Final response

In the following, the specific Author Comments (ACs) previously provided to the Short Comments (SCs) and Referee comments (RCs) are given again (questions are in blue, authors' responses are in black). In addition, it is now specified, if relevant, where the modifications were implemented in the revised manuscript (version with no apparent modifications).

Finally, the current Final response also includes the revised version of the manuscript with apparent additions in blue and removals in red and crossed.

Answer to the review comments by Dr Lieke Melsen

We would like to thank Dr Lieke Melsen for her detailed analysis of the article. It helped us to clarify some points in the manuscript.

Major

- Page 6, line 3-4: "Furthermore, this result can be completed by making the same plot for other transformations giving more weight on low flows. Figure 4 shows that square rooted (Fig. 4 (a) and (b)) and inverse (Fig. 4 (c) and (d)) transformations do encounter the same problems as with the logarithm for catchments that have an average log-transformed flow around zero." This statement is inconsistent with the figures. The square root transformation does show a completely different pattern. Please clarify

  This statement is clearly inconsistent as we forgot a word. Instead of "do encounter", we meant "do not encounter". We apologize for this mistake that totally change the meaning of the sentence.

  However, we can discuss a little this point: Dr Lieke Melsen is right when saying that the square rooted flows show a completely different pattern but, to a lesser degree, it is also the case for the inverted flows. Indeed, if the KGE' on inverted values shows negative values for catchments that have an average log-transformed flow around zero, it also shows negative values for a significant part of the other catchments (Fig. 4 (c) and (d)). These negative values are more due to the difference between inverted flows and the untransformed flows than to some numerical flaws in the KGE'.

  Added/Modified: Sect. 5.1, p 6, lines 11-12 (of the revised manuscript) *Figure 3 shows that square rooted (Fig. 3 (a) and (b)) and inverse (Fig. 3 (c) and (d)) transformations do* ***not*** *encounter the same problems as with the logarithm for catchments that have an average log-transformed flow around zero.*

- Related to that; table 1 states that square root transformation does not increase low flow weight, but to me it seems that it diminishes the weight of high-flows, thereby somehow increasing the weight of low flows. Please clarify.

  The reviewer is right, by decreasing the high-flow weight, the square root transformation indirectly increases the low-flow weight. We stated this for the square root transformation in order to highlight the fact that this transformation increases low-flow weights to a lesser extent compared to the inverse, Box-Cox or logarithmic transformations.

  Instead of using two columns, namely about low and high flows columns, we propose to keep only one column named "Increases low-flow weight" and to use a different number of + signs as an intensity representation (+ for square root, ++ for logarithm, Box-Cox and the inverted square root, added following Dr John Ding comment and +++ for inverse).

Added/Modified: *We removed the column "Decrease high-flow weight" and used different numbers of + in the column "Increase low-flow weight" in table 1 (Sect. 6.2, top of p 12).*

- Page 8, line 3: I understand "optimistic" refers to a higher model performance for KGE' when evaluated in l/s compared to m3/s. However, I don't really understand how Eq. 5 automatically implies this. In Eq. 5 I see that log(1000) is always added, but whether this leads to an improved or decreased model performance seems to me dependent on the bias in the model. Please clarify.

  To clarify the impact of the added log(1000) we can calculate the bias ratio of log-transformed flows in $\mathrm{l \cdot s^{-1}}$ regarding the average of log-transformed flows in $\mathrm{m^3 \cdot s^{-1}}$. Using Eq. 5, the bias ratio is equal to:

$$\beta_{\log}[\mathrm{l \cdot s^{-1}}] = \frac{\log(1000) + \mu_{\log,\mathrm{s}}[\mathrm{m^3 \cdot s^{-1}}]}{\log(1000) + \mu_{\log,\mathrm{o}}[\mathrm{m^3 \cdot s^{-1}}]} \tag{1}$$

  In the tested data set, $\mu_{\log,\mathrm{s}}[\mathrm{m^3 \cdot s^{-1}}]$ and $\mu_{\log,\mathrm{o}}[\mathrm{m^3 \cdot s^{-1}}]$ (the log-transformed flows averages of respectively the simulated and observed flows in $\mathrm{m^3 \cdot s^{-1}}$) are in majority between -4 and 4. Because log(1000) is higher than the flows averages (≈ 6.9), it will have a greater impact on the ratio calculation than the average flow itself which leads to a tendency to improve the ratio. As we use the KGE', the γ ratio is also affected and, because of the interaction between the average and the standard deviation of flows it is even more complicated to predict the ratio difference between $\mathrm{m^3 \cdot s^{-1}}$ and $\mathrm{l \cdot s^{-1}}$.

  To illustrate this on the data set used in the article we plotted the values of the three KGE' components for the log-transformed flows in $\mathrm{m^3 \cdot s^{-1}}$ and $\mathrm{l \cdot s^{-1}}$ (Fig. 1 of this answer). Fig. 1 (c) shows that the bias ratio tends to be improved in $\mathrm{l \cdot s^{-1}}$ especially for the catchments that have a bad bias ratio in $\mathrm{m^3 \cdot s^{-1}}$. The difference between the two flow units is more complicated in the case of the coefficient of variation ratios (Fig. 1 (b)).

