Dear Dr. Dimitri Solomatine,

Thank you for your decision regarding the revision of our manuscript (hess-2014-83):" Multiobjective sensitivity analysis and optimization of a distributed hydrologic model MOBIDIC" (Yang, J. et al.).

We have done carefully revised the entire manuscript based on the comments of the reviewers. We have addressed all the comments and incorporated nearly all the suggestions made by the reviewers.

I have enclosed the revised manuscript and the point-by-point responses to the reviewers. Should you need to contact me, please contact me by e-mail yangjing@ms.xjb.ac.cn

Thank you so much. I look forward to hearing from you.

Sincerely,

Yang et al.

Reply to review comments from Dr. Rafael Rosolem

[1]. The analyses of parameter sensitivity are conducted separately for each objective function, and ultimately a choice is made to identify sensitive and insensitive parameters in a multi-objective context. This is what we called "pseudo-multiobjective" sensitivity analysis in our studies. The last paragraph in Section 5.1 exemplifies the known difficulties associated with choosing sensitive parameters using those "pseudomultiobjective" approaches, ultimately leading to unavoidable degree of subjectivity. It would be nice if the authors could discuss how their method compares to one of the "pseudo" methods as well as with the fully multiobjective criteria approach proposed by Rosolem et al (2012,2013), and discuss advantages and limitations.

Our reply: In the paper, our approach is based on the idea: if one factor is sensitive to any objective, then it should be included in the optimization. We cited Rosolem et al (2012,2013) and compared these two approaches in lines 218 to 225:

It is worth noting: this sensitivity analysis approach applied here is not a fully multiobjective sensitivity analysis approach as proposed by Rosolem et al (2012 and 2013) which applies sensitivity analysis to all objectives in an integrated way and is objective. However compared to the fully multiobjctive sensitivity analysis approach (as proposed in Rosolem et al, 2012) which easily requires over 10,000 model runs, our approach is very computationally efficient as both Morris method and SDP method only need several hundred model runs, which is highly appreciable for physically based and distributed hydrologic models.

[2] The fact that both sensitivity analysis approaches show MARD results being nearly the same as those obtained with SRMSE may indicate that those were not conflicting/competing objective functions chosen by the authors. This could possible indicate a poor choice of original objective function. The authors recognize this result in the Conclusion section but did not discuss implications in depth.

Our reply: We added the discussion in lines 430 ~ 433:

The correlation coefficient is low (0.13) between SRMSE and MARD, and is even lower when these two objectives approach to their minima regions (i.e., SRMSE < 0.53 and MARD < 0.09). This might indicate a poor choice of the objective function, as also shown by similar sensitivity results for these two objective functions in Section 5.1

Reply to review comments from Anonymous Referee #1

1.A flowchart figure is suggested to add to clarify the methodology, i.e., the relationship between the two sensitivity methods.

Our reply: After a comparison, instead of a flow chart, we added some sentence in lines 206~210 to clarify the relationship between these two sensitivity methods:

In practice, especially for over-parameterized cases, Morris method is firstly suggested to screen out insensitive factors, and then SDP method is applied to quantify the contributions of the sensitive factors and their interactions. In this study, as model parameters are aggregated into nine factors (as listed Table 1), these two methods are applied individually.

2.The single objective optimization is performed with the Nelder–Mead Simplex algorithm, why not use the Genetic Algorithm and make the comparisons fairer. In fact, epsilon-NSGAII is also very effective for single objective optimization

Our reply: we agree that epsilon-NSGAII is also very effective for single objective optimization. However, as stated in lines 265 and 266 "*And SOO was done with the classic Nelder–Mead algorithm (Nelder and Mead, 1965) which is already coded into the MOBIDIC package*", our objective is to demonstrate "*Multiobjective sensitivity analysis and optimization provides an alternative way for future MOBIDIC modelling*" (lines 42 and 43) and how MOO result can be easily converted to SOO result.

