

# Multi-objective Parameter Optimization of Common Land Model Using Adaptive Surrogate Modelling

Wei Gong<sup>1</sup>, Qingyun Duan<sup>1</sup>, Jianduo Li<sup>1</sup>, Chen Wang<sup>1</sup>, Zhenhua Di<sup>1</sup>, Yongjiu Dai<sup>1</sup>,  
Aizhong Ye<sup>1</sup>, Chiyuan Miao<sup>1</sup>

<sup>1</sup>College of Global Change and Earth System Science (GCESS), Beijing Normal University, Beijing, 100875, and Joint Center for Global Change Studies, Beijing, 100875, China

([gongwei2012@bnu.edu.cn](mailto:gongwei2012@bnu.edu.cn), [qyduan@bnu.edu.cn](mailto:qyduan@bnu.edu.cn), [lijianduo@mail.bnu.edu.cn](mailto:lijianduo@mail.bnu.edu.cn),  
[wangchen@mail.bnu.edu.cn](mailto:wangchen@mail.bnu.edu.cn), [zhdi@bnu.edu.cn](mailto:zhdi@bnu.edu.cn), [yongjiudai@bnu.edu.cn](mailto:yongjiudai@bnu.edu.cn), [azye@bnu.edu.cn](mailto:azye@bnu.edu.cn),  
[miaocy@vip.sina.com](mailto:miaocy@vip.sina.com) )

## 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 to 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-large~~ number of model runs (typically  $10^5 \sim 10^6$ ). It makes parameter optimization computationally prohibitive. An uncertainty ~~qualification-quantification~~ framework was developed to meet the aforementioned challenges, which include the following steps: (1) useing parameter screening to reduce the number of adjustable parameters; (2) useing surrogate models to emulate the responses of dynamic models to the variation of adjustable parameters; (3) useing an adaptive strategy to ~~promote-improve~~ the efficiency of surrogate modeling

based optimization; (4) using 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 application of a case study of a land surface model – the 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. The result indicate that this framework can effeicientively achieve optimal parameters in a more effective way. Moreover, this result implies the possibility of calibrating other large complex dynamic models, such as regional-scale land surface models, atmospheric models and climate models. In the future work we are going to continue the development of the framework and improve its scalability, robustness and flexibility for other expensive models.~~

#### Keywords:

Land surface model; multi-objective optimization; parameter calibration; surrogate modeling; statistical emulator; adaptive sampling;

#### 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 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 (e.g., [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; Duan *et al.*, 1993; Duan *et al.*, 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 as 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 (e.g., [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; Liu *et al.*, 2004; 2005]. To take a multi-objective optimization approach 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.

Surrogate based optimization is one of the most commonly used approaches to optimizing large complex dynamic models. Several books and literature reviews have described the advances of surrogate based optimization in recent years [e.g., Jones, 2001; Ong *et al.*, 2005; Jin, 2011; Koziel and Leifsson, 2013; and Wang *et al.*, 2014]. Surrogate based optimization has been applied to economics, robotics, chemistry, physics, civil and environmental engineering, computational fluid dynamics, aerospace

designs, et al [Gorissen, 2010]. On the development of surrogate based optimization, Jones et al. [1998] proposed EGO (Effective Global Optimizer) for expensive models using ‘DACE stochastic process model’, namely Kriging interpolation method, as surrogate model. Castelletti et al. [2010] developed a multi-objective optimization method for water quality management using radial basis function, inverse distance weighted and n-dimensional linear interpolator as surrogates. Loshchilov et al. [2010] investigated the use of ranked-based Support Vector Machine and demonstrated that for surrogate based optimization capturing the relative value of the objective functions is more important than reducing the absolute fitting error. Pilá and Neruda [2013] developed a surrogate model selector for multi-objective surrogate-assisted optimization. In hydrology and water resources, Razavi et al. [2012] has summarized recent applications, advantages, and existing problems. Wang et al. [2014] evaluated the influence of initial sampling and adaptive sampling methods for surrogate-assisted optimization of a simple hydrological model, SAC-SMA model. Song et al. [2012] optimized the parameter of a distributed hydrological model-DTVGM model’s parameter with SCE-UA algorithm using MARS method [Friedman, 1991] as surrogate.

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 ~~of~~ the dynamic system models. In this paper, we will illustrate this parametric uncertainty quantification framework ~~with~~ the Common Land Model (CoLM), a widely used, physically based land surface model developed by Yongjiu Dai and colleagues [Dai *et al.*, 2003; Dai *et al.*, 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: section 2 introduces the basic information of CoLM, the study area and dataset, the adjustable parameters and the output variables to be analyzed; section 3 presents an inter-comparison of 5 surrogate modeling methods, and discusses how many model runs ~~will~~would be sufficient to build a surrogate model for optimization; section 4 carries out single and multiple objective optimization using an adaptive surrogate model based optimization strategy (ASMO); section 5 provides the discussion and conclusions.

## 2. Experiment setup

### Model and Parameters

Common Land Model (CoLM) proposed by Yongjiu Dai and colleagues [Dai *et al.*, 2003; Dai *et al.*, 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**.

[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 Root Mean Squared Error is used as the objective function in our analysis:

$$NRMSE_i = \frac{\sqrt{\sum_{j=1}^N (y_{i,j}^{sim} - y_{i,j}^{obs})^2}}{\sum_{j=1}^N y_{i,j}^{obs}} \quad (1)$$

where  $i$  is the index of output variables,  $j$  is the index of time step,  $N$  is the total number

of observations,  $y_{i,j}^{sim}$  and  $y_{i,j}^{obs}$  are the simulated and observed values, respectively.

Objective functions represent the performance of model simulation and a smaller objective function means better performance.

### 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<sup>2</sup> and its altitude varies from sea level to 5500m. 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, 38°03'N, altitude is 3032.8m 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 10cm, 20cm, 40cm and 80cm. In CoLM, the soil is divided into 10 layers and the simulated soil temperature and soil moisture are 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 minutes and the simulation outputs are averaged to 3 hours 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-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 [up to](#) 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~~[can](#) 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 [pitfalls](#) 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, [respectively](#). The inter-comparison results are shown in **Figure 1** and **Figure 2**, 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). **Figure 1** shows the error of [the training set, namely the NRMSE between the outputs predicted by the surrogate model and the outputs of the training samples](#), and **figure 2** shows the NRMSE of the testing set. [Since every sample set of each size was independently generated, we use the 2000 points set to test](#)

50, 100, 200, 400, 800 and 1200 points set, and use the 1200 one to test the 2000 one. 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 obtained by Li [2012] and Li et al. [2013] (see The list of screened sensitive parameters is are shown in Table 2).

