogate modeling based optimization method called ASMO (Wang et al., 2014). The major steps of ASMO are as follows: (1) construct a surrogate model with initial samples, and find the optimal parameter of the surrogate model. (2) Run the original model with this optimal parameter and get a new sample. (3) Add the new sample to the sample set and construct a new surrogate model, and go back to the 1st step. The effectiveness and efficiency of ASMO have been validated in Wang et al. (2014) using 6-D Hartman function and a simple hydrologic model SAC-SMA. As shown in the comparison between ASMO and SCE-UA, ASMO is more efficient that can get archive convergence with less model runs, while SCE-UA is more effective that can get closer to the true global optimal parameter. So making a choice between ASMO and SCE-UA is a "cost-benefit" trade-off: if the model is very cheap to run, SCE-UA is preferred because it is more effective to find the global optimum; while if the model is very expensive to run, ASMO is preferred because it can find a fairly good parameter within a limited time of model runs. Such parameter set might not be the true global optimum, but it is the "not bad" solution that is cheap enough we can afford.

HESSD
11, 6715–6751, 2014

We carried out multi-objective optimization with ASMO using weighting function defined in Eq. (2) and the optimization results are shown in Figs. 3 and 4. To compare, we also carried out the direct optimization using SCE-UA. Figure 3 presents the default parameter, the optimal parameter given by ASMO and that given by SCE-UA. Figure 4 shows the improvements given by ASMO and SCE-UA comparing to the default parameters. From Fig. 4 we can find that all of outputs are improved nearly 10 % except soil temperature, and the improvements made by ASMO is similar with SCE-UA. The results indicated that multi-objective optimization can indeed enhance the performance of CoLM using either ASMO or SCE-UA method. The ASMO method get converged after 11 iterations, namely, the total number of model runs is 411, while the number of model runs for SCE-UA is at 1000, which is the maximum model runs set for SCE-UA. Obviously ASMO is a more efficient method compared to SCE-UA in this case.

We also used the Taylor diagram (Taylor, 2001) to compare the simulation results during calibration period and validation period (see Figs. 5 and 6). The optimization results given by SCE and ASMO are compared against the performance of default parameterization scheme. Since only 2 years observation data of the 6 output variables are available, we use the first year (2008) data as the warm-up period, use the second year (2009) data as calibration period, and then use the previous 2008 year data as the validation period. The missing records have been removed from the comparison.

As indicated in Fig. 5, the performance of optimized parameter given by both SCE and ASMO (Case C and D in the Taylor diagram) are better than default parameterization scheme (Case B) expect soil temperature. Even though, the correlation coefficients given by all the three cases are higher than 0.9, indicating that this imperfection will not cause significant inconsistency of the land surface modelling. In Fig. 6, the performance of validation period is quite similar with that in calibration period, indicating that the optimal parameters are well identified and the over-fitting problem is avoided.

HESSD
11, 6715–6751, 2014

The four energy fluxes (sensible/latent heat, upward long-wave radiation, net radiation) and soil surface temperature have very good performance. However, the performance of soil moisture seems not satisfactory. The correlation coefficient of soil moisture of Case B (default parameter) is less than 0, while with the help of SCE and ASMO optimization the correlation coefficient is only slightly larger than 0. The possible reasons might be as follows: (1) the default soil parameters of CoLM is derived from the soil texture in 17-category FAO-STATSGO soil dataset (Ji and Dai, 2010), which provide top-layer (30 cm) and bottom-layer (30–100 cm) global soil textures and has a 30 s resolution. The resolution and accuracy of this dataset may be not accurate enough in A'rou frozen/thaw station. A finer soil database, such as "The Soil Database of China for Land Surface Modeling" (Shangguan et al., 2013), or an in-situ field survey for soil texture, should be used to improve the quality of default parameterization scheme. (2) Simulating frozen/thaw processes is a challenging task in land surface modeling, and we are still lack of knowledge about the details of the physical processes. Parameter optimization can improve the model performance if the model physics are correct, but optimization is helpless if the model structure is inconsistent with the true physical processes. Although CoLM's performance of simulating frozen soil and snow cover has been evaluated in the experiment in Valdai, Russia (Dai et al., 2003), the situation of Heihe in China is very different. For an instance, in CoLM the soil depth is set to 2.86 m globally, but actually the soil depth varies in different place. Fundamentally such a simplification may not introduce significant error to the simulation of energy flux, but it definitely influence the performance of hydrological processes such as soil moisture. Otherwise, the altitude of Heihe is much higher than Valdai, and the influence of human activities is also significantly different. All these reasons can potentially influence the performance of CoLM and can not be mitigated by parameter optimization, so we should revise the model physics before parameter optimization.

HESSD
11, 6715–6751, 2014

5 Conclusions

We have carried out multi-objective parameter optimization for a land surface model, CoLM, at the Heihe river basin. Although there have been many similar works, such as multi-objective calibration of hydrological models (Gupta et al., 1998; Vrugt et al., 2003), land surface models (Gupta et al., 1999), single column land–atmosphere coupled model (Liu et al., 2005), and SVAT model (Pollacco et al., 2013), the novel contribution of this research lies in the significant reduction of model runs. In previous research, a typical multi-objective optimization needs $10^5 \sim 10^6$ or even more model runs. For large complex dynamic models which are very expensive to run, it is impractical to parameter optimization because of lack of computational resources. In this research, we managed to achieve a multi-objective optimal parameter set with only 411 model runs. The performance of the optimal parameter set is similar with the one gotten from SCE method using more than 1000 model runs. Such a result indicates that the proposed framework in this paper is able to provide optimal parameters with only hundreds of model runs. Consequently this framework is suitable to be applied to more large complex dynamic system models, such as regional land surface models, atmospheric models and even global climate models.

Appendix A: Surrogate modelling approaches

A1 Multivariate Adaptive Regression Splines (MARS)

The Multivariate Adaptive Regression Splines (MARS) model is a kind of flexible regression model of high dimensional data (Friedman, 1991). It automatically divide the high-dimensional input space into different partitions with several knots and carry out linear or nonlinear regression in each partition. It takes the form of an expansion in

HESSD
11, 6715–6751, 2014

product spline basis functions as follows:

$$y = f(\boldsymbol{x}) = a_0 + \sum_{m=1}^{M} a_m \prod_{k=1}^{K_m} [s_{k,m}(x_{v(k,m)} - t_{k,m})]_+ \quad \text{(A1)}$$

where y is the output variable and $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ is the n-dimensional input vector; a_0 is a constant, a_m are weightings of each basis functions, m is the index of basis functions and M is the total number of basis functions; in each basis function $B_m(\boldsymbol{x}) = \prod_{k=1}^{K_m} [s_{k,m}(x_{v(k,\,m)} - t_{k,m})]_+$, k is the index of knots and K_m is the total number of knots; $s_{k,m}$ take on value ± 1 and indicate the right/left sense of associated step function, $v(k,m)$ is the index of the input variable in vector $\boldsymbol{x}$, and $t_{k,m}$ indicates the knot location of the kth knot in the mth basis function.