  In a nutshell, the KGE' value tends to be higher because of the artificial improvement of the bias ratio but the coefficients of variation ratio can vary differently and lead to a decrease of the KGE' value.

  In the manuscript, we will replace "optimistic" by "higher model performance" and add some sentences to better explain the reasons of this apparent improvement of performances.

  Added/Modified: Sect. 5.2, p 8, lines 8-13 *The **higher model performance** when using $l \cdot s^{-1}$ than when using $m^3 \cdot s^{-1}$ can be explained analytically. Considering Eq. 7, the formula of the bias ratio in $l \cdot s^{-1}$ regarding the averages in $m^3 \cdot s^{-1}$ is:*

$$\beta_{\log}[\mathrm{l \cdot s^{-1}}] = \frac{\log(1000) + \mu_{\log,\mathrm{s}}[\mathrm{m^3 \cdot s^{-1}}]}{\log(1000) + \mu_{\log,\mathrm{o}}[\mathrm{m^3 \cdot s^{-1}}]}$$

  Because log(1000) *is not negligible compared to the averages, adding this constant term would artificially improve β and, by extension, the KGE' value. The γ ratio* *is also affected and, due to the interactions between the standard deviation and the averages, modify differently the KGE' value.*

Minor

- Page 4, line 13 It would be good if the order of Box-Cox and adding a constant is changed in order to be consistent with results.

  It will be done.

  Added/Modified: *Done (see Sect. 3.3, p 4, lines 13-19)*

- Figure 2 is relevant and insightful, but it takes some time to understand all information. Perhaps, it can be stressed in the caption that left, simulated in shown in red and right observed is shown in red (as is also done for a figure later in the manuscript).

  Regarding this remark, we propose to replace " The red dots represent the catchments where the average of the log-transformed simulated (a) or observed (b) flows is around 0." by "In plot (a), the axis values represents the observed log-transformed flows averages and the color represents the simulated ones while in the plot (b), it is the opposite."

  Added/Modified: *We did this modification not only in Fig. 2 caption but also in the captions of the three other figures that are similar to Fig. 2 (namely Fig. 3, 7 and 9).*

- Page 6, line 7 "remain correct". Correct seems a vague term in this context (what is a correct objective function value?). Please consider rewording.

  The reviewer is right, the word "correct" is not well chosen, particularly because some of the NSE values in question are around zero which denotes a bad simulation. It will be replaced by "positive or around zero".

  Added/Modified: Sect. 5.1, p 7, lines 1-2 *This tends to confirm that the strongly negative KGE' values stem more from a numerical issue than an actual problem in simulated values because the NSE values in these catchments remain* ***positive or around zero****.*

- Consider to include the original KGE equation in Section 2 as well, especially because this information is relevant in the discussion of the modified Box-Cox. E.g. p. 9 l. 20, it will not affect the KGE because μ_s is not in the denominator in the original KGE (perhaps help the reader on this as well, e.g. on p. 10 just above the section Summary).

  It is a good suggestion, an equation will be added replacing the γ term of Eq. 1 by an α. The E_{KG} in Eq. 1 will be denoted E'_{KG} and proper reference to the KGE equation in page 9 and 10.

  Added/Modified: Sect. 2, p 2, line 20 *The KGE and KGE' criteria (Gupta et al., 2009; Kling et al., 2012, respectively denoted* E_{KG} *and* E'_{KG} *in Eq. 1 and Eq. 2)*

  Sect. 2, p 2, lines 23-24

$$E_{\mathrm{KG}} = 1 - \sqrt{(r-1)^2 + (\beta-1)^2 + (\alpha-1)^2}$$

$$E'_{\mathrm{KG}} = 1 - \sqrt{(r-1)^2 + (\beta-1)^2 + (\gamma-1)^2}$$

Sect. 2, p 3, lines 3-4 – *α, the ratio between the simulated and observed standard deviations evaluates the flows variability error:*

$$\alpha = \frac{\sigma_{\mathrm{s}}}{\sigma_{\mathrm{o}}}$$

Sect. 5.4, p 10, lines 10-11 *Because this instability is due to μ_{s} (which is only in the denominator of the γ ratio in Eq. 6), it will only affect the KGE'. The KGE is not affected because an α ratio is used instead of the γ ratio (Eq. 1 and 5)*

- Page 3, line 19 conversation -> conversion

  It will be fixed

  Added/Modified: Sect. 3.2, p 3, lines 23-24 *It can be easily demonstrated that γ, β and r remain identical when flow is expressed in any of these two units, since the division by 1000 necessary for the* ***conversion*** *is eliminated in the ratios.*