[Table 2]

As shown in Figure 1, for some cases, such as upward longwave radiation, the fitting ability of the training set does not significantly change significantly with sample size, but for soil moisture, larger sample size leads to better fitted surrogate models. Such phenomenon indicated that the variousness specific features of the response surfaces have significant influence on the fitting ability, and good surrogate models must have the ability to adaptively identify and mitigate the various properties to those features. As shown in Figure 1, GPR has the best fitting ability for almost every cases except soil temperature. Because as described in Appendix 2, the hyper-parameters used by GPR can be adaptively be adaptively determined using the maximum marginal likelihood method.

Figure 2 shows the NRMSE of the testing sets, indicating the risk of over-fitting. In Figure 2 we can get note more remarkable findings as follows: Figure 1 indicated that: (1) The error of a surrogate model decreases as the sample size increases. The marginal benefits of using additional samples becomes less or even negligible if the sample size is larger than 400; 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. (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 acceptable, while for other variables (e.g., soil temperature), SVM's performance is not satisfactory; (3) The convergence speed for the 6 output variables is are different. For net radiation, soil temperature and soil moisture, the fitting error decreases to nearly zero if the sampling points are more than 200; while for sensible heat, latent heat and upward long-wave radiation, the marginal benefit of adding more points is still



significant ~~for~~ beyond more than 200 sample points. 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 consistently give the best performance for all the 6 output variables, we choose GPR in the following multi-objective optimization analysis presented later.

[Figure 1]

[Figure 2]

## 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; Duan *et al.*, 1993; Duan *et al.*, 1994] is used to find the optima of the surrogate models. 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 model runs.

The converged optimal parameters from the given by single-objective optimization runs are shown in Figure 32. 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.

[Figure 32]

[Table 3]

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~~ obtained 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 sizes is also significant, comparing to the previous 4 outputs; (3) Surprisingly, for ~~four of the outputs, namely~~ some variables (~~e.g.,~~ sensible heat, upward longwave radiation, net radiation; ~~and~~ 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. Interestingly, the LH50's optimization result for sensible heat is even smaller than that of LH2000. This is because LH sampling is random and the LH 50 sampling may have produced a sample point very close to the global optimum, while the best sample point of LH2000 sampling may be further away from the global optimum. Consequently, the number of samples required for surrogate based optimization varies for different outputs because of the randomness of sampling designs, and the complexity of response surfaces. A more complex surface needs more sample points to build an effective surrogate model, compared to simple surface. Even so, this

result is very encouraging that with the help of surrogate models we can possibly 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 convergence is also listed in Table 3. The result indicated that SCE can get better, or similar optimal NRMSE, but the cost number of original-model runs is larger than that using surrogate model. If the original dynamic model costs too much CPU time to run, surrogate based optimization can be more efficiently than the 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 (e.g., [Boyle *et al.*, 2000; Boyle, 2000; Gupta *et al.*, 1998; Yapo *et al.*, 1998; Vrugt *et al.*, 2003; Bastidas *et al.*, 1999; Gupta *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/constraints: (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-dominated parameter sets. The Pareto optimal set usually contains hundreds of points, 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 but for large complex dynamic models such as regional or global land surface models CoLM, it is generally

impractical, and also unnecessary to run the model in an ensemble mode with hundreds of model runs. For regional or global land surface models coupled with atmospheric models, providing only one parameter set that has good simulation ability for most outputs is a more economical and convenient choice.

In multi-objective optimization, there have been many methods that can transform multiple objectives to single objective. Among them, the weighting function based method is the most intuitive and widely used one. In this paper, we assign higher weights to the outputs with larger errors. In the research of Liu et al. [2005], the RMSE of each outputs were normalized by the RMSE of default parameter set, and each normalized RMSE were assigned equal weights. Van Griensven and Meixner [2007] developed a weighting system based on Bayesian statistics to define ‘high probability regions’ that can give ‘good’ results for multiple outputs. However, both of Liu et al. [2005] and van Griensven and Meixner [2007] tended to assign higher weights to the outputs with lower RMSE, and lower weights to the outputs with higher RMSE. This tendency, although reasonable in the probability meaning, conflicts with our intuitive motivations that we want to emphasis on the poorly simulated outputs with large RMSE. Jackson et al. [2003] assumed Gaussian error in the data and model so that the outputs were in a joint Gaussian distribution, and the multi-objective ‘cost function’ was defined on the joint Gaussian distribution of multiple outputs. In Gupta et al. [1998], a multiple weighting function method is proposed to fully describe the Pareto frontier, if the frontier is convex and model simulation is cheap enough. If one outputs is more important than others, a higher weight should be assigned to it. Marler and Arora [2010] reviewed the applications, conceptual significance and pitfalls of weighting function based optimal methods, and gave some suggestions to avoid blind use of it.

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

NRMSE.

[Table 4]

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

$$F = \sum_{i=1}^n w_i RMSE_i \quad (2)$$

in which the  $RMSE_i$  is the Root Mean Squared Error of each output variable that defined as  $RMSE_i = \frac{1}{N} \sqrt{\sum_{j=1}^N (y_{i,j}^{sim} - y_{i,j}^{obs})^2}$ ,  $w_i$  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 1<sup>st</sup> step. The effectiveness and efficiency of ASMO have been validated in [Wang *et al.*, 2014] using 6D 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 converge 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 can~~ provide only the true approximate global optimum, but ~~this approach is the “not bad” solution that is much~~ cheaper enough we can afford than using traditional approaches such as SCE-UA.

We carried out multi-objective optimization with ASMO using weighting function defined in **equation (2)** and the optimization results are shown in **figure 34 and 45**. To compare, we also carried out the direct optimization using SCE-UA. **Figure 34** presents

the default parameter, the optimal parameter given by ASMO and that given by SCE-UA. **Figure 45** shows the improvements given by ASMO and SCE-UA comparing to the default parameters. From **Figure 45** we can find that all of the outputs are improved nearly 10% except soil temperature, and the improvements made by ASMO is similar with-to that by 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.

[Figure 34]

[Figure 45]

We also used the Taylor diagram [Taylor, 2001] to compare the simulation results during-for the calibration period and the validation period (see **figure 56 and 67**). The optimization results given by SCE-UA 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 **figure 56**, the performance of optimized parameters given by both SCE-UA and ASMO (Case C and D in the Taylor diagram) are better than default parameterization scheme (Case B) except soil temperature. Even though soil temperature simulation is degraded, the correlation coefficients given by all the three cases are higher than 0.9, indicating that this imperfection will not cause significant inconsistency ofin the land surface modelling. In **figure 67**, the performance of the validation period is shown quite similar withto that in the calibration period, indicating that the optimal parameters are well identified and the over-fitting problem is avoided.