MARS model is built in two stages: the forward pass and the backward pass. The forward pass builds an over-fitting model includes all input variables, while the backward pass removes the insensitive input variables one at a time. According to statistical learning theory, such a build-prune strategy can extract information from training data and meanwhile avoid the influence of noise (Hastie et al., 2009). Because of its pruning and fitting ability, MARS method can be used as parameter screening method (Gan et al., 2014; Li et al., 2013; Shahsavani et al., 2010), and also surrogate modeling method (Razavi et al., 2012; Song et al., 2012; Zhan et al., 2013).

A2 Gaussian Processes Regression (GPR)

Gaussian Processes Regression (GPR) (Rasmussen and Williams, 2006) is a new machine learning method based on statistical learning theory and Bayesian theory. It is suitable for high-dimensional, small-sample nonlinear regression problems. In function-space view, a Gaussian process can be completely specified by its mean function and

HESSD
11, 6715–6751, 2014

covariance function:

$$\begin{cases} m(\boldsymbol{x}) = E[f(\boldsymbol{x})] \\ k(\boldsymbol{x}, \boldsymbol{x}') = E[(f(\boldsymbol{x}) - m(\boldsymbol{x}))(f(\boldsymbol{x}') - m(\boldsymbol{x}'))] \end{cases} \tag{A2}$$

where $f(\boldsymbol{x})$ is the Gaussian process with n-dimensional input vector $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$, $m(\boldsymbol{x})$ is its mean function and $k(\boldsymbol{x}, \boldsymbol{x}')$ is its covariance function between two input vectors $\boldsymbol{x}$ and $\boldsymbol{x}'$. For short this Gaussian process can be written as $f(\boldsymbol{x}) = \mathrm{GP}(m(\boldsymbol{x}), k(\boldsymbol{x}, \boldsymbol{x}'))$.

Suppose a nonlinear regression model

$$y = f(\boldsymbol{x}) + \varepsilon \tag{A3}$$

where $\boldsymbol{x}$ is the input vector, y is the output variable, and ε is the independent identically distributed Gaussian noise term with zero mean and variance σ_n^2. Suppose $\boldsymbol{y}$ is the training outputs, $\mathbf{X}$ is the training input matrix in which each column is an input vector, $\boldsymbol{f}_*$ is the test outputs, $\mathbf{X}_*$ is the test input matrix, $K(\mathbf{X}, \mathbf{X})$, $K(\mathbf{X}, \mathbf{X}_*)$ and $K(\mathbf{X}_*, \mathbf{X}_*)$ denote covariance matrixes of all pairs of training and test inputs. We can easily write the joint distribution of training and test input and output as a joint Gaussian distribution:

$$\begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{f}_* \end{bmatrix} \sim N\left(\mathbf{0}, \begin{bmatrix} K(\mathbf{X}, \mathbf{X}) + \sigma_n^2 I & K(\mathbf{X}, \mathbf{X}_*) \\ K(\mathbf{X}_*, \mathbf{X}) & K(\mathbf{X}_*, \mathbf{X}_*) \end{bmatrix}\right) \tag{A4}$$

We can derive the mean and variance of predict output from Bayesian theory. The predictive equations are presented as follows:

$$E(\boldsymbol{f}_*) = K(\mathbf{X}_*, \mathbf{X})\left[K(\mathbf{X}, \mathbf{X}) + \sigma_n^2 I\right]^{-1} \boldsymbol{y} \tag{A5}$$

$$\mathrm{cov}(\boldsymbol{f}_*) = K(\mathbf{X}_*, \mathbf{X}_*) - K(\mathbf{X}_*, \mathbf{X})\left[K(\mathbf{X}, \mathbf{X}) + \sigma_n^2 I\right]^{-1} K(\mathbf{X}, \mathbf{X}_*) \tag{A6}$$

HESSD
11, 6715–6751, 2014

In this example, the outputs $\mathbf{y}$ is centered to zero so that the mean function is $m(\boldsymbol{x}) = 0$, while each element of covariance matrixes equals to the covariance function $k(\boldsymbol{x}, \boldsymbol{x}')$ of input pairs.

The covariance function is the crucial ingredient of Gaussian Processes Regression, as it encodes the prior knowledge about the input–output relationship. There are many kinds of covariance functions to choose and users can construct special type of cov-function depending on their prior knowledge. In this paper, we choose Martérn covariance function:

$$k(r) = \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\sqrt{2\nu} r}{l} \right)^{\nu} K_{\nu} \left(\frac{\sqrt{2\nu} r}{l} \right) \tag{A7}$$

where $r = |\boldsymbol{x} - \boldsymbol{x}'|$ is the Euclidian distance between input pair $\boldsymbol{x}$ and $\boldsymbol{x}'$, $K_{\nu}(.)$ is a modified Bessel function, ν and l positive hyper parameters, ν is the shape factor and l is the scale factor (or characteristic length). The Martérn covariance function is an isotopic cov-function that the covariance only depends on the distance between $\boldsymbol{x}$ and $\boldsymbol{x}'$. The shape scale ν controls the shape of cov-function: larger ν lead to a smoother process while small ν lead to a rougher one. If the shape scale $\nu \to \infty$ we obtain squared exponential covariance function $k(r) = \exp(-r^2/2l^2)$, which is also called radial basis function (RBF). The Martérn covariance function becomes a product of a polynomial and an exponential when ν is half-integer: $\nu = p + 1/2$. The most widely used cases are $\nu = 3/2$ and $\nu = 5/2$, as follows:

$$k_{\nu=3/2}(r) = \left(1 + \frac{\sqrt{3} r}{l} \right) \exp \left(-\frac{\sqrt{3} r}{l} \right) \tag{A8}$$

$$k_{\nu=3/2}(r) = \left(1 + \frac{\sqrt{5} r}{l} + \frac{5r^2}{3l^2} \right) \exp \left(-\frac{\sqrt{5} r}{l} \right) \tag{A9}$$

In this paper, a value of $\nu = 5/2$ was used.