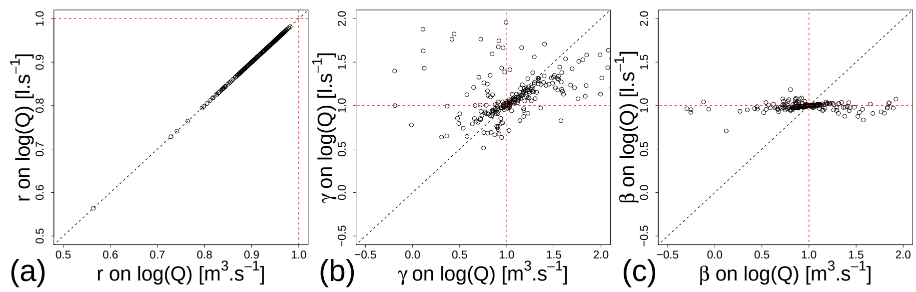

Figure 1: Values difference between cubic metres and litres per seconds of the three components of the KGE' calculated on log-transformed flows.

Answer to the review comments by Dr Björn Guse

We would like to thank Dr Björn Guse for his review of the article.

Minor comments:

- Page 5, line 28: The modelling period is subdivided into a calibration (2003-2008) and a validation period (2008-2013). Please make clear whether the year 2008 belongs to the calibration or validation period.

  Actually we made a mistake in this description. To be exact, the calibration period covers the period from July 2005 to June 2009 and the validation one covers the period from July 2009 to July 2013. We apologize for this mistake and will modify the article accordingly.

  Added/Modified: Sect. 4.1, p 5, lines 1-2 *The availability of data covers the **2005-2013** period.*

  Sect. 4.2, p 5, lines 6-8 *The available records are split into a calibration **(from July 2005 to June 2009)** and a validation **(from July 2009 to July 2013)** period following a standard split-sample test procedure (Klemeš, 1986).*

- Figs. 2-11: If possible, I recommend to add a header to the subplots. Since all these figures have a similar layout, it would be good to differentiate them in a clearer way. E.g. you may add "observations" and "simulations" as header to the subplots of Fig. 3.

  This is a good suggestion, it will help making the figures more understandable and complements Dr Lieke Melsen's minor comment number 2.

  Added/Modified: *We added these headers in Fig. 2, 3, 7 and 9.*

- Fig. 2-3: Maybe you can merge both figures by adding the information that you compare log-transformed and untransformed data.

  This is also a good idea as Fig. 2 has more value when compared to Fig. 3. We will do it.

  Added/Modified: *Done, we removed the former Fig.3 and added it to Fig.2 (c) and (d).*

Technical comments:

- I recommend to avoid paragraphs with only one or two sentences such as on P.4, L.20-27.

  We will try to aggregate the smallest paragraphs for a better understanding of the manuscript.

  Added/Modified: *Following this advice, we aggregated 4 couples of paragraphs. These aggregations are pictured by small red lines in the following marked-up manuscript version.*

- Page 13, L. 4: Please remove “pp.” in the Coron et al.

  It will be done

  Added/Modified: *Done*

Answer to the comments by S. Mylevaganam

We would like to thank S. Mylevaganam for his comments on the technical aspects of our manuscript and the Associate Editor for framing our answer regarding the issues which are out of the scope of the technical discussion.

- 1) From the reader's point of view, considering the current discussion, the current version of the manuscript needs to be reviewed by qualified referees who are specialized in the subject that is presented in the manuscript. Since this is a technical note, the need for more technical evaluation by the referees is required. Moreover, the referees' role should not be harbored to state whether or not the referees "like/hate/love" the manuscript.

  This comment is not about the paper content and we therefore refer to the Editor's reply.

- 2) As per the authors [see P-2 LN-2], Gupta et al. (2009) clearly demonstrated that discharge variability is not correctly taken into account for the evaluation. Therefore, Gupta et al. (2009) proposed a new criterion, Kling-Gupta efficiency (KGE), which provides direct assessment of four aspects of discharge time series, namely shape and timing, water balance and variability[see P-2 LN-6]. As far as I remember, in 2006, this piece of idea was introduced by a graduate student. Therefore, respecting Mr. Donald Trump's intention of preventing people from stealing someone's ideas/works/technologies, it would be more appropriate for the authors to evaluate the originality of Gupta et al. (2009)'s work.

  The objective of the paper was to discuss a specific application of a criterion previously published in an international journal after peer-review (Journal of Hydrology) and which has been widely used then. If there is a possible concern about the KGE criterion, the reviewer should directly contact the authors or the editors of Journal of Hydrology. We agree with the Editor that this discussion and our article are not the right place to discuss this issue. We are not aware of the graduate student work mentioned by the reviewer, for which no detail is given and therefore we are not able to evaluate the originality of the work.

