Response letter to reviews of hess-2021-514

"Pitfalls and a feasible solution for using KGE as an informal likelihood function in MCMC methods: DREAM(ZS) as an example"

Yan Liu, Jaime Fernández-Ortega, Matías Mudarra, Andreas Hartmann

We thank the editor and reviewers for their comments that help us to improve our manuscript. In the following, the reviewer comments are shown in regular font, and our point-by-point replies are shown in italic and blue font. Upon revision we have made the following major changes to the manuscript and Supplementary Material:

1. *We changed the setting of case study 1 to derive the pseudo-analytical posterior distribution. Our adapted KGE approach is able to reproduce the shape of the posterior distribution of model parameters;*
2. *We added a calibration using GLUE method (NSE was used as the objective function) and another formal likelihood function using log-transformation in case study 2. Results show that GLUE method has a very wide uncertainty compared to other applied approaches. The log-transformation works well for low flow, while our adapted KGE approach have a good and balanced performance for both low and high flows;*
3. *We changed case study 3 to simultaneously calibrate model parameters using discharge and three solutes. Results show that our adapted KGE approach performs better for discharge and solutes. Most importantly, it can improve the performance for the solutes up to 44%. The adapted KGE approach has advantages compared to the formal likelihood function for calibrations combining different types of observations where the amount and unit of each data variable are different;*
4. *We provide the total uncertainty in case study 3 and put the parameter uncertainty in the supplement. The total uncertainty band of the adapted KGE approach can cover more very low and very high observations especially for solutes.*

Reply to the comments of editor

Thank for your submission and comments to the Reviewers 1 and 2. According to the reviewers' comments, there are important issues you need to address. Among them I would mention specially:

We thank the editor for summarizing the major points to improve our study.

1) Introducing one case study showing that your approach is advantageous over the formal likelihood approach.

   Response: We changed case study 3 to include calibrations using both discharge and solutes. It shows our adapted KGE approach is advantageous for calibrations using multiple types of observation data. In case study 2, we also compared the adapted KGE with another formal likelihood function using the log-transformation approach. Results show that our adapted KGE approach can have a good and balanced performance for both high and low flows. Details please refer to our reply to major comment #1 of reviewer #1 and major comment #1 of reviewer #2.

2) Including one analysis with the known analytical solution of the posterior distribution and compare with the results derived using your adapted KGE approach.

   Response: We changed case study 1 to derive the pseudo-analytical posterior distribution and to show the true model parameters. Our adapted KGE approach can reproduce a similar shape of the posterior distribution. Details please refer to our reply to major comment #2 of reviewer #1.

3) Providing more details on how a formal likelihood approach is compared with your KGE approach.

   Response: We compared the performance on low and high flows between the formal likelihood function, log-transformation and our adapted KGE approach in case study 2. We compared the performance for both discharge and solutes between the formal likelihood function and our adapted KGE approach in case study 3. We also compared the capability of the uncertainty that covers observations between the two approach in case study 3.

After these major revisions, you work can be reconsidered for potential publication at HESS, after further review by editors and referees.

We hope that after the careful revision our revised manuscript can reach the standard for the HESS publication.

Reply to the comments of reviewer #1

The authors proposed an informal likelihood function based on KGE (with modifications), and demonstrated its performance against a formal likelihood function based on RMSE in DREAM_ZS with three cases. There are several key questions that were not clearly answered.

We thank the reviewer for the constructive comments.

1. Why should one use the KGE-based informal likelihood function? Why Gamma distribution? It seems that it is not advantageous over the formal likelihood function in the three case studies. It would be essential to design a case where the formal likelihood function would fail while the KGE-based one still works. Simply introducing a new metric (without solving challenging problems) has no significance.

   Response: The motivation of proposing this adapted KGE is that KGE is widely used as the performance measure in hydrological studies. It is also commonly used as objective function for model calibration. However, there are two problems when using the original KGE in MCMC-type calibrations, which we circumvent with the Gamma distribution: (i) it ensures the monotonically increase of probability density even with negative KGE values, and (ii) it achieves a proper nonlinearity with the increase of model performance. They can lead to an efficient and proper chain evolution. Another advantage of the Gamma distribution is that it does not require the definition of additional parameters, maintaining the good performance compared to the formal likelihood function.