HESSD
11, 6715–6751, 2014

To adaptively determine the values of hyper parameters l and σ_n, we use maximum marginal likelihood method. Because of the properties of Gaussian distribution, the log-marginal likelihood can be easily obtained as follows:

$$\log[p(\boldsymbol{y}|\mathbf{X})] = -\frac{1}{2}\boldsymbol{y}^T\left(K+\sigma_n^2 I\right)^{-1}\boldsymbol{y} - \frac{1}{2}\log\left|K+\sigma_n^2 I\right| - \frac{n}{2}\log 2\pi \tag{A10}$$

where $K = K(\mathbf{X},\mathbf{X})$. In the training process of GPR, we used SCE-UA optimization method (Duan et al., 1993) to find the best l and σ_n.

A3 Random Forests (RF)

Random Forests (RF) (Breiman, 2001) are a combination of Classification and Regression Trees (CART) (Breiman et al., 1984). Generally speaking, tree-based methods split the feature space into a set of rectangles and fit the samples in each rectangle with a class label (for classification problems) or a constant value (for regression problems). In this paper only regression tree was discussed. Suppose $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ is the n-dimensional input feature vector and y is the output response, the regression tree can be expressed as follows:

$$\hat{f}(\boldsymbol{x}) = \sum_{m=1}^{M} c_m I(\boldsymbol{x} \in R_m) \tag{A11}$$

$$I(\boldsymbol{x} \in R_m) = \begin{cases} 1, & \boldsymbol{x} \in R_m \\ 0, & \boldsymbol{x} \notin R_m \end{cases} \tag{A12}$$

where M is the total number of rectangles, m is the index of rectangle, R_m is its corresponding region, c_m is a constant value equals to the mean value of y in region R_m. To effectively and efficiently find the best binary partition, a greedy algorithm is used to determine the feature to split and the location of split point. This greedy algorithm can be very fast especially for large dataset.

HESSD
11, 6715–6751, 2014

Because of the major disadvantages of a single tree, such as over-fitting, lack of smoothness and high variance, many improved methods have been proposed, such as MARS and random forests. Random forests construct many trees using randomly selected outputs and features, and synthetic the output of all the trees to give the prediction result. Random forests only have two parameters: the total number of trees t, and the selected feature number $\hat{m}$. Constructing random forests needs following steps:

1. Bootstrap aggregating (Bagging): from total N samples $(\boldsymbol{x}_i, y_i), i = 1, 2, \ldots, N$, randomly select one point at one time with replacement, and replicate N times to get a resample set containing N points. This set is called a bootstrap replication. We need t bootstrap replications for each tree.

2. Tree construction: for each splitting of each tree, randomly select $\hat{m}$ features from the total M, and select the best fitting feature among the $\hat{m}$ to split. The $\hat{m}$ selected features should be replaced in every splitting step.

3. The prediction result of random forests is given by averaging the output of t trees.

$$\hat{f}_{\mathrm{rf}}(\boldsymbol{x}) = \sum_{j=1}^{t} \hat{f}_j(\boldsymbol{x}) \tag{A13}$$

Random forests have outstanding performance in very high dimensional problems, such as medical diagnosis and document retrieval. Such problems usually have hundreds or thousands of input variables (features), with each one only provides a little information. A single classification or regression model usually has very poor skill that only slightly better than random prediction. However, by combining many trees trained using random features, a random forest can give improved accuracy. For big-data problems that have more than 100 input features and more than one million training samples, random forests become the only choice because of its outstanding efficiency and effectiveness.

HESSD
11, 6715–6751, 2014

A4 Support Vector Machine (SVM)

Support Vector Machine (SVM) is an appealing machine learning method for classification and regression problems depending on the statistical learning theory (Vapnik, 1998, 2002). The SVM method can avoid over-fitting problem because it employs the structural risk minimization principle. It is also efficient for big-data because of its scarcity. A brief introduction to support vector regression is presented below.

The aim of SVM is to find a function $f(\boldsymbol{x})$ that can fit the output y with minimum risk given a N point training set $(\boldsymbol{x}_i, y_i), i = 1, 2, \ldots, N$. Take a simple linear regression model for example, the function $f(\boldsymbol{x})$ can be:

$$f(\boldsymbol{x}) = \boldsymbol{w}^{\mathrm{T}}\boldsymbol{x} + b \tag{A14}$$

where $\boldsymbol{w}$ is the weighting vector and $\boldsymbol{x}$ is the n-dimensional input feature vector. This function is actually determined by a small subset of training samples called support vectors (SVs).

Nonlinear problems can be transferred to linear problems by applying a nonlinear mapping from low-dimensional input space to some high-dimensional feature space:

$$f(\boldsymbol{x}) = \boldsymbol{w}^{\mathrm{T}}\phi(\boldsymbol{x}) + b \tag{A15}$$

where $\phi(\boldsymbol{x})$ is the mapping function. The inner product of mapping function is called Kernel Function: $K(\boldsymbol{x}, \boldsymbol{x}') = \phi(\boldsymbol{x})^{\mathrm{T}}\phi(\boldsymbol{x}')$ and this method is called Kernel method. The commonly used kernel functions are: linear kernel function, polynomial, sigmoid and radial basis function (RBF). In this paper we use RBF kernel:

$$K(\boldsymbol{x}, \boldsymbol{x}') = \exp(-\gamma|\boldsymbol{x} - \boldsymbol{x}'|^2) \tag{A16}$$

where $|\boldsymbol{x} - \boldsymbol{x}'|$ is the Euclidian distance between $\boldsymbol{x}$ and $\boldsymbol{x}'$, γ is a user defined parameter that controls the smoothness of $f(\boldsymbol{x})$.

HESSD
11, 6715–6751, 2014

To qualify the "risk" of function $f(\boldsymbol{x})$, a loss function is defined as follows:

$$|y - f(\boldsymbol{x})|_{\varepsilon} = \begin{cases} 0, & \text{if } |y - f(\boldsymbol{x})| \le \varepsilon \\ |y - f(\boldsymbol{x})| - \varepsilon, & \text{otherwise} \end{cases} \tag{A17}$$

The loss function means regression errors less than tolerance ε are not penalized. The penalty-free zone is also called ε-tube or ε-boundary. As explained in statistical learning theory (Vapnik, 1998), the innovative loss function is the key point that SVM can balance empirical risk (risk of large error in the training set) and structure risk (risk of an over-complex model, or over-fitting). The problem of simultaneously minimizing both empirical risk (represented by regression error) and structure risk (represented by the width of ε-tube) can be written as a quadratic optimization problem:

$$\begin{aligned} &\min_{\boldsymbol{w},b,\xi,\xi^*} \frac{1}{2}\boldsymbol{w}^{\mathrm{T}}\boldsymbol{w} + C\sum_{i=1}^{n}\xi_i + C\sum_{i=1}^{n}\xi_i^* \\ &\text{subject to } \boldsymbol{w}^{\mathrm{T}}\phi(\boldsymbol{x}_i) + b - y_i \le \varepsilon + \xi_i \\ &\qquad y_i - \boldsymbol{w}^{\mathrm{T}}\phi(\boldsymbol{x}_i) - b \le \varepsilon + \xi_i^* \\ &\qquad \xi_i, \xi_i^* \ge 0, i = 1, 2, \ldots, n \end{aligned} \tag{A18}$$