- 3) As per the authors, The KGE' criterion (Kling et al., 2012, denoted EKG in Eq. 1) is written as a sum of the distances to 1 (perfect value) of three components of the modelling error [see P-2 LN-20]. What is meant by "sum" of the "distances" to 1? What is the mathematical formula that is used to compute the distance? If we consider a three dimensional space (i.e., x-axis=ratio-1, y-axis=ratio-2, z-axis=ratio-3), isn't the square root component merely the distance from the origin (i.e., [1, 1, 1])?

  These questions show that our sentence is not precise enough. Mathematically speaking, the KGE' is a linear transformation of the Euclidian distance from the ideal point (i.e., [1, 1, 1]) in the three-dimensional space defined by the three ratios (Eq. (2) to (4)). In Eq. (1), this Euclidian distance is represented by the square root component and the computed linear transformation of this distance is $f : x \mapsto 1 - x$. This function is used to allow the KGE' to have the same range of values as for the NSE. We thank the reviewer for pointing out this and will attempt to be more precise in the manuscript by modifying the page 2 line 20 sentence.

Added/Modified: Sect. 2, p 2, lines 20-22 *The KGE and KGE' criteria (Gupta et al., 2009; Kling et al., 2012, respectively denoted* E_{KG} *and* E'_{KG} *in Eq. 1 and Eq. 2) are written as a linear transformation (*$f : x \mapsto 1 - x$*) of the Euclidian distance to an ideal value (i.e. [1,1,1]) in a three dimensional space defined by three components of the modelling error:*

- 4) What is the physical meaning of equation (1)? Let's say that the right-hand side of the equation (1) has two components. The first component is "1". The second component is the square root component that includes the ratios (e.g., beta). Why would you subtract the second component (i.e., square root component) from the first component (i.e., 1)? What is the physical meaning of the second component? What is the physical meaning of the first component? If the second component of the equation (1) represents the distance (see the definition), as per dimensional theories, the first component needs to be a distance. Otherwise, the operator (i.e., negative sign) becomes meaningless. What is the distance represented by the first component? What is the origin for the distance that is represented by the first component?

  First of all, regarding the dimensional theories, the KGE' expression is right. Indeed, a Euclidian distance in a space of dimensions without units (it is the case of the three ratios that form the KGE') is dimensionless. Thus, linearly transforming this dimensionless Euclidian distance is not wrong mathematically speaking.

  However, the choice of the transformation f can be discussed. As said in the answer of the reviewer comment 3), the distance is subtracted to 1 to have the same range of values as the NSE criterion. This is clearly due to legacy because a lot of rainfall-runoff modellers are used to the NSE and to analyse its values. This subtraction can be discussed as the Euclidian distance stands for itself as an evaluation criterion but, because the transformation of this distance in the KGE' is linear, the interpretation of the KGE' values remains the same as for the Euclidian distance. Consequently, it has no impact on the evaluation of the performance of the model.

  Regarding the physical meaning of the KGE', we will answer in our response of comment 6).

- 5) Does your equation (3) evaluate the water balance error? What is meant by water balance? What is the range of your beta value? Assume that we have the following monthly observed flow values :5,5,4,5,5,5,5,5,6,5,5,5. Assume that we have the following monthly simulated flow values :5,5,6,5,5,5,5,5,5,5,5,4.As per your equation(3), if we consider all the flow values, the beta value is 1.However, if we consider the first six months, the value of the beta is not equal to 1. Is your beta value time dependent?

  The beta ratio represents the quantitative aspect of the simulation. If it is greater than 1, the model overestimates the discharge and if it is lower than 1, the model underestimate the discharge. We agree that the value of beta depends on the time as it is the case of all the other ratios. We used a split-sample test to limit the impact of this time dependency.

To avoid misunderstanding we will replace "water balance" by "bias".

Added/Modified: Sect.2, p 3, line 1 – *β, the bias term, evaluates the* ***bias between observed and simulated flows****:*

- 6) Assume that the ratio-1=1 (i.e., equation (2)), ratio-2=1 (i.e., equation (3)), and ratio-3=0.5 (i.e., equation (4)). As per your equation (1), the value of KGE is 0.5. Now, assume that the ratio-1=1 (i.e., equation (2)), ratio-2=0.5 (i.e., equation (3)), and ratio-3=1 (i.e., equation (4)). As per your equation (1), the value of KGE is 0.5. What is the physical meaning of the KGE values?

  The physical meaning of the KGE value itself is not well defined. It is simply an aggregated representation of the model error over the studied period. To understand its value, the modeller needs to have a look on the three components of the criterion separately. This is stated in the Gupta et al. (2009) publication and it is often done by the KGE users (for example in the work of Ficchí et al., 2016, cited in the present manuscript). Moreover, depending of the modeller's objectives, Gupta et al. (2009) also proposed to weight each component of the KGE.

- 7) As per your equation (4), the ratio-3 is a function of your beta value. In other words, your ratio-3 is a function of ratio-2 (i.e., equation (3)). This gives an indication that the ratio-3 that is accounted in your equation (1) repeats the influence of ratio-2 in equation (1).

