

Response letter to Anonymous Referee #2

Comment:

This paper presents an application of an efficient parameter optimization technique developed by the authors. As the method is recently published and parts of it have already been presented elsewhere, the paper's contribution lies in the analysis of different surrogate model construction methods with respect to their application to multi-objective optimization of a land surface model. To a certain extent, the paper achieves this aim. There are however a number of improvements that I would expect to see before publication as a final paper.

Response:

First, we would like to thank the editor and all the reviewers for your kind, helpful comments on this manuscript. We have enclosed a revised version and two response letters. Hopefully they can appropriately address the concerns in the review letters.

Comment:

The most important improvements required are:

1. Acknowledgement and literature review of existing surrogate-based optimisation Techniques.
2. Discussion of the generality of the conclusions and assumptions underlying the results obtained
3. Clarification of some points which are not sufficiently well presented
4. Editing of typos, editing of language and filling in of placeholders
e.g. 'Artificial Neural Network (ANN) (REF) is a time-honored machine learning method comparing to the former four' p6736, L3

Response:

Thank you for your helpful comments. The revised manuscript has been improved in the following aspects.

1. A literature review about surrogate-assisted optimization and its application in hydrology has been added in the introduction section.
2. A paragraph about limitations of current work and future research was added to section 5: Discussion and conclusions.
3. Clarification of 'weighting functions', 'uncertainty quantification framework' and 'elbow points' as well as other topics were added to the revised version.
4. The placeholder 'REF' was replaced by [Jain et al., 1996] and typo 'machine' was also corrected. Other typos were also corrected, as listed in the end of the response letter.

Comment:

While the paper is already of interest, these changes would improve the quality of the manuscript and give the reader a clearer impression of the context and utility of the method proposed by the authors.

More detailed discussion follows:

1. Acknowledgement and literature review of existing surrogate-based optimization

techniques (dating back to at least 2001)

Given that the paper aims to promote the use of the author's new adaptive surrogate model based optimization (ASMO) strategy (p6718 L25), I would expect to see acknowledgement and a brief overview of existing surrogate-based optimization techniques within the main text, even if the authors have mentioned this literature in other papers/journals.

The paper appears to contain only one other reference to existing surrogate-based optimisation techniques: 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. p6729, L18 Appendix A1 '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 quick search for the terms "surrogate optimization" brings up at least the following references, dating back to at least 2001 and including several reviews, a book and open-source implementations.

Ong, YewSoon, P B Nair, A J Keane, and K W Wong. 2005. "Surrogate-Assisted Evolutionary Optimization Frameworks for High-Fidelity Engineering Design Problems." In *Knowledge Incorporation in Evolutionary Computation SE - 15*, edited by Yaochu Jin, 167:307–31. *Studies in Fuzziness and Soft Computing*. Springer Berlin Heidelberg. doi:10.1007/978-3-540-44511-1_15.

Koziel, Slawomir, and Leifur Leifsson, eds. 2013. *Surrogate-Based Modeling and Optimization*. New York, NY: Springer New York. doi:10.1007/978-1-4614-7551-4.

Forrester, Alexander I.J., and Andy J. Keane. 2009. "Recent Advances in Surrogate-Based Optimization." *Progress in Aerospace Sciences* 45 (1-3): 50–79. doi:10.1016/j.paerosci.2008.11.001.

Jin, Yaochu. 2011. "Surrogate-Assisted Evolutionary Computation: Recent Advances and Future Challenges." *Swarm and Evolutionary Computation* 1 (2): 61–70. doi:10.1016/j.swevo.2011.05.001.

Jones, Donald R. 2001. "A Taxonomy of Global Optimization Methods Based on Response Surfaces." *Journal of Global Optimization* 21 (4). Kluwer Academic Publishers: 345–83. doi:10.1023/A:1012771025575.

Queipo, Nestor V., Raphael T. Haftka, Wei Shyy, Tushar Goel, Rajkumar Vaidyanathan, and P. Kevin Tucker. 2005. "Surrogate-Based Analysis and Optimization." *Progress in Aerospace Sciences* 41 (1): 1–28. doi:10.1016/j.paerosci.2005.02.001.

Zhou, Aimin, Bo-Yang Qu, Hui Li, Shi-Zheng Zhao, Ponnuthurai Nagarathnam Suganthan, and

Qingfu Zhang. 2011. "Multiobjective Evolutionary Algorithms: A Survey of the State of the Art." *Swarm and Evolutionary Computation* 1 (1): 32–49. doi:10.1016/j.swevo.2011.03.001.