The problem can be transferred to the dual problem:

$$\begin{aligned} &\min_{\boldsymbol{w},b,\xi,\xi^*} \frac{1}{2}(\boldsymbol{\alpha} - \boldsymbol{\alpha}^*)^{\mathrm{T}}\mathbf{K}(\boldsymbol{\alpha} - \boldsymbol{\alpha}^*) + \varepsilon\sum_{i=1}^{n}(\alpha_i + \alpha_i^*) + \sum_{i=1}^{n} y_i(\alpha_i - \alpha_i^*) \\ &\text{subject to } \boldsymbol{e}^{\mathrm{T}}(\boldsymbol{\alpha} - \boldsymbol{\alpha}^*) = 0 \\ &\qquad y_i - \boldsymbol{w}^{\mathrm{T}}\phi(\boldsymbol{x}_i) - b \le \varepsilon + \xi_i^* \\ &\qquad 0 \le \alpha_i, \alpha_i^* \le C, i = 1, 2, \ldots, n \end{aligned} \tag{A19}$$

where $\mathbf{K}$ is the kernel function matrix with $K_{ij} = K(\boldsymbol{x}_i, \boldsymbol{x}_j)$. Solving the dual problem and we can get the predictive function:

$$f(\boldsymbol{x}) = \sum_{i=1}^{n}\left(-\alpha_i + \alpha_i^*\right)K(\boldsymbol{x}_i, \boldsymbol{x}) + b \tag{A20}$$

HESSD
11, 6715–6751, 2014

where the vectors ($\alpha^* - \alpha$) are the support vectors (SVs).

A5 Artificial Neural Network (ANN)

Artificial Neural Network (ANN) (REF) is time-hornored marchine learning method comparing to the former four. It is a data-driven process that can solve complex nonlinear relationships between input and outpur data. A nerual network is constructed by many interconnected neurons. Each neuron can be mathematically described as a linear weighing function and a nonlinear activation function:

$$I_i = \sum_{j=1}^{n} w_{ij} x_j \tag{A21}$$

$$f_i(I) = \frac{1}{1 + \exp(-I_i)} \tag{A22}$$

where x_j is the jth input variable, w_{ij} is the weight and I_i is the weighted sum of the ith neuron. The output of the ith neuron $f_i(I)$ is given by the nonlinear activation function of the weighted sum input. Here we use Sigmoid function.

Minsky and Papert (1969) shows that single layer neural network can only solve linear problem. Cybenko (1989) extended ANN to multiple layer and demostrated that multi-layer ANN can infinitely approximate any nonlinear function (the universal approximation theorem). The training procedure of ANN is optimizing the value of weights. There are many training methods for ANN and we use the Levenberg–Marquardt (LM) (Marquardt, 1963) algorithm, a modification of the classic Newton algorithm provided in matlab ANN toolbox.

Acknowledgements. This research is supported by Natural Science Foundation of China (Grant No. 41075075, No. 41375139 and No. 51309011), Chinese Ministry of Science and Technology 973 Research Program (No. 2010CB428402) and Beijing Normal University Research Fund (No.2013YB47). Special thanks are due to "Environmental & Ecological Science Data Center for West China, National Natural Science Foundation of China" (http:

HESSD
11, 6715–6751, 2014

//westdc.westgis.ac.cn) for providing the meteorological forcing data, to the group of Shaomin Liu at State Key Laboratory of Remote Sensing Science, School of Geography and Remote Sensing Science of Beijing Normal University for providing the surface flux validation data.

References

Bastidas, L. A., Gupta, H. V., Sorooshian, S., Shuttleworth, W. J., and Yang, Z. L.: Sensitivity analysis of a land surface scheme using multicriteria methods, J. Geophys. Res.-Atmos., 104, 19481–19490, doi:10.1029/1999JD900155, 1999.

Bonan, G. B.: A Land Surface Model (LSM version 1.0) for Ecological, Hydrological, and Atmospheric Studies: technical Description and User's Guide, NCAR, Boulder, CO, USA, doi:10.5065/D6DF6P5X, 1996.

Boyle, D. P.: Multicriteria Calibration of Hydrological Models, Ph.D. thesis, Department of Hydrology and Water Resources, University of Arizona, Tucson, USA, 2000.

Boyle, D. P., Gupta, H. V., and Sorooshian, S.: Toward improved calibration of hydrologic models: combining the strengths of manual and automatic methods, Water Resour. Res., 36, 3663–3674, doi:10.1029/2000WR900207, 2000.

Breiman, L.: Random forests, Mach. Learn., 45, 5–32, 2001.

Breiman, L., Friedman, J., Stone, C. J., and Olshen, R. A.: Classification and Regression Trees, Chapman and Hall/CRC, Bota Raton, Florida, USA, 1984.

Cybenko, G.: Approximation by superpositions of a sigmoidal function, Math. Control Signal., 2, 303–314, doi:10.1007/BF02551274, 1989.

Dai, Y. and Zeng, Q.: A Land Surface Model (IAP94) for climate studies, Part I: formulation and validation in off-line experiments, Adv. Atmos. Sci., 14, 433, doi:10.1007/s00376-997-0063-4, 1997.

Dai, Y., Xue, F., and Zeng, Q.: A Land Surface Model (IAP94) for climate studies, Part II: implementation and preliminary results of coupled model with IAP GCM, Adv. Atmos. Sci., 15, 47–62, 1998.

Dai, Y. J., Zeng, X. B., Dickinson, R. E., Baker, I., Bonan, G. B., Bosilovich, M. G., Denning, A. S., Dirmeyer, P. A., Houser, P. R., Niu, G. Y., Oleson, K. W., Schlosser, C. A., and Yang, Z. L.: The Common Land Model, B. Am. Meteorol. Soc., 84, 1013–1023, 2003.