  In the publication that introduces the KGE', Kling et al. (2012) stated that: "For the variability ratio γ we used $\frac{CV_s}{CV_o}$ instead of $\frac{\sigma_s}{\sigma_o}$, which was proposed in the original version of the KGE-statistic (Gupta et al., 2009). This ensures that the bias and variability ratios are not cross-correlated, which otherwise may occur when e.g. the precipitation inputs are biased.". In other words if the bias ratio is bad, the ratio of standard deviation will also be affected. To avoid the impact of average discharge error on the variability component, the standard deviation ratio is normalised by the bias ratio.

  Added/Modified: Sect.2, p 3, line 6 *These coefficients of variation are used to avoid the impact of bias on the variability indicator (Kling et al., 2012):*

Answer to the comments by Dr John Ding

On Table 1: A fifth transformation of the low flow (comment SC1)

A combination of the first two transformations, the square root and the inverse, leads back in time to the 1960s to a RoSR (reciprocal of the square root) transformation of the dry weather flow. This physics-based double transformation was arrived at independently by Chapman (1964) in Australia, Ishihara and Tagaki (1965) in Japan, and this writer, Ding (1966) in Canada. All this, along with the fourth statistical one, the 1-parameter Box-Cox (1964) transformation, appeared in a 3-year time span from 1964 to 1966. I'll be interested in the authors' view on this parametric-free RoSR transformation.

A graphical illustration (comment SC2)

It's most gratifying to read the positive response from the authors to my suggestion of looking at a classical RoSR (i.e., ISR, the inverse square root) transformation. Their discussion paper has appeared at a most opportune time, as I have ready a slide presentation on the Budyko evapotranspiration framework. This happens to include an illustration of the log- and RoSR-transformations of a dry weather flow hydrograph. Two of these slides are shown here, Figure 9 and (Section) 7. In the latter, the storage exponent N appears in a nonlinear storage-discharge relation, $Q \approx S^N$.

Addendum (comment SC3)

Unbeknown to me until now, Figure 9 hints at a new data transformation targeting hydrographs. This is called the NISR, the negative inverse square-root transformation of the flow $Q : -\frac{1}{\sqrt{Q}}$.

Re-ranking the transformations (comment SC5)

The NISR (negative inverse square-root) transformation can be generalized to a NIR (negative inverse root) one defined below:

$$J_N(Q) = -\frac{1}{\sqrt[N]{Q}} = -Q^{\frac{-1}{N}} \tag{2}$$

$$J_0(Q) = \log Q \tag{3}$$

Some may dismiss the J_2 transformation as simply a sign change of the classical ISR (inverse square-root) one. Indeed they are correct. But as Leonardo Da Vince (1452-1519) once said, "Simplicity is the ultimate sophistication."

For example, J_2 happens to be a subset of the 1-parameter Box-Cox (1964) transformation (Eq. 6 in Santos et al. paper) when parameter $\lambda = -\frac{1}{2}$

$$f_{BC}^{\lambda=-\frac{1}{2}}(Q) = \frac{Q^\lambda - 1}{\lambda} = 2(1 + J_2) \tag{4}$$

The difference between these two ISR-type transformed values is:

$$f_{BC}^{\lambda=-\frac{1}{2}} - J_2 = 2 + J_2 = 2 - \frac{1}{\sqrt{Q}} \tag{5}$$

This has a maximum value of 2.

Figure 1 shows the modifications of the four transformation methods being considered in the authors' Table 1. These are the original logarithmic, a fixed-parameter Box-Cox, the inverse negated, and the square root both inverted and negated, and labelled J_0, $f_{BC}^{\lambda=-\frac{1}{2}}$, J_1 and J_2, respectively. All four transformation curves share the same inverted U-shape, being an advantage for comparison purposes. These show their relative impact on the transformed flow values, most obviously on the lower ends.

Answer to Dr J. Ding comment SC1

We would like to thank Dr J. Ding for reading our manuscript and for his comment.

We had not heard before about this transformation which is, actually, very interesting.

Regarding only Table 1, the inverted square root transformation shows exactly the same pros and con as the inverse transformation. It allows to decrease high-flow weight and increase low-flow weight in the KGE' calculation. A KGE' calculated on this transformation is also dimensionless and shows no issue when the flow average is around 1 (see Fig. 2 of this comment) and, as for inverse transformation, the inverted square root one needs specific attention for zero flows.

However, if we only consider numerical characteristics, the inverted square root transformation presents two advantages compared to invert transformation. The first one is that, even if it is sensitive to the constant added to avoid zero flows, this sensitivity is lower than the inverse transformation's sensitivity (as shown in Fig. 3). The second one is that the inverse transformation can be very extreme and totally erase the weight of high flows. The inverted square root can be seen as "smoother" than inverse.

In a nutshell, we consider the inverted square root transformation as a good compromise to replace logarithm transformation. We are grateful to Dr J. Ding for his suggestion and we propose to add this transformation in Table 1 and to add comments in Sect. 6.2. Obviously, we will acknowledge his contribution in the text.