<http://www.mathworks.com/matlabcentral/fileexchange/38530-surrogate-model-optimization-toolbox>

Optimization methods that aim to provide satisfactory solutions given a limited computational budget should also be mentioned, in light of the comment that "Such parameter set might not be the true global optimum, but it is the "not bad" solution that is cheap enough we can afford." (P6725, L23) e.g. Tolson, B. A., and C. A. Shoemaker (2007), Dynamically dimensioned search algorithm for computationally efficient watershed model calibration, *Water Resour. Res.*, 43, W01413, doi:10.1029/2005WR004723.

Response:

Thank you for your helpful suggestion. We added literature review on surrogate based optimization for LSMs in the introduction section. Please note that this paper is not intended to propose new methods or theories, but to integrate existing techniques to improve the simulation ability of a LSM. See below on the revision:

"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. Pilát 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."

Comment:

2. Discussion of the generality of the conclusions and assumptions underlying the results obtained The application uses 40 parameters (p6719 L11) of a single land surface model

(CoLM) applied to a single column case study (p6716, L17). It concludes (p6716 L19-22) that: "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."

This seems like a big jump given that doing so might involve:

- scalability of the technique to more parameters
- suitability of the response surface of the "other large complex dynamic models" to be fit by the surrogate technique
- suitability of the runtime of these larger models (411 may still be prohibitively large?)
- case-specific requirements as to how close it is necessary to be to the optimal value
- given that the result is not actually optimal, but rather 'similar' with the one gotten from SCE method using more than 1000 model runs' (p6728, L12)
- availability of software that can be used with those larger models - given that software availability is not discussed at all
- varying impacts of considering only sensitive parameters, where insensitive parameters may have significant interactions with sensitive parameters
- effect of stochasticity in sampling points to build and adapt the surrogate

...

At the very least, it would be useful for the paper to try to explain the factors affecting performance of the method, and its corresponding limitations.

Response:

Thank you for your constructive suggestions.

The last sentence of the abstract was revised as follows.

"The result indicates that this framework can efficiently archive 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."

The following sentence was removed from the conclusions.

"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 revised section 5, we added a lot of discussion about the factors affecting the performance, the method's limitations, and future works.

"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 modeling 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 modeling, 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."

Comment:

3. Clarification of some points which are not sufficiently well presented

- p6716 the paper refers to the framework alternatively as a "uncertainty qualification framework" and "uncertainty quantification framework". In any case it is unclear how uncertainty is addressed at all, given that the result is a single set of optimal parameters corresponding to a single weighted objective function of multiple outputs.

Response:

To be consistent, we use "uncertainty quantification framework" in the revised version.

This "uncertainty quantification framework" includes but not limited to sensitivity analysis, parameter screening, surrogate modeling, single/multi-objective optimization, confidence interval analysis and risk analysis. Parameter specification is one major source of model uncertainty, and parameter uncertainty is the most effective way to reduce the uncertainty. So in this paper, we kept using the term "uncertainty quantification framework".

Comment:

- Discussion of the sufficient number of points should recognize the statistical fact that error commonly continues to decrease as sample size increases and that the sufficient number of points therefore depends on the required error for a particular purpose. It therefore seems misleading to say that: 'error becomes stable' p6721, L28 'elbow points' p6722, L5

It may also be of interest to mention that the absolute error in the surrogate's estimate of the objective function is less important than the ability of the surrogate to capture the relative value of the objective function across parameter space.

Loshchilov, Ilya, Marc Schoenauer, and Michèle Sebag. 2010. "Comparison-Based Optimizers Need Comparison-Based Surrogates", September. Springer-Verlag, 364-73. <http://dl.acm.org/citation.cfm?id=1885031.1885071>.

Response:

Thank you for your suggestion and the draft was revised as follows:

(1) "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." => "The marginal benefits of additional samples becomes less or even negligible if the sample size is larger than 400."

(2) "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." => "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 more than 200 sample points.”

(3) The suggested reference was added to the literature review in section 1, as shown in the response to comment 1.

Comment:

- The paper combines multiple objectives using weights. It states: "assign more weight to the objective function output, if that output is simulated more poorly as compared to other outputs" (p6724, L26) Clarification is needed to the effect that the weighting is calculated based on performance of the default parameters, i.e. outputs that need more improvement are emphasized.

Otherwise I would expect that poor performance might be an indicator of error in data or model structure, in which case it is customary to assign weight inversely proportional to the error. The reference cited (Liu et al. 2005) divides each objective function by its performance with the default parameters, which appears to have the opposite effect-of giving less weight to poorly performing outputs?

Response:

The reviewer raised an interesting concern about the weighing system. The reviewer #1 also asked similar question. Actually in [Liu et al., 2005], W_i was proportional to $1/f_i(\text{default})$, which means assign larger weights to smaller error outputs. [Liu et al., 2005]'s explanation is (1) Define objective function $f_i = \text{RMSE}_i$, and then the normalized objective $f'_i = \text{NRMSE}_i$, in which the RMSE was normalized by RMSE simulated by default parameters. (2) Assign equal weights to the normalized objectives. Consequently, the weights assigned to RMSE_i was actually $1/\text{RMSE}_i(\text{default})$. The consideration of [Liu et al., 2005] was to averagely assign weights to each output, considering their magnitude of error, and make sure their weights were approximately the same after normalization.