HESSD
11, 6715–6751, 2014

Dai, Y. J., Dickinson, R. E., and Wang, Y. P.: A two-big-leaf model for canopy temperature, photosynthesis, and stomatal conductance, J. Climate, 17, 2281–2299, 2004.
Dickinson, R. E., Henderson-Sellers, A., and Kennedy, P. J.: Biosphere–Atmosphere Transfer Scheme (BATS) Version 1e as Coupled to the NCAR Community Climate Model, NCAR, Boulder, CO, USA, 1993.
Duan, Q. Y., Sorooshian, S., and Gupta, V. K.: Effective and efficient global optimization for conceptual rainfall–runoff models, Water Resour. Res., 28, 1015–1031, doi:10.1029/91WR02985, 1992.
Duan, Q. Y., Gupta, V. K., and Sorooshian, S.: Shuffled complex evolution approach for effective and efficient global minimization, J. Optimiz. Theory App., 76, 501–521, doi:10.1007/BF00939380, 1993.
Duan, Q. Y., Sorooshian, S., and Gupta, V. K.: Optimal use of the SCE-UA global optimization method for calibrating watershed models, J. Hydrol., 158, 265–284, doi:10.1016/0022-1694(94)90057-4, 1994.
Friedman, J. H.: Multivariate adaptive regression splines, Ann. Stat., 19, 1–14, 1991.
Gan, Y., Duan, Q., Gong, W., Tong, C., Sun, Y., Chu, W., Ye, A., Miao, C., and Di, Z.: A comprehensive evaluation of various sensitivity analysis methods: a case study with a hydrological model, Environ. Modell. Softw., 51, 269–285, doi:10.1016/j.envsoft.2013.09.031, 2014.
Gupta, H. V., Sorooshian, S., and Yapo, P. O.: Toward improved calibration of hydrologic models: multiple and noncommensurable measures of information, Water Resour. Res., 34, 751–763, doi:10.1029/97WR03495, 1998.
Gupta, H. V., Bastidas, L. A., Sorooshian, S., Shuttleworth, W. J., and Yang, Z. L.: Parameter estimation of a land surface scheme using multicriteria methods, J. Geophys. Res., 104, 19491–19503, doi:10.1029/1999JD900154, 1999.
Hastie, T., Tibshirani, R., and Friedman, J.: The Elements of Statistical Learning, 2nd edn., Springer, New York, USA, 2009.
Henderson-Sellers, A., Mcguffie, K., and Pitman, A. J.: The Project for Intercomparison of Land-surface Parametrization Schemes (PILPS): 1992 to 1995, Clim. Dynam., 12, 849–859, doi:10.1007/s003820050147, 1996.
Hu, Z. Y., Ma, M. G., Jin, R., Wang, W. Z., Huang, G. H., Zhang, Z. H., and Tan, J. L.: WATER: Dataset of Automatic Meteorological Observations at the A'rou Freeze/Thaw Observation Station, Cold and Arid Regions Environmental and Engineering Research Institute, Chinese

HESSD
11, 6715–6751, 2014

Academy of Sciences, Lanzhou, China, doi:10.3972/water973.0279.db, available at: http://westdc.westgis.ac.cn/water/50a8fe84-deeb-49b0-b8d8-8f784bc27a89, last access: 11 June 2014, 2003.

Ji, D. and Dai, Y.: The Common Land Model (CoLM) Technical Guide, GCESS, Beijing Normal University, Beijing, China, 2010.

Leplastrier, M., Pitman, A. J., Gupta, H., and Xia, Y.: Exploring the relationship between complexity and performance in a land surface model using the multicriteria method, J. Geophys. Res., 107, 4443D20, doi:10.1029/2001JD000931, 2002.

Li, J., Duan, Q. Y., Gong, W., Ye, A., Dai, Y., Miao, C., Di, Z., Tong, C., and Sun, Y.: Assessing parameter importance of the Common Land Model based on qualitative and quantitative sensitivity analysis, Hydrol. Earth Syst. Sci., 17, 3279–3293, doi:10.5194/hess-17-3279-2013, 2013.

Liang, X., Wood, E. F., Lettenmaier, D. P., Lohmann, D., Boone, A., Chang, S., Chen, F., Dai, Y., Desborough, C., Dickinson, R. E., Duan, Q., Ek, M., Gusev, Y. M., Habets, F., Irannejad, P., Koster, R., Mitchell, K. E., Nasonova, O. N., Noilhan, J., Schaake, J., Schlosser, A., Shao, Y., Shmakin, A. B., Verseghy, D., Warrach, K., Wetzel, P., Xue, Y., Yang, Z., and Zeng, Q.: The Project for Intercomparison of Land-surface Parameterization Schemes (PILPS) phase 2(c) Red-Arkansas River basin experiment: 2. Spatial and temporal analysis of energy fluxes, Global Planet. Change, 19, 137–159, doi:10.1016/S0921-8181(98)00045-9, 1998.

Liu, Y. Q., Bastidas, L. A., Gupta, H. V., and Sorooshian, S.: Impacts of a parameterization deficiency on offline and coupled land surface model simulations, J. Hydrometeorol., 4, 901–914, doi:10.1175/1525-7541(2003)004<0901:IOAPDO>2.0.CO;2, 2003.

Liu, Y. Q., Gupta, H. V., Sorooshian, S., Bastidas, L. A., and Shuttleworth, W. J.: Exploring parameter sensitivities of the land surface using a locally coupled land–atmosphere model, J. Geophys. Res., 109, D21101D21, doi:10.1029/2004JD004730, 2004.

Liu, Y. Q., Gupta, H. V., Sorooshian, S., Bastidas, L. A., and Shuttleworth, W. J.: Constraining land surface and atmospheric parameters of a locally coupled model using observational data, J. Hydrometeorol., 6, 156–172, doi:10.1175/JHM407.1, 2005.

Lohmann, D., Lettenmaier, D. P., Liang, X., Wood, E. F., Boone, A., Chang, S., Chen, F., Dai, Y., Desborough, C., Dickinson, R. E., Duan, Q., Ek, M., Gusev, Y. M., Habets, F., Irannejad, P., Koster, R., Mitchell, K. E., Nasonova, O. N., Noilhan, J., Schaake, J., Schlosser, A., Shao, Y., Shmakin, A. B., Verseghy, D., Warrach, K., Wetzel, P., Xue, Y., Yang, Z., and Zeng, Q.: The Project for Intercomparison of Land-surface Parameterization Schemes (PILPS) phase 2(c)