Answer to Dr J. Ding comments SC2, SC3 and SC5

First of all, we would like to thank again Dr J. Ding for his valuable contribution to the discussion.

- Regarding John Ding's SC3 comment and the use of negative or positive inverted root we can argue that, in the context of this manuscript, the change of sign has no impact. Indeed, following the example of the inverted flows ($-J_1$ in comment SC5), the mean and the standard deviation are linked to the ones of J_1 by the following relations, respectively:

$$\begin{aligned} \mu_{-Q^{-1}} &= -\mu_{Q^{-1}} \\ \sigma_{-Q^{-1}} &= \sigma_{Q^{-1}} \end{aligned} \tag{6}$$

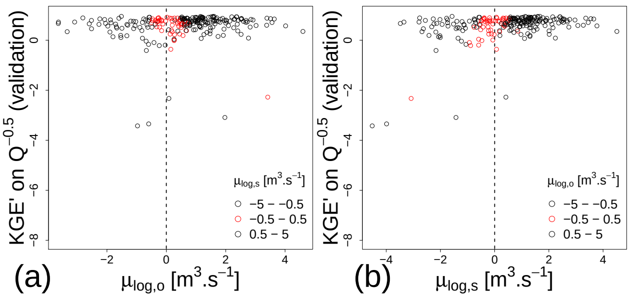

Figure 2: Values of KGE' on inverse square rooted transformed flows versus the mean of the log-transformed observed (a) and simulated (b) flows. Each dot represents the performance obtained in validation for one catchment after calibration with the KGE' on untransformed flows as an objective function. The red dots represent the catchments where the average of the log-transformed simulated (a) or observed (b) flows is around 0.

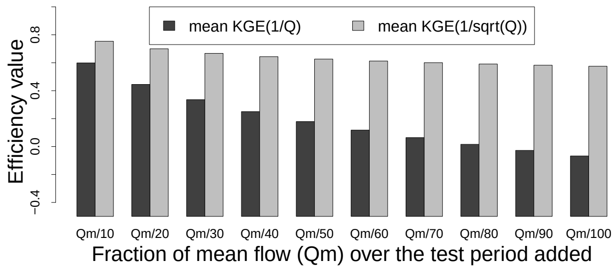

Figure 3: Sensitivity of KGE' to the fraction of average flows that is added to flows to avoid zero flows in the inverted square root transformation for 240 catchments over the validation period.

with $\mu_{Q^{-1}}$ and $\sigma_{Q^{-1}}$ respectively the mean and the standard deviation of inverted flows and $\mu_{-Q^{-1}}$ and $\sigma_{-Q^{-1}}$ respectively the mean and the standard deviation of negative inverted flows.

The consequence of Eq. 6 in this comment, using Eq. 2 to 4 in the manuscript is that:

$$\begin{array}{rcl} r_{-Q^{-1}} & = & r_{Q^{-1}} \\ \beta_{-Q^{-1}} & = & \beta_{Q^{-1}} \\ \gamma_{-Q^{-1}} & = & \gamma_{Q^{-1}} \end{array} \qquad (7)$$

and, using Eq.1 of the manuscript:

$$KGE'(-Q^{-1}) = KGE'(Q^{-1}) \qquad (8)$$

The sign of the transformation has, thus, no importance in the KGE' (and also KGE) calculation. For this reason we will keep positive transformations in our manuscript.

We show the equivalence of KGE' values for both aforementioned transformations on our data set in Fig. 4 of this answer to Dr Ding's comment.

- The choice of N in the transformations J_N proposed by Dr J. Ding can be interesting. One may choose the N value according to the weight intended on low flows. The higher N, the lower the weight on low flows. In addition, regarding Dr. J. Ding comment SC2, the value of N can also be deduced from the observation of recession curve in the simulated catchment.

- Regarding comment SC5, we will add this generic inverted root transformation in table 1 of the manuscript instead of the inverted square root as we stated it in comment AC1. We will also add a comment on its parametrization in the text.

- About the correspondence between the inverted root transformation and the Box-Cox transformation, Dr J. Ding is right arguing that, if $\lambda = -\frac{1}{N}$, a linear relation links the Box-Cox transformation and the inverted root transformation. However, to obtain this linear relation, λ has to be negative and, as much as we know, hydrologists who use the Box-Cox transformation always use a positive λ value because it allows to avoid issues with zero flows. As a consequence, we will keep the Box-Cox transformation in the table as it is.

PS: As an answer of SC2 comment sentence “For example, why the Box-Cox (1964) transformation has gone mainstream in hydrology, but the ISR (1964-66) has not, as if some of us hadn’t tried or hard enough.”, we can hypothesize that the greater interest for Box-Cox is due to its property to avoid the zero-flow issue. However, it is also possible that the use of Box-Cox is also due to legacy as it is the case for NSE criterion.

