Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2019-327-RC1, 2019


Hydrology and Earth System Sciences
Discussions
Open Access
EGU



# *Interactive comment on* "Technical note: Inherent benchmark or not? Comparing Nash-Sutcliffe and Kling-Gupta efficiency scores" *by* Wouter J. M. Knoben et al.

**Hoshin Gupta (Referee)**

hoshin.gupta@hwr.arizona.edu

Received and published: 23 July 2019

Attached as a pdf file.

Please also note the supplement to this comment:
https://www.hydrol-earth-syst-sci-discuss.net/hess-2019-327/hess-2019-327-RC1-supplement.pdf

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2019-327, 2019.



Printer-friendly version

Discussion paper

**Review of HESS Technical note: "*Inherent benchmark or not? Comparing Nash- Sutcliffe and Kling-Gupta efficiency scores*", by Wouter J, M Knoben, JE Freer and RA Woods**

**Review Provided by Hoshin Gupta (23rd July 2019)**

**Summary of the Paper:** The paper makes perhaps three main points:

**Main Point Number (1): On Use of the "Mean Flow Benchmark" to interpret NSE and KGE**

- The NSE normalizes model performance to an interpretable scale such that NSE = 1 indicates perfect correspondence between simulations and observations, NSE = 0 indicates that the model simulations have the same explanatory power as the mean of the observations, and NSE < 0 indicates that the model is a worse predictor than the mean of the observations.
- NSE = 0 is regularly used as a benchmark to distinguish '*good*' and '*bad*' models, although this threshold could be considered a low level of predictive skill and is also a relatively arbitrary choice.
- KGE addresses several shortcomings in NSE and is increasingly used for model calibration and evaluation. Like NSE, KGE = 1 indicates perfect agreement between simulations and observations.
- Some users have tried to assign a similar scale/threshold as with NSE to be used in interpretation of KGE scores. Many authors use positive KGE values as indicative of '*good*' model performance, and negative KGE values as indicative of '*bad*' performance.
- However, this paper shows that placing the threshold for '*good*' model performance at KGE = 0 is generally correct (i.e., positive KGE values do indicate improvements upon the mean flow benchmark) but not complete. In fact, *negative KGE values do not necessarily indicate a model that performs worse than the mean flow benchmark*. The authors show this in mathematical terms, and then present results from a synthetic experiment to highlight that NSE and KGE values are not directly comparable and that understanding of the NSE metric does not translate well into understanding of the KGE metric.
- Mathematically, if the model simulations of the system responses are in fact constant over time and equal to the mean of the observed flows (the mean flow benchmark), we actually have KGE ≈ −0.41.

**Main Point Number (2): On the Need to Explicitly Consider Benchmark Performance**

- NSE and KGE values cannot be directly compared and should not be treated as approximately equivalent. There is no unique relationship between NSE and KGE values and where NSE values fall in the KGE component space depends in part on the coefficient of variation (CV) of the observations.
- NSE values that are traditionally seen as high do not necessarily translate into high KGE values. Hydrologists who choose to use the KGE metric need to develop new understanding of how this metric should be interpreted and not let themselves be guided by their understanding of NSE.

**Fig. 1.**