In this paper, our consideration is as follows: Consider two outputs A and B, if we want to optimize A without considering B, we assign $W_a = 1$ and $W_b = 0$. Similarly, if we want to consider A twice as important as B, assign $W_a = 2/3$ and $W_b = 1/3$. In this case study, every output is important but we want to improve the worst ones, so a larger weight was assigned to outputs with larger NRMSE.

A paragraph about our thinking about weighing systems, as well as our plan of future works on transforming multi-obj to single-obj, has been added to section 4.2.

“In multi-objective optimization, there have been many methods that can transform multiple objectives to a 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 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 output is more important than the 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.”

Comment:

4. Editing of typos, editing of language and filling in of placeholders There are sufficiently many improvements necessary that the manuscript should be thoroughly reviewed in full. This includes but is not limited to pages 6716, 6718, 6723-6728, 6730-6733, 6736

Response:

The suggestion is very helpful and the manuscript has been revised thoroughly, the typos are listed below.

Line 23: need => needs; huge=>large

Line 25: quantification

Line 26: aforementioned challenges => aforementioned challenges, which include the following steps

Line 26, 27, 29, 30: use => using

Line 29: promote => improve

Line 32: case study of a => application of a

Line 33: Common Land Model => the Common Land Model

Line 81: multi-objective optimization => multi-objective optimization approach

Line 82: “know how to” deleted

Line 109: “only” deleted

Line 114: in => of

Line 114: models => model

Line 115: on => with

Line 123: on => of

Line 125: will => would

Line 128: conclusion => discussion and conclusions

Line 145: Normalized Mean Squared Error => Normalized Root Mean Squared Error

Line 183: even millions => even up to millions

Line 189: may => can

Line 192: pitfall => pitfalls

Line 199: The sample sizes are set to 50, 100, 200, 400, 800, 1200, and 2000. => add “respectively” to the end of the sentence.

Line 224: The error of surrogate model => The error of a surrogate model

Line 231: good => acceptable

Line 232: speed => speeds
Line 232: is => are
Line 240: "among all methods," deleted
Line 240: stable => consistently
Line 241: "following" deleted
Line 242: multi-objective optimization analysis. => multi-objective optimization analysis presented later.
Line 257: "in the following text" deleted
Line 260: sample points => model runs
Line 262: converged => optimal
Line 262: from the => given by
Line 263: "runs" deleted
Line 275: "significantly" deleted
Line 277: given => obtained; "the" deleted
Line 282: size => sizes
Line 299: cost => number; "original" deleted; larger than => larger than that
Line 302: efficiency => efficient; SCE => the SCE
Line 305 and 306: for P6 => of P6
Line 307: "measures" deleted
Line 308: "objective" deleted
Line 319: conditions => constrains
Line 321: non-inferior => non-dominated
Line 376:
Such parameter set might not be the true global optimum, but it is the "not bad" solution that cheap enough we can afford. =>
Such parameter set can provide only the approximate global optimum, but this approach is much cheaper than using traditional approaches such as SCE-UA.
Line 387: with => to that by
Line 390: "at" deleted
Line 398: during => for the; validation period => the validation period
Line 399, 406, 420, 468: SCE => SCE-UA
Line 405: parameter => parameters
Line 407: expect => except
Line 407: Even though => Even though soil temperature simulation is degraded
Line 410: of => in; validation period => the validation period
Line 411: is quite similar with => is shown quite similar to; calibration period => the calibration period
Line 424: provide => provides
Line 426: A'rou frozen/thaw station => A'rou station, where frequent freezing and thawing occur.
Line 430: frozen/thaw => freezing/thawing
Line 436: is very different => can be very different
Line 436: for an instance => for instance
Line 438: place => places

Line 440: otherwise => further

Line 442: can't => cannot

Line 458: many similar works => other studies

Line 465: it is impractical to parameter optimization => parameter optimization is impractical

Line 468: gotten from => obtained by

Line 470: with only hundreds of model runs => much efficiently

Line 543: test input and output => testing inputs and outputs

Line 545: predict output => predicted outputs

Line 556: positive hyper parameters => are positive hyper parameters

Line 572: Random Forests are => Random Forest is

Line 586, 588, 598: random forests => a random forest

Line 587: output => outputs

Line 599: outstanding performance in => outstanding performance for

Line 601: with each one only provides a little => but each feature provides only a little

Line 604: using => with

Line 642: REF => [Jain et al., 1996]

Line 643: machine => machine