3.In section 5.2, the high flows are underestimated (that can be observed in Fig. 9) because the logarithm scale of the observed and simulated flows (SRMSE and MARD) are chosen. It needs to justify that the purpose of the hydrologic model is not for the flood forecasting. Particularly, MARD seems to more address the normal flows.

Our reply: we added this in lines 331~333.

It is worth noting that we use the logarithms of the flows instead of flows to avoid overfitting flow peaks (Boyle et al., 2000; Shafii and De Smedt, 2009) as flood forecasting is not our main focus

4.In section 5, the authors used SDP method to discuss multiobjective sensitivity analysis quantitatively. This study did not give a threshold for the sensitivity index (vertical axis in Fig. 4) that the factors could be screened out clearly, i.e., has a very low sensitivity index (Fig. 4) while it is chosen as a sensitive factor.

Our reply: though SDP is a quantitative approach, it cannot estimate contributions from higher order interactions (see section 3.1.2). The successful use of SDP depends on the uncertainty that SDP explains. Instead of using a threshold, it is more meaningful to combine the results with other knowledge (e.g., study area and the model itself). In Figure 4, for objectives SRMSE and MARD, it explains over 80%, we can easily identify the five sensitive parameters; for WBI, it explains only 58% total uncertainty, we selected rWcmax based on comparison with Morris results and evaporation characteristic though sensitive indices (main effect and quasi total effect) of rWcmax are low. See our explanation in section 5.1.

Reply to review comments from Anonymous Referee #2

1. [line 17, p. 3510] The authors state "a factor can be a model parameter or a group of model parameters". When they say, a "group": do they mean both spatially distributed parameters, and sets of spatially distributed parameters? This could use clarification.

Our reply: We clarified this in lines 155and 157:

A factor can be a model parameter or a group of distributed model parameters with the same parameter name, and in this paper it is a change to be applied to a group of model parameters

2. [line 19, p. 3512] The authors' methodology screens out and excludes certain model parameters. The authors should review and cite vanWerkhoven et al (2009), Advances in Water Resources. The van Werkhoven study showed that if the wrong metrics were used for reduction of parameters, the Pareto sets can drastically change.

Our reply: We reviewed and cited this paper in 75 ~ 77

and van Werkhoven et al (2009) demonstrates how the calibration result responds to reduced parameter sets with different objectives and different metrics of parameter exclusion

3. [section 3.2] In this section, the review by Efstratiadis and Koutsoyiannis (2010), Hydrological Sciences Journal, should be cited. Additionally there is a typo in the second reference to Kollat and Reed.

Our reply: We cited this paper in lines 97 and 98, and the typo was corrected.

A good review of MOO applications in hydrological modelling is given by Efstratiadis and Koutsoyiannis (2010)..

4. I commend the authors on comparing the results to single objective optimization (equation 6).

Our reply: Thanks.

5. I would recommend some more discussion of Figure 3. Why was it that the factors look like they are appearing in "groups"? I found it difficult to interpret the results of this figure.

Our reply: *Factors in the same group have a similar effect on studied objective functio* . This phenomenon (“group”) is objective dependent and exists for SRMSE and MARD, not for WBI. We added our discussion in lines 352 and 353, and 364 and 365.

In lines 352 and 353:

Factors in the same group have a similar effect on studied objective function

In lines 364 and 365:

For MARD, the results are nearly the same to SRMSE. And this means factors behave similarly to these two objective functions.

6. Several comments about the multiobjective calibration:

6a. When the authors say "converged" what do they mean? They do this both for the SOO and MOO results; mathematical convergence to the true optimal solution cannot be proven.

Our reply: For “Converged”, normally it means the algorithm reaches the stop criteria (e.g., difference in objective functions between two consecutive iterations smaller than a given value) given by the users to let the optimization algorithm know when to stop. However, we agree that convergence to the true optimal solution cannot be proven. Therefore, to avoid confusion, we changed “converged” to “led to” or “stopped”.