   As suggested, we changed case study 3 to simultaneously calibrate model parameters for discharge and three solutes to show that the adapted KGE is superior to the formal likelihood function for this type of model calibration. The adapted KGE approach improves the mean performance regarding Cl, NO_3 and SO_4 by 7%, 10% and 44% compared to the formal likelihood function using discharge and solutes and it has a slightly higher general performance for discharge (mean KGE is around 0.93). The adapted KGE also better represents the variability of the observations, especially for discharge, Cl and SO_4 where the variability metric (α) is centered around 1. The adapted KGE approach envelops most of the very high and very low solute concentration values (Cl and SO_4) in the total uncertainty band. This indicates that it can better represent the uncertainty when using multiple types of data for calibration such as discharge and three solutes in this case study.

   In addition, in case study 2 we compared the performance of the adapted KGE with GLUE, the formal likelihood function and the log-transformation. This shows that log-transformation works better for the low flow and the formal likelihood function has a better performance in high flow, while the adapted KGE combines advantages of the formal likelihood function and the log-transformation, leading to a good and balanced performance for low and high flows. The adapted KGE also has a higher general performance concerning the mean KGE of the evaluation and a better performance for variability. The efficiency and convergence rate are similar between the formal likelihood function and the adapted KGE for real-world calibrations that has different uncertainties.

   We have added the above discussion in the revised manuscript.

2. No theoretical analysis has been provided. At least one case where analytical form of posterior is available should be considered to verify whether the new likelihood can obtain the right answer.

   Response: We changed the generation of the virtual experiment in case study 1 to derive the pseudo-analytical posterior distribution of model parameters. Since the analytical posterior distribution of model parameters of a hydrological model is hardly achievable, we use the following procedure to obtain the pseudo-analytical posterior distribution. Firstly, we run the model with "true" parameters to obtain the simulated discharge. Secondly, assuming a normal distribution for error residuals (a common assumption for hydrological modeling), we generate random values from a normal distribution (mean=0, standard deviation=5% of the mean simulated discharge) and add *these random values as measurement errors to the simulated discharge to form the observations. Finally, due to no uncertainty in input data and model structure, using this setting for measurement errors we can use the formal likelihood function to derive the pseudo-analytical posterior distribution of model parameters. In case study 1, the adapted KGE approach (*KGE_{gamma}*) shows similar magnitude and shape as the pseudo-analytical posterior distributions of all model parameters derived from the formal likelihood function using the special virtual setting. This suggests that our adapted KGE approach can explore the right parameter posterior distributions. We have added the above discussion in the revised manuscript.*

3. The numbers of unknown parameters are generally small. A case with more than 20 unknown parameters (>100 would be better) is suggested to demonstrate its performance in more challenging settings.

   Response: Our approach was developed based on lumped or semi-distributed hydrological models, where the number of model parameters is usually smaller than 20 for which $DREAM_{(SZ)}$ *is common calibration tool (Liu et al., 2021; Shafii et al., 2014; Vrugt et al., 2008, 2009). Other new likelihood measures are also usually tested with simple analytical models or models with similar complexity as ours (Knoben et al., 2019; Schwemmle et al., 2020).*

4. Comparison with other informal likelihood functions (NSE, GLUE, etc.) is lacking.

   Response: We added the calibration using the GLUE method (NSE as the objective function) in case study 2. The results show that the adapted KGE approach performs better than GLUE. Additionally, as suggested in major comment 1 by reviewer #2 we also compared the performance between the adapted KGE and another formal likelihood function using the log-transformation. Generally, the adapted KGE approach has a good and balanced performance for both low and high flows. Please also refer to our reply to this major comment 1 and our reply to reviewer #2.

Minor comments

1. Lines 47-48: confused about what is N about.

   Response: N is the variable symbol that was used as a parameter. We reformulated it to "Freer et al. (1996) introduced a parameter as an exponent symbolled with N".

2. Lines 57-60: The proposal should not affect the shape of posterior if the chain is sufficiently long.

   Response: We agree that if the chain is long enough, the 'true' shape of posterior can be explored. However, in practice one needs to consider efficiency due to the computational cost. This means a limited number of realizations will be performed. Using the original KGE, the differentiation of very good (e.g. KGE=0.8) and good (e.g. KGE=0.6) in the standard MCMC is small. This will lead to a very fast convergence (indicating by the diagnostic index), which means using the limited realizations and its converged chains will result in a very flat posterior distribution, i.e. the exploration of the shape of posterior is largely affected. We added the explained condition "considering the computational cost with a limited number of realizations in practice" in the revised manuscript.