HESSD
11, 6715–6751, 2014

Red–Arkansas River basin experiment: 3. Spatial and temporal analysis of water fluxes, Global Planet. Change, 19, 161–179, doi:10.1016/S0921-8181(98)00046-0, 1998.
Marquardt, D. W.: An algorithm for least-squares estimation of nonlinear parameters, J. Soc. Ind. Appl. Math., 11, 431–441, 1963.
Mckay, M. D., Beckman, R. J., and Conover, W. J.: A comparison of three methods for selecting values of input variables in the analysis of output from a computer code, Technometrics, 21, 239–245, doi:10.2307/1268522, 1979.
Minsky, M. and Papert, S. A.: Perceptrons: an Introduction to Computational Geometry, MIT Press, Cambridge, Massachusetts, USA, 1969.
Pollacco, J. A. P., Mohanty, B. P., and Efstratiadis, A.: Weighted objective function selector algorithm for parameter estimation of SVAT models with remote sensing data, Water Resour. Res., 49, 6959–6978, doi:10.1002/wrcr.20554, 2013.
Rasmussen, C. E. and Williams, C. K. I.: Gaussian Processes for Machine Learning, MIT Press, Massachusetts, USA, 2006.
Razavi, S., Tolson, B. A., and Burn, D. H.: Review of surrogate modeling in water resources, Water Resour. Res., 48, W7401, doi:10.1029/2011WR011527, 2012.
Shahsavani, D., Tarantola, S., and Ratto, M.: Evaluation of MARS modeling technique for sensitivity analysis of model output, Procedia – Social and Behavioral Sciences, 2, 7737–7738, doi:10.1016/j.sbspro.2010.05.204, 2010.
Shangguan, W., Dai, Y., Liu, B., Zhu, A., Duan, Q., Wu, L., Ji, D., Ye, A., Yuan, H., Zhang, Q., Chen, D., Chen, M., Chu, J., Dou, Y., Guo, J., Li, H., Li, J., Liang, L., Liang, X., Liu, H., Liu, S., Miao, C., and Zhang, Y.: A China data set of soil properties for land surface modeling, Journal of Advances in Modeling Earth Systems, 5, 212–224, doi:10.1002/jame.20026, 2013.
Song, X., Zhan, C., and Xia, J.: Integration of a statistical emulator approach with the SCE-UA method for parameter optimization of a hydrological model, Chinese Sci. Bull., 57, 3397–3403, 2012.
Taylor, K. E.: Summarizing multiple aspects of model performance in a single diagram, J. Geophys. Res.-Atmos., 106, 7183–7192, doi:10.1029/2000JD900719, 2001.
Vapnik, V. N.: Statistical Learning Theory, John Wiley & Sons, NewYork, USA, 1998.
Vapnik, V. N.: The Nature of Statistical Learning Theory, 2nd edn., Springer, New York, 2002.
Vrugt, J. A., Gupta, H. V., Bouten, W., and Sorooshian, S.: A shuffled complex evolution metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters, Water Resour. Res., 39, 1201, doi:10.1029/2002WR001642, 2003a.

HESSD
11, 6715–6751, 2014

Vrugt, J. A., Gupta, H. V., Bastidas, L. A., Bouten, W., and Sorooshian, S.: Effective and efficient algorithm for multiobjective optimization of hydrologic models, Water Resour. Res., 39, 1214, doi:10.1029/2002WR001746, 2003b.

Wang, C., Duan, Q., Gong, W., Ye, A., Di, Z., and Miao, C.: The effect of initial sampling designs and sample sizes on adaptive surrogate modeling based optimization, Environ. Modell. Softw., in review, 2014.

Wang, Q. J.: The genetic algorithm and its application to calibrating conceptual rainfall–runoff models, Water Resour. Res., 27, 2467–2471, 1991.

Wood, E. F., Lettenmaier, D. P., Liang, X., Lohmann, D., Boone, A., Chang, S., Chen, F., Dai, Y., Dickinson, R. E., Duan, Q., Ek, M., Gusev, Y. M., Habets, F., Irannejad, P., Koster, R., Mitchel, K. E., Nasonova, O. N., Noilhan, J., Schaake, J., Schlosser, A., Shao, Y., Shmakin, A. B., Verseghy, D., Warrach, K., Wetzel, P., Xue, Y., Yang, Z., and Zeng, Q.: The Project for Intercomparison of Land-surface Parameterization Schemes (PILPS) phase 2(c) Red-Arkansas River basin experiment: 1. Experiment description and summary intercomparisons, Global Planet. Change, 19, 115–135, doi:10.1016/S0921-8181(98)00044-7, 1998.

Xia, Y., Pittman, A. J., Gupta, H. V., Leplastrier, M., Henderson-Sellers, A., and Bastidas, L. A.: Calibrating a land surface model of varying complexity using multicriteria methods and the Cabauw dataset, J. Hydrometeorol., 3, 181–194, doi:10.1175/1525-7541(2002)003<0181:CALSMO>2.0.CO;2, 2002.

Yapo, P. O., Gupta, H. V., and Sorooshian, S.: Multi-objective global optimization for hydrologic models, J. Hydrol., 204, 83–97, doi:10.1016/S0022-1694(97)00107-8, 1998.

Zhan, C., Song, X., Xia, J., and Tong, C.: An efficient integrated approach for global sensitivity analysis of hydrological model parameters, Environ. Modell. Softw., 41, 39–52, 2013.

HESSD
11, 6715–6751, 2014

Table 1. Adjustable parameters and their categories, meanings and ranges.