[Figure 56]

[Figure 67]

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-UA 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 the 17-category FAO-STATSGO soil dataset [Ji and Dai, 2010], which provides top-layer\_(30cm) and bottom-layer\_(30-100cm) global soil textures and has a 30 seconds resolution. The resolution and accuracy of this dataset may be not accurate enough in A'rou frozen/thaw station, where frequent freezing and thawing occur. 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 freeze/thawing 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 underlying 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 can be very different. For an instance, in CoLM the soil depth is set to 2.86m globally, but actually the soil depth varies in different places. 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. OtherwiseFurther, 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't not be

mitigated by parameter optimization, ~~so we should revise the model physics before parameter optimization.~~

In the optimization results, five of the outputs were improved but only soil temperature became worse. In multi-objective optimization, a compromise is necessary. In this case study, soil temperature requires small P6 and large 36, which conflict with all other outputs. Consequently, improving every output is impossible and some output might be sacrificed. If the cost is affordable and the gain is big enough, such compromise might be worthwhile. In this case study, the smallest weight was assigned to soil temperature. In the optimal solution, the RMSE of soil temperature increases from 2.66 degree to 2.90 degree (only 0.24 degrees larger), but other outputs RMSE can all be improved by about 10%. We think the sacrifice of soil temperature is worthwhile because a negligible degradation of one output can lead to significant improvement of all other outputs.

## 5. Discussion and 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~~ other studies, 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 researches, 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, ~~parameter optimization 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 obtained from by~~ SCE-UA 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~~ much efficiently. ~~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.~~

In the future work, we are going to extend the uncertainty quantification framework to other large complex dynamic models, such as regional-scale land surface models, atmospheric models and climate models. We will look into testing the scalability of the screening, surrogate modelling and optimization techniques on more complex models with more adjustable parameters. We will also investigate the influence of uniformity and stochasticity of initial sampling points, and compare the suitability of different sampling methods. In addition to examining the main and total effects of the parameters, we will also evaluate the interactions among parameters. We will continue to improve the effectiveness, efficiency, flexibility and robustness of Gaussian Processes Regression approach for surrogate modelling, and test with more complex models. Since weighting function based multi-objective optimization methods are simple, intuitive and effective, an inter-comparison of different weighting systems can be an interesting topic worthy of further research. Further, we intend to investigate ways to identify Pareto optimal parameter sets using a surrogate based optimization approach.

Discussion and collaborations are warmly welcomed on this and ongoing works. The computer code used in this study is available from the first author, which going to be published as part of the ‘UQlab’ software package in the future.

## 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 the Fundamental Research Funds for the Central Universities - 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://westdc.westgis.ac.cn>) for providing the meteorological forcing data, to the group of Prof. 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.

批注 [g1]: Change the funding resource.

## Appendix A. Surrogate modelling approaches

### A.1 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 product spline basis functions as follows:

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

where  $y$  is the output variable and  $\mathbf{x} = (x_1, x_2, \dots, x_n)$  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  $B_m(\mathbf{x}) = \prod_{k=1}^{K_m} [s_{km}(x_{v(k,m)} - t_{km})]_+$ ,  $k$  is the index of knots and  $K_m$  is the total number of knots;  $s_{km}$  take on value  $\pm 1$  and indicate the right/left sense of associated step function,  $v(k, m)$  is the index of the input variable in vector  $\mathbf{x}$ , and  $t_{km}$  indicates the knot location of the  $k$ -th knot in the  $m$ -th 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].

### A.2 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 the function-space view, a Gaussian process can be completely specified by its mean function and covariance function:

$$\begin{cases} m(\mathbf{x}) = E[f(\mathbf{x})] \\ k(\mathbf{x}, \mathbf{x}') = E[(f(\mathbf{x}) - m(\mathbf{x}))(f(\mathbf{x}') - m(\mathbf{x}'))] \end{cases} \quad (\text{A.2})$$

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

Suppose a nonlinear regression model

$$y = f(\mathbf{x}) + \varepsilon \quad (\text{A.3})$$

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

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{f}_* \end{bmatrix} \sim N \left( \mathbf{0}, \begin{bmatrix} K(X, X) + \sigma_n^2 I & K(X, X_*) \\ K(X_*, X) & K(X_*, X_*) \end{bmatrix} \right) \quad (\text{A.4})$$

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

$$E(\mathbf{f}_*) = K(X_*, X)[K(X, X) + \sigma_n^2 I]^{-1} \mathbf{y} \quad (\text{A.5})$$

$$\text{cov}(\mathbf{f}_*) = K(X_*, X_*) - K(X_*, X)[K(X, X) + \sigma_n^2 I]^{-1} K(X, X_*) \quad (\text{A.6})$$

In this example, the outputs  $\mathbf{y}$  is centered to zero so that the mean function is  $m(\mathbf{x}) = 0$ , while each element of covariance matrixes equals to the covariance function  $k(\mathbf{x}, \mathbf{x}')$  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 covariance function depending on their prior knowledge. In this paper, we choose Martén

554 covariance function:

$$k(r) = \frac{2^{1-\nu}}{\Gamma(\nu)} \left( \frac{\sqrt{2\nu}r}{l} \right)^\nu K_\nu \left( \frac{\sqrt{2\nu}r}{l} \right) \quad (\text{A.7})$$

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

$$k_{\nu=3/2}(r) = \left( 1 + \frac{\sqrt{3}r}{l} \right) \exp \left( -\frac{\sqrt{3}r}{l} \right) \quad (\text{A.8})$$

$$k_{\nu=5/2}(r) = \left( 1 + \frac{\sqrt{5}r}{l} + \frac{5r^2}{3l^2} \right) \exp \left( -\frac{\sqrt{5}r}{l} \right) \quad (\text{A.9})$$

批注 [g2]: A typo, v=3/2 shuld be v=5/2

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

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

$$\log[p(\mathbf{y}|\mathbf{X})] = -\frac{1}{2}\mathbf{y}^T(K + \sigma_n^2 I)^{-1}\mathbf{y} - \frac{1}{2}\log|K + \sigma_n^2 I| - \frac{n}{2}\log 2\pi \quad (\text{A.10})$$

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

### 571 A.3 Random Forests (RF)

572 Random Forests (RF) [Breiman, 2001] are-is a combination of Classification and  
 573 Regression Trees (CART) [Breiman *et al.*, 1984]. Generally speaking, Tree-based  
 574 methods split the feature space into a set of rectangles and fit the samples in each  
 575 rectangle with a class label (for classification problems) or a constant value (for

regression problems). In this paper only regression tree was discussed. Suppose  $\mathbf{x} = (x_1, x_2, \dots, x_n)$  is the  $n$ -dimensional input feature vector and  $y$  is the output response, the regression tree can be expressed as follows:

$$\hat{f}(\mathbf{x}) = \sum_{m=1}^M c_m I(\mathbf{x} \in R_m) \quad (\text{A.11})$$

$$I(\mathbf{x} \in R_m) = \begin{cases} 1, & \mathbf{x} \in R_m \\ 0, & \mathbf{x} \notin R_m \end{cases} \quad (\text{A.12})$$

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.