3. Line 82: if the types of observations are different and with different magnitudes, how to calculate the ED metric?

   Response: Since the adapted KGE is informal, we can combine multiple KGEs with each KGE for one type of observations (such as the weighted sum). The ED metric will be 1 subtracts the combined KGE. The combination of KGE will be based on the importance of each type of information defined by the user. It will be like using multi-objectives. For example, in case study 3, we used the equal weight for discharge and each solute to calculate the combined KGE.

4. There is no need to include results of KGE_ori, as they are obviously wrong.

   Response: Thanks for the suggestion. In revision, we only show KGE_ori in case study 1 to demonstrate the problems using the original KGE and focus on comparing the adapted KGE with different formal likelihood functions in case study 2 and 3.

5. Figures 6 (h-g), curves of KGE_ori and formal are quite different, why? A synthetic case with similar settings is needed to check which one failed to capture the truth.

   Response: We have removed KGE_ori as suggested. But differences of the posterior distributions of model parameters between formal, formal$_{log}$ and KGE$_{gamma}$ are due to the parameter interactions after checking the auto-correlation between model parameters. This is a common problem in hydrological modeling. Adding more information may further constrain the model parameter. As introduced in case study 3, we did the simultaneous calibration of discharge and solute concentrations. The adapted KGE approach outperforms the formal likelihood function, especially for solute concentrations. Please also refer to our reply to major comment 1.

6. Line 364: capable to->capable of

   Response: Revised.

7. What is equation of the likelihood function based on RMSE? There are also many forms of formal likelihood function (e.g., Table B1 in J.A. Vrugt / Environmental Modelling & Software 75 (2016) 273e316)

   Response: We added the equation in the revised manuscript. It is the first, "lik=11", in Table B1.

Reply to the comments of reviewer #2

General comments:

This study suggests an approach to adapt the KGE through transformation with a Gamma distribution so that it can be better used as an informal likelihood function in calibration procedures. The study finds that the results and inference behavior when using this adapted KGE measure are very similar to the case when using the RMSE as a likelihood function. In a synthetic case study, it is also shown that the presented approach successfully re-infers the known true parameter values.

The manuscript presents an elegant and innovative approach to a solution for a very relevant problem and could therefore be of high value in many fields. The manuscript is very well written, carefully composed and logically structured, and all in all very convincing.

We thank the review for the constructive comments and the positive evaluation of our work.

However, it is a bit too brief I feel in some respects and would need to be extended by some theoretical considerations among others (see comments below).

Major comments:

It is not clear to me what the “formal likelihood function” is in this case. The authors say that it is the RMSE, but it would be useful to show an equation that explicitly states which assumptions w.r.t. distribution type (I assume the normal distribution) and standard deviation this corresponds to. For example, something along the lines of: using the RMSE is equivalent to assuming independently normally distributed errors at each time step in a formal Bayesian inference approach and assuming that the standard deviation is equal to a certain (which?) value at each time step, ideally including the full equation. As is mentioned by the authors, the RMSE is very sensitive to large flows and would not be a typical measure used in formal likelihood approaches in my opinion. There are assumptions that usually work better, such as a standard deviation that is proportional to the predicted streamflow, for example. For a comprehensive overview of the different assumptions on the standard deviation of the residual error (and associate transformations) in formal likelihood approaches, see for example McInerny et al. (2017). In my view, it would make more sense to use one of their suggested approaches as the “formal” approach in this study.

Response: We provided the equation for calculating the formal likelihood (lik=11 in Table B1, Vrugt, 2016) in the revised manuscript. It assumes the error residuals are independent and normally distributed. Thanks for the very nice reference discussing different error models. As discussed by McInerney et al. (2017), there is no perfect error model that fits for all catchments and simultaneously optimizes all performance metrics (such as for both low flow and high flow). The log-transformation can work well for low flows. Therefore, we also added the comparison between the adapted KGE and the formal likelihood using the log-transformation in case study 2. Regarding the acceptance and convergence rates, the adapted KGE approach is similar as the formal likelihood function; both are better than the log-transformation in case study 2. Regarding the discharge calibration, the log-transformation generally works well for the low flow as expected and the formal likelihood functions works well for the high flow, while the adapted KGE approach have a good and balanced performance for both high and low flows.