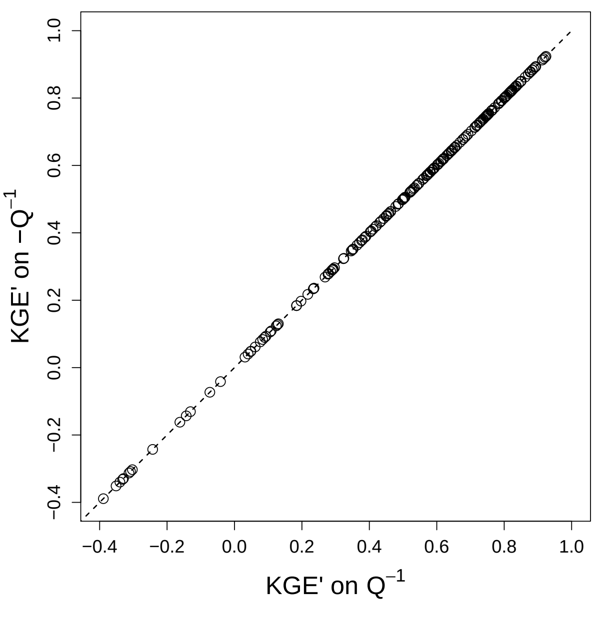

Figure 4: Equality of the KGE’ values of the GR4J simulations using inverse transformation and negative inverse transformation on the 240 tested catchments.

Added/Modified

Sect. 6.2, p 11, lines 3-7 *The inverted root is an example of used transformation that is not tested in the article but leads to increase the weight of low flows (Chapman,* *1964; Ishihara and Takagi, 1965; Ding 1966). It can be parametrised with the value of the power in the root ($Q^{-\frac{1}{N}}$). Depending on the value of N, there will be more or less weight on low flows. The higher N is and the less the weight on low flows is. This N value can also be determined with the recession curves of observed flows.*

We add a line in Table 1 (top of p 12) to describe the pros and cons of the parametric inverted root proposed by Dr Ding.

Technical note: Pitfalls in using log-transformed flows within the KGE criterion

Léonard Santos, Guillaume Thirel, and Charles Perrin

Irstea, HYCAR Research Unit, 1 rue Pierre-Gilles de Gennes, 92160 Antony, France

Correspondence to: Léonard Santos (leonard.santos@irstea.fr)

Abstract. Log-transformed discharge is often used to calculate performance criteria to better focus on low flows. This prior transformation limits the heteroscedasticity of model residuals and was largely applied in criteria based on squared residuals, like Nash-Sutcliffe efficiency (NSE). In the recent years, NSE has been shown to have mathematical limitations and Kling-Gupta efficiency (KGE) was proposed as an alternative to provide more balance between the expected qualities of a model (namely representing the water balance, flow variability and correlation). As in the case of NSE, several authors used the KGE criterion (or its improved version KGE') with a prior logarithmic transformation on flows. However, we show that the use of this transformation is not adapted to the case of the KGE (or KGE') criterion and may lead to several numerical issues, potentially resulting in a biased evaluation of model performance. We present the theoretical underpinning aspects of these issues and concrete modelling examples, showing that KGE' computed on log-transformed flows should be avoided. Alternatives are discussed.

1 Introduction

In the context of rainfall-runoff modelling, evaluating the quality of the models' outputs is essential. Deterministic simulations are commonly evaluated using efficiency criteria such as Nash-Sutcliffe efficiency (NSE, Nash and Sutcliffe, 1970). The choice of the criteria obviously depends on the modeller's objective. For example, one may wish to focus on the overall water balance evaluation, or more specifically on the simulation of different flow ranges, typically high, intermediate or low flows. For these different objectives, given that the model residuals are generally not homoscedastic and often depend on the flow magnitude, one common option to focus more closely on specific flow ranges is to apply various prior transformations on the simulated and observed discharge time series to distort the range of errors, which consequently changes the relative weight of different flow ranges in the criterion. This is commonly done within the NSE criterion, which has been one of the most popular criteria used in hydrological modelling in the past few decades. NSE is the distance to 1 of the ratio between the mean square error of the model and the variance of observed flows. Compared to the basic criterion computed on untransformed flows, a prior squared transformation on flows would put even more weight on high flows, a logarithmic or inverse transformation would put more weight on low flows while a square root transformation would have an intermediate effect (Krause et al., 2005; Oudin et al., 2006; De Vos and Rientjes, 2010; Pushpalatha et al., 2012).

However, the Nash-Sutcliffe criterion was shown to have limitations. Indeed, using a decomposition of NSE based on the correlation, bias and ratio of variances, Gupta et al. (2009) clearly demonstrated that discharge variability is not correctly taken into account for the evaluation. Therefore, Gupta et al. (2009) proposed a new criterion, Kling-Gupta efficiency (KGE), which was then improved into a modified criterion called KGE' (Kling et al., 2012). KGE combines the previous components of NSE (correlation, bias, ratio of variances or coefficients of variation) in a more balanced way. It corrects the underestimation of variability and provides direct assessment of four aspects of discharge time series, namely shape and timing, water balance and variability.