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. A Random forests construct many trees using randomly selected outputs and features, and synthetic the outputs of all the trees to give the prediction result. A 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  $(\mathbf{x}_i, y_i), i = 1, 2, \dots, N$ , 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 a random forests is given by averaging the output of  $t$  trees.

$$\hat{f}_{rf}(\mathbf{x}) = \sum_{j=1}^t \hat{f}_j(\mathbf{x}) \quad (\text{A.13})$$

Random forests have outstanding performance for very high dimensional

problems, such as medical diagnosis and document retrieval. Such problems usually have hundreds or thousands of input variables (features), ~~with but~~ each ~~one-feature only~~ provides only 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-with~~ 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.

#### A.4 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  $f(\mathbf{x})$  that can fit the output  $y$  with minimum risk given a  $N$  point training set  $(\mathbf{x}_i, y_i), i = 1, 2, \dots, N$ . Take a simple linear regression model for example, the function  $f(\mathbf{x})$  can be:

$$f(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b \quad (\text{A.14})$$

where  $\mathbf{w}$  is the weighting vector and  $\mathbf{x}$  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:

$$f(\mathbf{x}) = \mathbf{w}^T \phi(\mathbf{x}) + b \quad (\text{A.15})$$

where  $\phi(\mathbf{x})$  is the mapping function. The inner product of mapping function is called Kernel Function:  $K(\mathbf{x}, \mathbf{x}') = \phi(\mathbf{x})^T \phi(\mathbf{x}')$  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:

$$K(\mathbf{x}, \mathbf{x}') = \exp(-\gamma|\mathbf{x} - \mathbf{x}'|^2) \quad (\text{A.16})$$

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

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

$$|y - f(\mathbf{x})|_\varepsilon = \begin{cases} 0, & \text{if } |y - f(\mathbf{x})| \leq \varepsilon \\ |y - f(\mathbf{x})| - \varepsilon, & \text{otherwise} \end{cases} \quad (\text{A.17})$$

The loss function means regression errors less than tolerance  $\varepsilon$  are not penalized. The penalty-free zone is also called  $\varepsilon$ -tube or  $\varepsilon$ -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  $\varepsilon$ -tube) can be written as a quadratic optimization problem:

$$\begin{aligned} \min_{\mathbf{w}, b, \xi, \xi^*} \quad & \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^n \xi_i + C \sum_{i=1}^n \xi_i^* \\ \text{subject to} \quad & \mathbf{w}^T \phi(\mathbf{x}_i) + b - y_i \leq \varepsilon + \xi_i \\ & y_i - \mathbf{w}^T \phi(\mathbf{x}_i) - b \leq \varepsilon + \xi_i^* \\ & \xi_i, \xi_i^* \geq 0, i = 1, 2, \dots, n \end{aligned} \quad (\text{A.18})$$

The problem can be transferred to the dual problem:

$$\begin{aligned} \min_{\mathbf{w}, b, \xi, \xi^*} \quad & \frac{1}{2} (\boldsymbol{\alpha} - \boldsymbol{\alpha}^*)^T \mathbf{K} (\boldsymbol{\alpha} - \boldsymbol{\alpha}^*) + \varepsilon \sum_{i=1}^n (\alpha_i + \alpha_i^*) \\ & + \sum_{i=1}^n y_i (\alpha_i - \alpha_i^*) \\ \text{subject to} \quad & \mathbf{e}^T (\boldsymbol{\alpha} - \boldsymbol{\alpha}^*) = 0 \\ & y_i - \mathbf{w}^T \phi(\mathbf{x}_i) - b \leq \varepsilon + \xi_i^* \\ & 0 \leq \alpha_i, \alpha_i^* \leq C, i = 1, 2, \dots, n \end{aligned} \quad (\text{A.19})$$

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

$$f(\mathbf{x}) = \sum_{i=1}^n (-\alpha_i + \alpha_i^*) K(\mathbf{x}_i, \mathbf{x}) + b \quad (\text{A.20})$$

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

## A.5 Artificial Neural Network (ANN)

Artificial Neural Network (ANN) [Jain et al., 1996] is time-honored

批注 [g3]:

machine learning method comparing to the former four. It is a data-driven process that can solve complex nonlinear relationships between input and output data. A neural network is constructed by many interconnected neurons. Each neuron can be mathematically described as a linear weighing function and a nonlinear activation function:

$$I_i = \sum_{j=1}^n w_{ij}x_j \quad (\text{A.21})$$

$$f_i(I) = \frac{1}{1 + \exp(-I_i)} \quad (\text{A.22})$$

where  $x_j$  is the  $j$ -th input variable,  $w_{ij}$  is the weight and  $I_i$  is the weighted sum of the  $i$ -th neuron. The output of the  $i$ -th 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 demonstrated 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.

## References:

- Bastidas, L. A., H. V. Gupta, S. Sorooshian, W. J. Shuttleworth, and Z. L. Yang (1999), Sensitivity analysis of a land surface scheme using multicriteria methods, *Journal of Geophysical Research: Atmospheres*, 104(D16), 19481-19490.
- Bonan, G. B. (1996), A Land Surface Model (LSM Version 1.0) for Ecological, Hydrological, and Atmospheric Studies: Technical Description and User's Guide, NCAR, Boulder, CO, USA.
- Boyle, D. P., H. V. Gupta, and S. Sorooshian (2000), Toward improved calibration of hydrologic models: Combining the strengths of manual and automatic methods, *Water Resour Res*, 36(12), 3663-3674.
- Boyle, D. P. (2000), Multicriteria calibration of hydrological models, Ph.D. dissertation thesis, University of Arizona, Tucson, USA.
- Breiman, L., J. Friedman, C. J. Stone, and R. A. Olshen (1984), *Classification and Regression Trees*, Chapman and Hall/CRC, Boca Raton, Florida, USA.
- Breiman, L. (2001), Random forests, *Mach Learn*, 45(1), 5-32.



Castelletti, A., F. Pianosi, R. Soncini-Sessa, and J. P. Antenucci (2010), A multiobjective response surface approach for improved water quality planning in lakes and reservoirs, *Water Resour Res*, 46(6), W6502.