6b. The authors claim eNSGAII took more model simulations. This seems like an obvious result – eNSGAII is a population based technique for finding solutions to multiple objective problems. It's like comparing apples and oranges, so to say.

Our reply: It is obvious but it can still gives some readers some idea on how different these two algorithms are in terms of computational cost.

6c. What were the eNSGAII parameters? Injection rate? Initial population size? This is going to affect the computational efficiency (see 6b).

Our reply: We set all eNSGAII parameters to their recommended values, e.g., injection rate 0.25. Initial population size was set to 128, and initial population was generated with Sobol' quasi-random sampling technique. We added these in lines 256 ~ 260, and lines 407 ~ 409:

In lines 256 ~ 260

In this study, two changes were made to the original ε-NSGAII: 1) the initial population is generated with Sobol' quasi-random sampling technique to improve the coverage of parameter space; 2) the code is parallelized and interfaced with MOBIDIC to improve the computational speed.

In lines 407 ~ 409

For MOO, we set the initial population size 128 to obtain a good coverage of the factor space and other ε-NSGAII parameters to their recommended values, …

6d. Was the MOO optimization repeated for multiple random seeds? Was the single objective optimization repeated for different starting points? This is standard practice, and if the authors did not do this it may call the optimization results into question. For example, they claim "Nelder-Meld [...] was dependent on starting point" [line 4, p.3521]. This implies that they tried multiple runs, but I'm not sure.

Our reply: The original paper is based on MOO run from one random seed and SOO run from one starting point. In this revised version, another MOO and SOO were tested (see lines 453~457, and 496~498). The SOO was applied starting from a point nearby the

MOO single optimum and verified our statement "Nelder-Mead [...] was dependent on starting point"

In lines 453~457:

Compared with MOO, obviously, SOO was trapped in the local optima as seen in top-left plot. Another SOO was done with starting point close to the optimum of MOO, and now the optimum of SOO is very close to that of MOO, which means optimization with Nelder-Mead algorithm was dependent of starting point

In lines 496~498:

The result of MOO above is based on a single random seed. The result of MOO with another random seed is similar to the above except that range of rW_{cmax} is narrower (however its effect on the simulation result is limited due to its low sensitivity discussed)

6e. If they did NOT try multiple runs, one justification for this is the computational time that the simulation model takes to run. However, this changes the tone of the study a bit. Now, the use of eNSGAII is to just get good parameter sets – you can't make as many claims about convergence if you don't repeat the study several times for multiple random seeds.

Our reply: Although multi-random-seed MOO is very appealing and we did another MOO run, we don't think it is practical for fully distributed and physically based models due to the computational cost. What we did here is to choose an algorithm (here is eNSGAII) which has been proved robust in the literature, and improved initial coverage of initial population. We added in lines 496~502:

The result of MOO above is based on a single random seed. The result of MOO with another random seed is similar to the above except that range of rW_{cmax} is narrower (however its effect on the simulation result is limited due to its low sensitivity discussed). Multiple-rand-seed MOO is always appealing, but it might not be practical to fully distributed and physically based models which is normally time-consuming in computation. What one can do is to choose a reliable and robust algorithm based on literature review.

7. In the results, is there an approach to choose one solution, and navigate the tradeoffs? The authors may want to refer to Kollat and Reed, 2007, Environmental Modelling and Software for one possible approach. One criticism of multiobjective calibration is that users can eventually only use one parameter set, so approaches should be designed to try to facilitate that choice of parameters.

Our reply: We cited the paper in lines 86~88. And just a note: this paper shows how MOO result can be converted to SOO result by assigning different weights to these objective functions without further optimization.

Lines 86~88:

Although there are criticisms of MOO such as that only one parameter set can be used for decision making, recently researches (e.g., Kollat and Reed, 2007) start to provide the answers