In addition, we compared the adapted KGE with the GLUE method suggested in major comment 4 of reviewer #1. Please also refer to our reply to that for more details.

On a related note, the standard deviation of the additive error in formal likelihood approaches is an important parameter that needs to be used in prediction as well. The authors infer the posterior parameter distributions of the model parameters and then use these posteriors for prediction. This is fine if only parametric uncertainty is relevant, but by this, they completely neglect all other sources of uncertainty. The residual uncertainty (i.e., additive error) is very important since it represents the lumped effect of the input uncertainties, model structural uncertainties and observational uncertainties (present here at least in case study 3 as mentioned by the authors). The neglection of all these uncertainties is also the reason for the very narrow distribution of the performance metrics in prediction (Fig. 7 and 9). If actual streamflow predictions including error bands were shown, we would probably see that the observations are not covered at all by the error bands, which is a serious shortcoming if we are interested in reliable predictions.

Response: Thanks for the suggestion. We provided the total uncertainty in case study 1, and both parameter and total uncertainty in case study 3. As stated by the reviewer, many observations cannot be covered by the parameter uncertainty (Fig. S2) in case study 3, especially for solute concentrations. The total uncertainty of the formal likelihood and our adapted KGE approach are both much wider than the parameter uncertainty. But compared to the formal likelihood function, the total uncertainty derived from the adapted KGE shown in Fig. 9 can cover most of the observations including very low and high values, particularly for Cl and SO_4 concentrations. It demonstrates that the adapted KGE approach is advantageous for calibrations combining multiple types of observation data, especially when the data amount and unit are much different. Therefore, for prediction it may provide more reliable simulations.

Technical comments:

Line 44: It is not clear to me what you mean by “they can mimic the weight to small improvements in NSE”.

Response: We changed it to “they can mimic the weight such that small improvements in NSE can also be distinctly identified leading to the chain evolution”.

Line 55: did you mean “unsatisfactory”?

Response: Revised.

Line 55-57: I find this sentence incomprehensible

Response: We changed it to “Using NSE as the likelihood function, the number of measurements cannot be considered. Therefore, with increasing number of measurements, the information added to the performance measure is little, thus preventing the improvement of chain evolution”.

Line 60: “rates” instead of “rate”

Response: Revised.

Line 65: replace “theoretically statistical” with “statistically sound”, also in other instances if needed

Response: Revised.

References

Knoben, W. J. M., Freer, J. E. and Woods, R. A.: Technical note: Inherent benchmark or not? Comparing Nash-Sutcliffe and Kling-Gupta efficiency scores, Hydrol. Earth Syst. Sci., 23(10), 4323–4331, doi:10.5194/hess-23-4323-2019, 2019.

Liu, Y., Wagener, T. and Hartmann, A.: Assessing Streamflow Sensitivity to Precipitation Variability in Karst-Influenced Catchments With Unclosed Water Balances, Water Resour. Res., 57(1), doi:10.1029/2020WR028598, 2021.

McInerney, D., Thyer, M., Kavetski, D., Lerat, J. and Kuczera, G.: Improving probabilistic prediction of daily streamflow by identifying Pareto optimal approaches for modeling heteroscedastic residual errors, Water Resour. Res., 53(3), 2199–2239, doi:10.1002/2016WR019168, 2017.

Schwemmle, R., Demand, D. and Weiler, M.: Technical note: Diagnostic efficiency – specific evaluation of model performance, Hydrol. Earth Syst. Sci. Discuss., (2008), 1–15, doi:10.5194/hess-2020-237, 2020.

Shafii, M., Tolson, B. and Matott, L. S.: Uncertainty-based multi-criteria calibration of rainfall-runoff models: A comparative study, Stoch. Environ. Res. Risk Assess., 28(6), 1493–1510, doi:10.1007/s00477-014-0855-x, 2014.

Vrugt, J. A., ter Braak, C. J. F., Clark, M. P., Hyman, J. M. and Robinson, B. A.: Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with Markov chain Monte Carlo simulation, Water Resour. Res., 44(12), 1–15, doi:10.1029/2007wr006720, 2008.

Vrugt, J. A., ter Braak, C. J. F., Gupta, H. V. and Robinson, B. A.: Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches in hydrologic modeling?, Stoch. Environ. Res. Risk Assess., 23(7), 1011–1026, doi:10.1007/s00477-008-0274-y, 2009.