Cybenko, G. (1989), Approximation by superpositions of a sigmoidal function, *Mathematics of Control, Signals and Systems*, 2(4), 303-314.

Dai, Y., and Q. Zeng (1997), A Land Surface Model (IAP94) for Climate Studies Part I: Formulation and Validation in Off-Line Experiments, *Adv Atmos Sci*, 14(4), 433-460.

Dai, Y., F. Xue, and Q. Zeng (1998), A Land Surface Model (IAP94) for Climate Studies Part II: Implementation and Preliminary Results of Coupled Model with IAP GCM, *Adv Atmos Sci*, 15(4), 47-62.

Dai, Y. J., X. B. Zeng, R. E. Dickinson, I. Baker, G. B. Bonan, M. G. Bosilovich, A. S. Denning, P. A. Dirmeyer, P. R. Houser, G. Y. Niu, K. W. Oleson, C. A. Schlosser, and Z. L. Yang (2003), The Common Land Model, *B Am Meteorol Soc*, 84(8), 1013-1023.

Dai, Y. J., R. E. Dickinson, and Y. P. Wang (2004), A two-big-leaf model for canopy temperature, photosynthesis, and stomatal conductance, *J Climate*, 17(12), 2281-2299.

Dickinson, R. E., A. Henderson-Sellers, and P. J. Kennedy (1993), Biosphere-atmosphere Transfer Scheme (BATS) Version 1e as Coupled to the NCAR Community Climate Model, NCAR, Boulder, CO, USA.

Duan, Q. Y., S. Sorooshian, and V. K. Gupta (1992), Effective and Efficient Global Optimization for Conceptual Rainfall-Runoff Models, *Water Resour Res*, 28(4), 1015-1031.

Duan, Q. Y., V. K. Gupta, and S. Sorooshian (1993), Shuffled Complex Evolution Approach for Effective and Efficient Global Minimization, *J Optimiz Theory App*, 76(3), 501-521.

Duan, Q. Y., S. Sorooshian, and V. K. Gupta (1994), Optimal Use of the SCE-UA Global Optimization Method for Calibrating Watershed Models, *J Hydrol*, 158(3-4), 265-284.

Friedman, J. H. (1991), Multivariate Adaptive Regression Splines, *Ann Stat*, 19(1), 1-14.

Gan, Y., Q. Duan, W. Gong, C. Tong, Y. Sun, W. Chu, A. Ye, C. Miao, and Z. Di (2014), A comprehensive evaluation of various sensitivity analysis methods: A case study with a hydrological model, *Environ Modell Softw*, 51(0), 269-285.

Gorissen, D. (2010), Grid-enabled Adaptive Surrogate Modeling for Computer Aided Engineering, Ph.D. thesis, Ghent University, Ghent, Flanders, Belgium.

Gupta, H. V., S. Sorooshian, and P. O. Yapo (1998), Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information, *Water Resour Res*, 34(4), 751-763.

Gupta, H. V., L. A. Bastidas, S. Sorooshian, W. J. Shuttleworth, and Z. L. Yang (1999), Parameter estimation of a land surface scheme using multicriteria methods, *J Geophys Res*, 104(D16), 19491-19503.

Hastie, T., R. Tibshirani, and J. Friedman (2009), *The Elements of Statistical Learning 2nd*, Springer, New York, USA.

Henderson-Sellers, A., K. McGuffie, and A. J. Pitman (1996), The Project for Intercomparison of Land-surface Parametrization Schemes (PILPS): 1992 to 1995, *Clim Dynam*, 12(12), 849-859.

Hu, Z. Y., M. G. Ma, R. Jin, W. Z. Wang, G. H. Huang, Z. H. Zhang, and J. L. Tan (2003), WATER: dataset of automatic meteorological observations at the A'rou freeze/thaw observation station, Cold and Arid Regions Environmental and Engineering Research Institute, Chinese Academy of Sciences, Lanzhou, China.

Jackson, C., Y. Xia, M. K. Sen, and P. L. Stoffa (2003), Optimal parameter and uncertainty estimation of a land surface model: A case study using data from Cabauw, Netherlands, *Journal of Geophysical*

- 717 *Research: Atmospheres*, 108(D18), 4583.
- 718 Jain, A. K., M. Jianchang, and K. M. Mohiuddin (1996), Artificial neural networks: a tutorial, *Computer*,  
719 29(3), 31-44.
- 720 Ji, D., and Y. Dai (2010), The Common Land Model (CoLM) technical guide, GCESS, Beijing Normal  
721 University, Beijing, China.
- 722 Jin, Y. (2011), Surrogate-assisted evolutionary computation: Recent advances and future challenges,  
723 *Swarm and Evolutionary Computation*, 1(2), 61-70.
- 724 Jones, D. R., M. Schonlau, and W. J. Welch (1998), Efficient global optimization of expensive black-  
725 box functions, *J Global Optim*, 13(4), 455-492.
- 726 Jones, D. R. (2001), A Taxonomy of Global Optimization Methods Based on Response Surfaces, *J*  
727 *Global Optim*, 21(4), 345-383.
- 728 Koziel, S., and L. Leifsson (2013), *Surrogate-Based Modeling and Optimization*, Springer New York,  
729 New York, NY.
- 730 Lepastrier, M., A. J. Pitman, H. Gupta, and Y. Xia (2002), Exploring the relationship between  
731 complexity and performance in a land surface model using the multicriteria method, *J Geophys Res*,  
732 107(4443D20).
- 733 Li, J. (2012), Screening and Analysis of Parameters Most Sensitive to CoLM, Master thesis, BNU,  
734 Beijing, China.
- 735 Li, J., Q. Y. Duan, W. Gong, A. Ye, Y. Dai, C. Miao, Z. Di, C. Tong, and Y. Sun (2013), Assessing  
736 parameter importance of the Common Land Model based on qualitative and quantitative sensitivity  
737 analysis, *Hydrol. Earth Syst. Sci.*, 17(8), 3279-3293.
- 738 Liang, X., E. F. Wood, D. P. Lettenmaier, D. Lohmann, A. Boone, S. Chang, F. Chen, Y. Dai, C.  
739 Desborough, R. E. Dickinson, Q. Duan, M. Ek, Y. M. Gusev, F. Habets, P. Irannejad, R. Koster, K. E.  
740 Mitchell, O. N. Nasonova, J. Noilhan, J. Schaake, A. Schlosser, Y. Shao, A. B. Shmakin, D. Verseghy,  
741 K. Warrach, P. Wetzel, Y. Xue, Z. Yang, and Q. Zeng (1998), The Project for Intercomparison of Land-  
742 surface Parameterization Schemes (PILPS) phase 2(c) Red-Arkansas River basin experiment:: 2. Spatial  
743 and temporal analysis of energy fluxes, *Global Planet Change*, 19(1 - 4), 137-159.
- 744 Liu, Y. Q., L. A. Bastidas, H. V. Gupta, and S. Sorooshian (2003), Impacts of a parameterization  
745 deficiency on offline and coupled land surface model simulations, *J Hydrometeorol*, 4(5), 901-914.
- 746 Liu, Y. Q., H. V. Gupta, S. Sorooshian, L. A. Bastidas, and W. J. Shuttleworth (2004), Exploring  
747 parameter sensitivities of the land surface using a locally coupled land-atmosphere model, *J Geophys*  
748 *Res*, 109(D21101D21), 13.
- 749 Liu, Y. Q., H. V. Gupta, S. Sorooshian, L. A. Bastidas, and W. J. Shuttleworth (2005), Constraining land  
750 surface and atmospheric parameters of a locally coupled model using observational data, *J*  
751 *Hydrometeorol*, 6(2), 156-172.
- 752 Lohmann, D., D. P. Lettenmaier, X. Liang, E. F. Wood, A. Boone, S. Chang, F. Chen, Y. Dai, C.  
753 Desborough, R. E. Dickinson, Q. Duan, M. Ek, Y. M. Gusev, F. Habets, P. Irannejad, R. Koster, K. E.  
754 Mitchell, O. N. Nasonova, J. Noilhan, J. Schaake, A. Schlosser, Y. Shao, A. B. Shmakin, D. Verseghy,  
755 K. Warrach, P. Wetzel, Y. Xue, Z. Yang, and Q. Zeng (1998), The Project for Intercomparison of Land-  
756 surface Parameterization Schemes (PILPS) phase 2(c) Red - Arkansas River basin experiment:: 3.  
757 Spatial and temporal analysis of water fluxes, *Global Planet Change*, 19(1 - 4), 161-179.
- 758 Loshchilov, I., M. Schoenauer, and M. E. L. Sebag (2010), Comparison-based Optimizers Need  
759 Comparison-based Surrogates, in Proceedings of the 11th International Conference on Parallel Problem  
760 Solving from Nature: Part I, pp. 364-373, Berlin, Heidelberg.