Num	Para	Units	Category	Physical meaning	Feasible range
P1	dewmx		canopy	maximum dew ponding of leaf area	[0.05, 0.15]
P2	hksati	$mm\,s^{-1}$	soil	maximum hydraulic conductivity	[0.001, 1]
P3	porsl	–	soil	porosity	[0.25, 0.75]
P4	phi0	mm	soil	minimum soil suction	[50, 500]
P5	wtfact		soil	fraction of shallow groundwater area	[0.15, 0.45]
P6	bsw	–	soil	Clapp and Hornberger "b" parameter	[2.5, 7.5]
P7	wimp		soil	water impermeable if porosity less than wimp	[0.01, 0.1]
P8	zlnd	m	soil	roughness length for soil surface	[0.005, 0.015]
P9	pondmx	mm	soil	maximum ponding depth for soil surface	[5, 15]
P10	csoilc	–	soil	drag coefficient for soil under canopy	[0.002, 0.006]
P11	zsno	m	snow	roughness length for snow	[0.0012, 0.0036]
P12	capr		soil	tuning factor of soil surface temperature	[0.17, 0.51]
P13	cnfac		canopy	Crank Nicholson factor	[0.25, 0.5]
P14	slti		canopy	slope of low temperature inhibition function	[0.1, 0.3]
P15	hlti		canopy	1/2 point of low temperature inhibition function	[278, 288]
P16	shti		canopy	slope of high temperature inhibition function	[0.15, 0.45]
P17	sqrtdi	$m^{-0.5}$	canopy	the inverse of square root of leaf dimension	[2.5, 7.5]
P18	effcon	$mol\,CO^2\,mol^{-1}$ quanta	canopy	quantum efficiency of vegetation photosynthesis	[0.035, 0.35]
P19	vmax25	$mol\,CO^2\,m^{-2}\,s^{-1}$	canopy	maximum carboxylation rate at 25 °C	$[10^{-6}, 200^{-6}]$
P20	hhti		canopy	1/2 point of high temperature inhibition function	[305, 315]
P21	trda		canopy	temperature coefficient of conductance-photosynthesis model	[0.65,1.95]
P22	trdm		canopy	temperature coefficient of conductance-photosynthesis model	[300, 350]
P23	trop		canopy	temperature coefficient of conductance-photosynthesis model	[250, 300]
P24	gradm		canopy	slope of conductance-photosynthesis model	[4, 9]
P25	binter		canopy	intercept of conductance-photosynthesis model	[0.01, 0.04]
P26	extkn		canopy	coefficient of leaf nitrogen allocation	[0.5, 0.75]
P27	chil		canopy	leaf angle distribution factor	[−0.3, 0.1]
P28	ref(1,1)		canopy	VIS reflectance of living leaf	[0.07, 0.105]
P29	ref(1,2)		canopy	VIS reflectance of dead leaf	[0.16, 0.36]
P30	ref(2,1)		canopy	NIR reflectance of living leaf	[0.35, 0.58]
P31	ref(2,2)		canopy	NIR reflectance of dead leaf	[0.39, 0.58]
P32	tran(1,1)		canopy	VIS transmittance of living leaf	[0.04, 0.08]
P33	tran(1,2)		canopy	VIS transmittance of dead leaf	[0.1, 0.3]
P34	tran(2,1)		canopy	NIR transmittance of living leaf	[0.1, 0.3]
P35	tran(2,2)		canopy	NIR transmittance of dead leaf	[0.3, 0.5]
P36	z0 m	m	canopy	aerodynamic roughness length	[0.05, 0.3]
P37	ssi		snow	irreducible water saturation of snow	[0.03, 0.04]
P38	smpmax	mm	soil	wilting point potential	[−2.e5, −1.e5]
P39	smpmin	mm	soil	restriction for min of soil potential	[−1.e8, −9.e7]
P40	trsmx0	$mm\,s^{-1}$	canopy	maximum transpiration for vegetation	[1.e-4, 1.e-2]

HESSD
11, 6715–6751, 2014

Table 2. Screened parameters of CoLM in A'rou Station (Li et.al., 2013).

Output variables (fluxes)	Screened parameters
Sensible Heat	P2, P4, P6, P30, P34, P36
Latent Heat	P2, P3, P4, P6, P18, P30, P36, P38
Upward Longwave Radiation	P6, P17, P36
Net radiation	P6, P17, P30, P34, P36
Soil Temperature	P3, P6, P36
Soil Moisture	P3, P6

HESSD
11, 6715–6751, 2014

Table 3. The NRMSE between simulated and observed outputs after single objective optimization.

	Sensible heat	Latent heat	Upward longwave radiation	Net radiation	Soil temperature	Soil moisture
Default	0.8586	0.5833	0.059	0.2357	0.0096	0.4527
SCE	0.745	0.4921	0.038	0.1963	0.0073	0.3900
Optimized	(1492)	(1354)	(458)	(982)	(473)	(210)
LH50	0.7672	0.5255	0.0377	0.1913	0.008	0.4222
LH100	0.7841	0.5785	0.0372	0.1908	0.0077	0.413
LH200	0.7821	0.5885	0.0374	0.1928	0.0069	0.3947
LH400	0.7798	0.5627	0.0374	0.1928	0.007	0.3971
LH800	0.7683	0.5024	0.0377	0.1909	0.0068	0.3956
LH1200	0.7760	0.5150	0.0374	0.1919	0.0068	0.3962
LH2000	0.7705	0.5048	0.0375	0.1912	0.0070	0.3946

HESSD
11, 6715–6751, 2014

Table 4. Weights assigned to each output variables.

Flux name	Label	Unit	RMSE	NRMSE	Weights
Sensible heat	fsena	$W\,m^{-2}$	49.14	0.8586	0.3905
Latent heat	lfevpa	$W\,m^{-2}$	43.59	0.5833	0.2653
Upward longwave radiation	orlg	$W\,m^{-2}$	19.43	0.0590	0.0268
Net radiation	sabvg	$W\,m^{-2}$	42.78	0.2357	0.1072
Soil temperature	tss	K	2.66	0.0096	0.0044
Soil moisture	wliq	$kg\,m^{-2}$	21.14	0.4527	0.2059

HESSD
11, 6715–6751, 2014

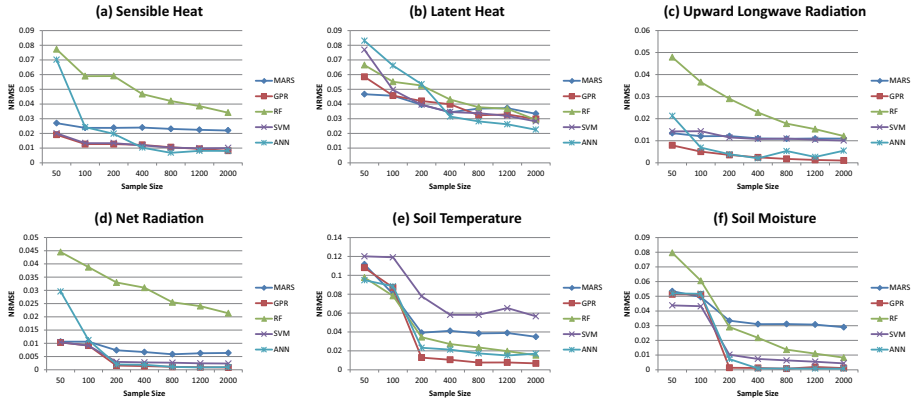

Figure 1. Inter-comparison of 5 surrogate modelling methods.

HESSD
11, 6715–6751, 2014

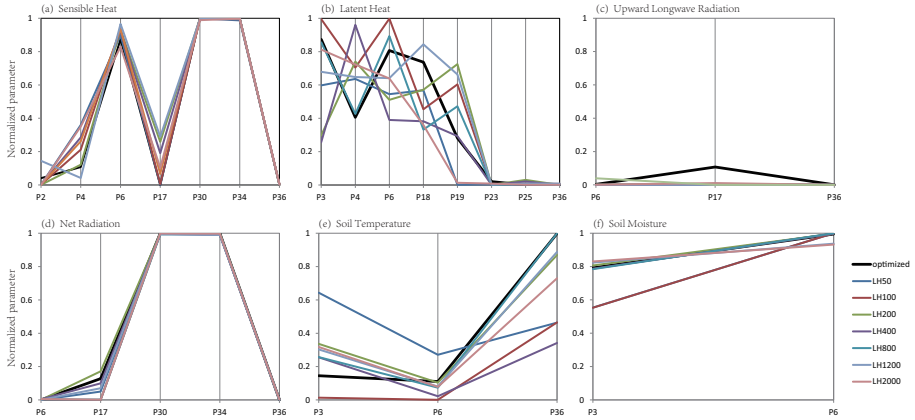

Figure 2. Single-objective optimization result: optimal parameters.