- Marler, R. T., and J. Arora (2010), The weighted sum method for multi-objective optimization: new insights, *Struct Multidiscip O*, 41(6), 853-862.
- Marquardt, D. W. (1963), An Algorithm for Least-Squares Estimation of Nonlinear Parameters, *Journal of the Society for Industrial and Applied Mathematics*, 11(2), 431-441.
- McKay, M. D., R. J. Beckman, and W. J. Conover (1979), A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code, *Technometrics*, 21(2), 239-245.
- Minsky, M., and S. A. Papert (1969), *Perceptrons: An Introduction to Computational Geometry*, MIT Press, Cambridge, Massachusetts, USA.
- Ong, Y., P. B. Nair, A. J. Keane, and K. W. Wong (2005), Surrogate-Assisted Evolutionary Optimization Frameworks for High-Fidelity Engineering Design Problems, In ‘Studies in Fuzziness and Soft Computing’, edited by Y. Jin, pp. 307-331, Springer Berlin Heidelberg.
- Pil á, M., and R. Neruda (2013), Surrogate model selection for evolutionary multiobjective optimization, paper presented at Evolutionary Computation (CEC), 2013 IEEE Congress on, 2013-01-01.
- Pollacco, J. A. P., B. P. Mohanty, and A. Efstratiadis (2013), Weighted objective function selector algorithm for parameter estimation of SVAT models with remote sensing data, *Water Resour Res*, 49(10), 6959-6978.
- Rasmussen, C. E., and C. K. I. Williams (2006), *Gaussian Processes for Machine Learning*, MIT Press, Massachusetts, USA.
- Razavi, S., B. A. Tolson, and D. H. Burn (2012), Review of surrogate modeling in water resources, *Water Resour Res*, 48(7), W7401.
- Shahsavani, D., S. Tarantola, and M. Ratto (2010), Evaluation of MARS modeling technique for sensitivity analysis of model output, *Procedia - Social and Behavioral Sciences*, 2(6), 7737-7738.
- Shangguan, W., Y. Dai, B. Liu, A. Zhu, Q. Duan, L. Wu, D. Ji, A. Ye, H. Yuan, Q. Zhang, D. Chen, M. Chen, J. Chu, Y. Dou, J. Guo, H. Li, J. Li, L. Liang, X. Liang, H. Liu, S. Liu, C. Miao, and Y. Zhang (2013), A China data set of soil properties for land surface modeling, *Journal of Advances in Modeling Earth Systems*, 5(2), 212-224.
- Song, X., C. Zhan, and J. Xia (2012), Integration of a statistical emulator approach with the SCE-UA method for parameter optimization of a hydrological model, *Chinese Sci Bull*, 57(26), 3397-3403.
- Taylor, K. E. (2001), Summarizing multiple aspects of model performance in a single diagram, *Journal of Geophysical Research: Atmospheres*, 106(D7), 7183-7192.
- van Griensven, A., and T. Meixner (2007), A global and efficient multi-objective auto-calibration and uncertainty estimation method for water quality catchment models, *JOURNAL OF HYDROINFORMATICS*, 9(4), 277-291.
- Vapnik, V. N. (1998), *Statistical Learning Theory*, John Wiley & Sons, New York, USA.
- Vapnik, V. N. (2002), *The Nature of Statistical Learning Theory*, 2nd ed., Springer, New York.
- Vrugt, J. A., H. V. Gupta, W. Bouten, and S. Sorooshian (2003), A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters, *Water Resour Res*, 39(12018), 1201.
- Vrugt, J. A., H. V. Gupta, L. A. Bastidas, W. Bouten, and S. Sorooshian (2003), Effective and efficient algorithm for multiobjective optimization of hydrologic models, *Water Resour Res*, 39(12148), 1214.
- Wang, C., Q. Duan, W. Gong, A. Ye, Z. Di, and C. Miao (2014), An evaluation of adaptive surrogate modeling based optimization with two benchmark problems, *Environ Modell Softw*, 60(0), 167-179.
- Wang, Q. J. (1991), The Genetic Algorithm and Its Application to Calibrating Conceptual Rainfall-

# Surrogate based parameter optimization of CoLM

805   Runoff Models, *Water Resour Res*, 27(9), 2467-2471.

806   Wood, E. F., D. P. Lettenmaier, X. Liang, D. Lohmann, A. Boone, S. Chang, F. Chen, Y. Dai, R. E.

807   Dickinson, Q. Duan, M. Ek, Y. M. Gusev, F. Habets, P. Irannejad, R. Koster, K. E. Mitchel, O. N.

808   Nasonova, J. Noilhan, J. Schaake, A. Schlosser, Y. Shao, A. B. Shmakin, D. Verseghy, K. Warrach, P.

809   Wetzel, Y. Xue, Z. Yang, and Q. Zeng (1998), The Project for Intercomparison of Land-surface

810   Parameterization Schemes (PILPS) Phase 2(c) Red - Arkansas River basin experiment:: 1. Experiment

811   description and summary intercomparisons, *Global Planet Change*, 19(1 - 4), 115-135.

812   Xia, Y., A. J. Pittman, H. V. Gupta, M. Leplastrier, A. Henderson-Sellers, and L. A. Bastidas (2002),

813   Calibrating a land surface model of varying complexity using multicriteria methods and the Cabauw

814   dataset, *J Hydrometeorol*, 3(2), 181-194.

815   Yapo, P. O., H. V. Gupta, and S. Sorooshian (1998), Multi-objective global optimization for hydrologic

816   models, *J Hydrol*, 204(1-4), 83-97.

817   Zhan, C., X. Song, J. Xia, and C. Tong (2013), An efficient integrated approach for global sensitivity

818   analysis of hydrological model parameters, *Environ Modell Softw*, 41(0), 39-52.

819

820

821

## Surrogate based parameter optimization of CoLM

### Figure list:

**Figure 1:** Inter-comparison of 5 surrogate modelling methods, [error of training set](#).

**Figure 2:** Inter-comparison of 5 surrogate modelling methods, [error of testing set](#).

**Figure 23:** Single-objective optimization result: optimal parameters.

**Figure 34:** Optimal value of CoLM given by multi-objective optimization (comparing default parameter, optimal parameter given by ASMO and SCE-UA)

**Figure 45:** Comparing the improvements given by ASMO and SCE.

**Figure 56:** Taylor diagram of simulated fluxes during calibration period (Jan-1-2009 to Dec-31-2009).

**Figure 67:** Taylor diagram of simulated fluxes during validation period (Here we use warm-up period as validation period, Jan-1-2008 to Dec-31-2008).

**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	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	zlnl	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	sqrtldi	m <sup>-0.5</sup>	canopy	the inverse of square root of leaf dimension	[2.5, 7.5]
P18	effcon	mol CO <sub>2</sub> /mol quanta	canopy	quantum efficiency of vegetation photosynthesis	[0.035, 0.35]
P19	vmax25	mol CO <sub>2</sub> /m <sup>2</sup> s	canopy	maximum carboxylation rate at 25°C	[10 <sup>-6</sup> , 200 <sup>-6</sup> ]
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]

# Surrogate based parameter optimization of CoLM

P35	tran(2,2)		canopy	NIR transmittance of dead leaf	[0.3, 0.5]
P36	z0m	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	canopy	maximum transpiration for vegetation	[1.e-4, 1. e-2]

838

839

840 **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	<del>P2, P3, P4, P6, P18, P30, P36, P38</del> <del>P3, P4, P6, P18, P19, P23, P25, P36</del>
Upward Longwave Radiation	P6, P17, P36
Net radiation	P6, P17, P30, P34, P36
Soil Temperature	P3, P6, P36
Soil Moisture	P3, P6

批注 [g4]: This screened parameters were suggested by [Li, J. (2012)]. It is corrected to be consistent with [Li, J. (2012)] and other parts of this paper.

841

842



# Surrogate based parameter optimization of CoLM

843

844 **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	0.0590	0.2357	0.0096	0.4527
SCE	0.7450	0.4921	0.0380	0.1963	0.0073	0.3900
Optimized	(1492)	(1354)	(458)	(982)	(473)	(210)
LH50	0.7672	0.5255	0.0377	0.1913	0.0080	0.4222
LH100	0.7841	0.5785	0.0372	0.1908	0.0077	0.4130
LH200	0.7821	0.5885	0.0374	0.1928	0.0069	0.3947
LH400	0.7798	0.5627	0.0374	0.1928	0.0070	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

845

846

Surrogate based parameter optimization of CoLM

847

848

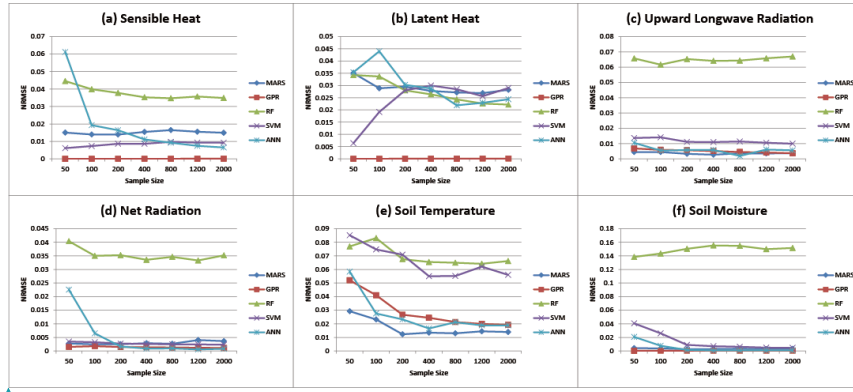
**Table 4:** Weights assigned to each output variables.

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

849

850

# Surrogate based parameter optimization of CoLM



**Figure 1:** Inter-comparison of 5 surrogate modelling methods, error of training set.

带格式的：字体：（默认）Times New Roman，加粗

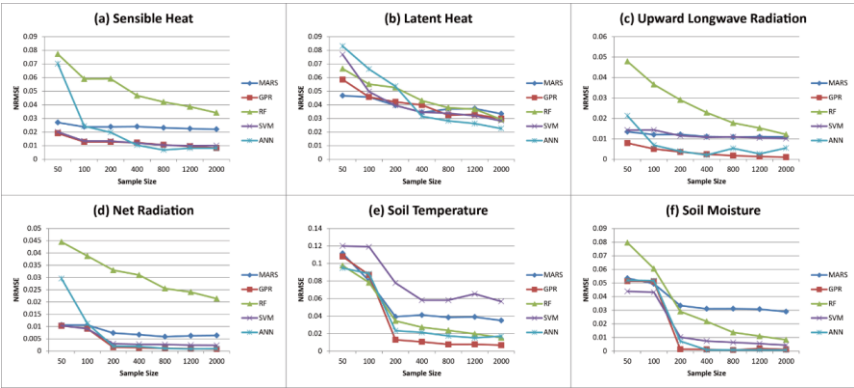
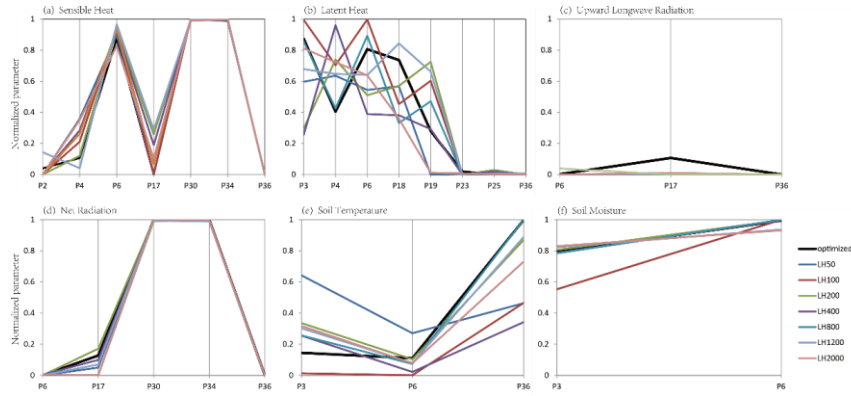