HESSD
11, 6715–6751, 2014

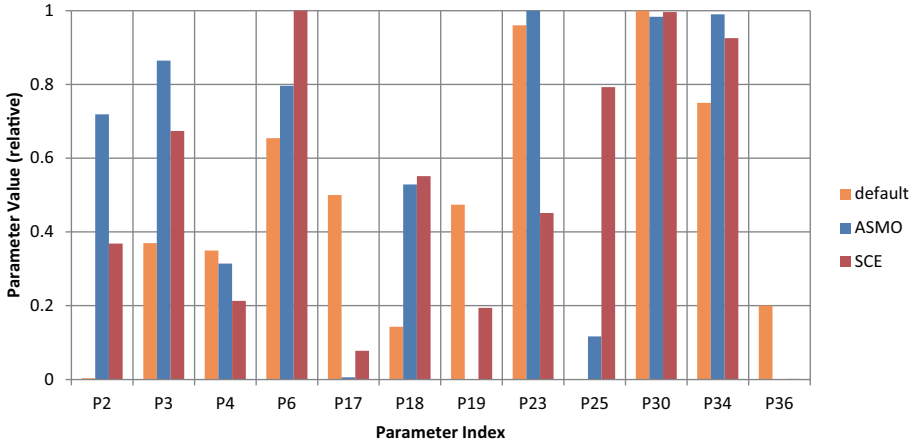

Figure 3. Optimal value of CoLM given by multi-objective optimization (comparing default parameter, optimal parameter given by ASMO and SCE-UA).

HESSD
11, 6715–6751, 2014

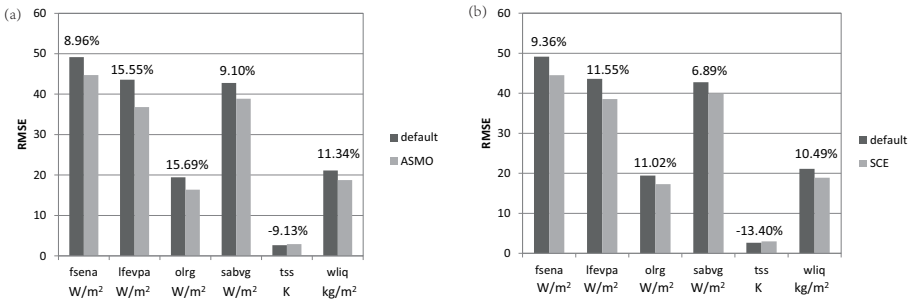

Figure 4. Comparing the improvements given by ASMO and SCE.

HESSD
11, 6715–6751, 2014

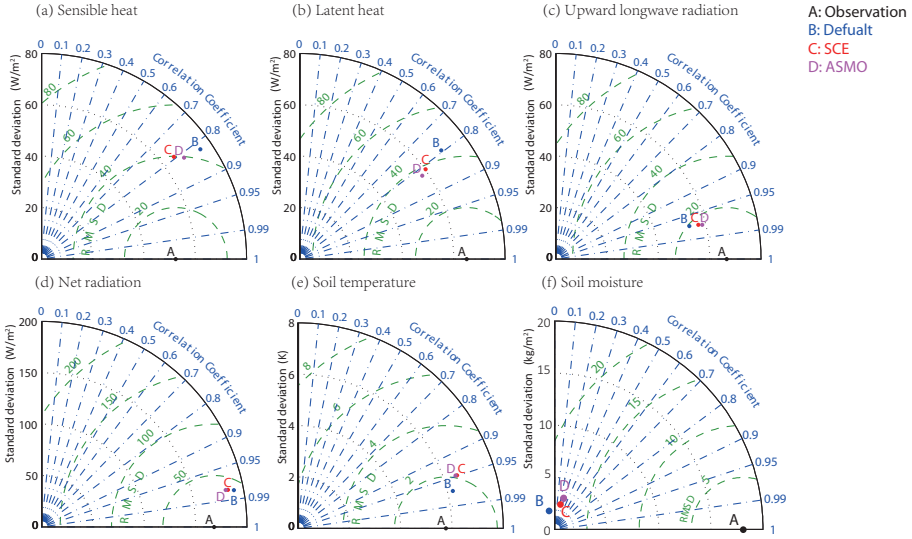

Figure 5. Taylor diagram of simulated fluxes during calibration period (1 January 2009 to 31 December 2009).

HESSD
11, 6715–6751, 2014

Multi-objective optimization of CoLM's parameters

W. Gong et al.

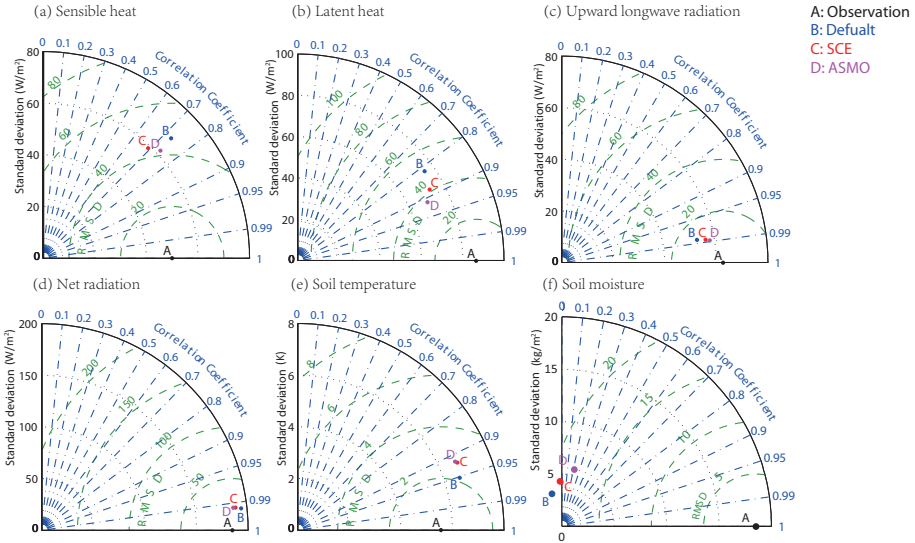

Figure 6. Taylor diagram of simulated fluxes during validation period (here we use warm-up period as validation period, 1 January 2008 to 31 December 2008).

HESSD
11, 6715–6751, 2014