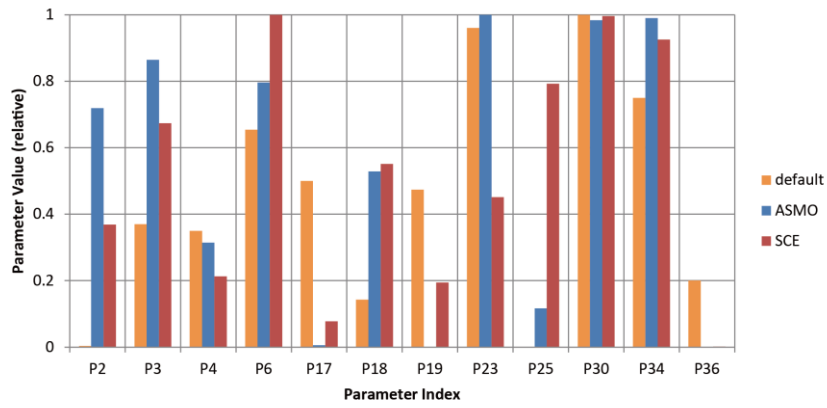
Figure 21: Inter-comparison of 5 surrogate modelling methods, error of testing set.

# Surrogate based parameter optimization of CoLM



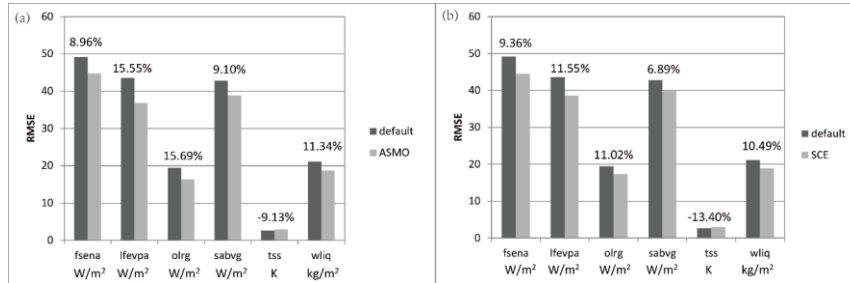
**Figure 2:** Single-objective optimization result: optimal parameters.

Surrogate based parameter optimization of CoLM



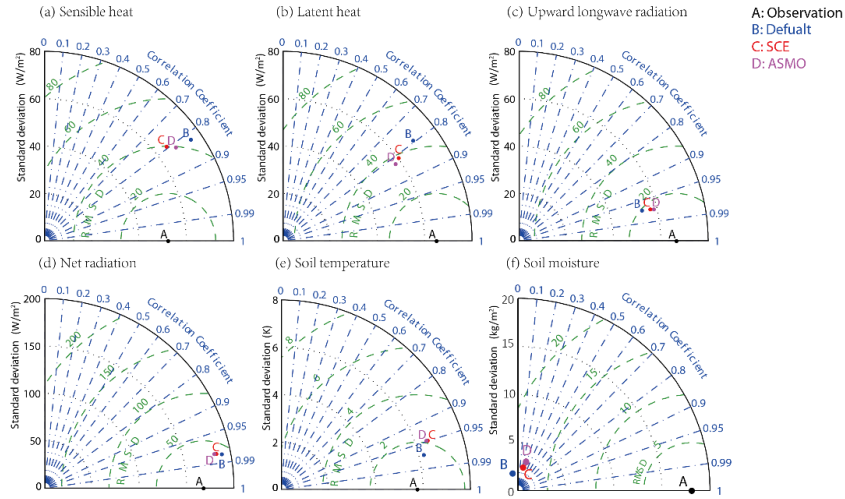
**Figure 3:** Optimal value of CoLM given by multi-objective optimization (comparing default parameter, optimal parameter given by ASMO and SCE-UA)

# Surrogate based parameter optimization of CoLM



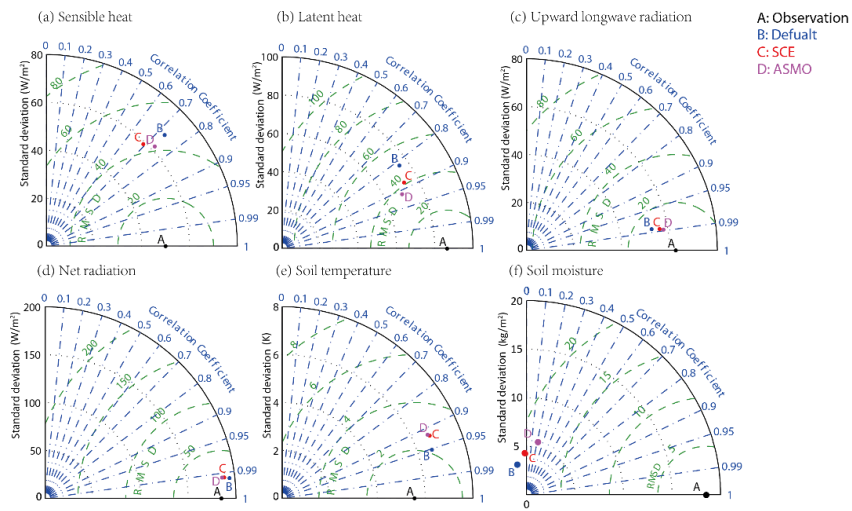
**Figure 4:** Comparing the improvements given by ASMO and SCE.

# Surrogate based parameter optimization of CoLM



**Figure 5:** Taylor diagram of simulated fluxes during calibration period (Jan-1-2009 to Dec-31-2009).





**Figure 6:** Taylor diagram of simulated fluxes during validation period (Here we use warm-up period as validation period, Jan-1-2008 to Dec-31-2008).