Given that this criterion tends to be sensitive to large errors, some users chose to apply prior transformations on flows before computing KGE, e.g. to put more weight on low flows, as done with NSE. For example, Pechlivanidis et al. (2014) applied the logarithmic transformation to use it as a benchmark for fitting a model on low flows. Seeger and Weiler (2014) used it as an objective function. Beck et al. (2016) used the untransformed and log-transformed flows in NSE, R^2 and KGE as an evaluation of different global models, and Quesada-Montano et al. (2018) also used it as an evaluation criterion of the HBV model outputs.

In this technical note we show that the use of a logarithmic transformation when computing KGE or KGE', applied in a similar way as with NSE, introduces numerical flaws and should be avoided. After reviewing the mathematical formulation of KGE', we expose the theoretical aspects explaining these flaws and illustrate them with modelling examples. Then we suggest alternatives to circumvent this issue. The tests will be carried out using KGE' but they are also valid for the initial KGE formulation.

2 The KGE and KGE' ~~formulation~~formulations

The KGE ~~' criterion (Kling et al., 2012, denoted E_{KG} in Eq. 1) is~~ and KGE' criteria (Gupta et al., 2009; Kling et al., 2012, respectively denoted E_{KG} and E'_{KG} in Eq. 1 and Eq. 2) are written as a ~~sum of the distances to~~ linear transformation ($f: x \mapsto 1-x$) of the Euclidian distance to an ideal value (i.e. [1~~(perfect value) of~~ ,1,1]) in a three dimensional space defined by three components of the modelling error:

$$E_{\mathrm{KG}} = 1 - \cancel{\sqrt{(r-1)^2+(\beta-1)^2+(\gamma-1)^2}}\sqrt{(r-1)^2+(\beta-1)^2+(\alpha-1)^2} \quad (1)$$

$$E'_{\mathrm{KG}} = 1 - \sqrt{(r-1)^2+(\beta-1)^2+(\gamma-1)^2} \quad (2)$$

in which:

- r, the Pearson correlation coefficient, evaluates the error on shape and timing between observed (Q_{o}) and simulated (Q_{s}) flows:

$$r = \frac{\mathrm{cov}(Q_{\mathrm{o}}, Q_{\mathrm{s}})}{\sigma_{\mathrm{o}}^2 \sigma_{\mathrm{s}}^2} \quad (3)$$

- β, the bias term, evaluates the ~~water balance error~~ bias between observed and simulated flows:

$$\beta = \frac{\mu_s}{\mu_o} \tag{4}$$

- α, the ratio between the simulated and observed standard deviations evaluates the flows variability error:

$$\alpha = \frac{\sigma_s}{\sigma_o} \tag{5}$$

- γ, the ratio between the simulated and observed coefficients of variation (CV) also evaluates the flows variability error. These coefficients of variation are used to avoid the impact of bias on the variability indicator (Kling et al., 2012) :

$$\gamma = \frac{\mu_o \sigma_s}{\sigma_o \mu_s} \tag{6}$$

where cov is the covariance between observation and simulation, μ is the mean and σ is the standard deviation, with subscripts o and s standing for observed and simulated, respectively.

The KGE' values range between $-\infty$ and 1, as for NSE, and it is positively oriented.

3 Issues associated with the use of a prior logarithmic transformation

3.1 Instability when the moments of log-transformed flows become close to zero

Because the three terms, γ, β and r are ratios, they can become overly sensitive to the denominator values (here μ_o, μ_s, σ_o or σ_s) if they become close to zero. In this case, a small absolute variation in the moments' values can negatively impact the related ratio and thus produce very negative KGE' values. It is generally very unlikely to obtain values of σ_o, σ_s, μ_s, μ_o so close to zero to produce numerical instability when using untransformed flows. However, when a prior logarithmic transformation is applied, the values of $\mu_{\log,o}$ or $\mu_{\log,s}$ (more rarely $\sigma_{\log,o}$ or $\sigma_{\log,s}$) computed on transformed values can become equal or very close to zero (because $\log(1) = 0$). The corresponding ratios r, β or γ would therefore become very large, leading to strongly negative KGE' values. Thus a small relative difference can lead to very different conclusions. In this case, the score value does not adequately represent the qualities of the model simulation.

3.2 Dependence on the flow unit chosen

KGE' and NSE criteria are dimensionless. This means that using discharge values expressed in litres per second or in cubic metres per second has no impact on the criteria values. It can be easily demonstrated that γ, β and r remain identical when flow is expressed in any of these two units, since the division by 1000 necessary for the ~~conversation~~ conversion is eliminated in the ratios. ~~–~~ When using a prior logarithmic transformation, the NSE criterion is not affected because the squared differences of flows eliminates the multiplicative conversion coefficients in the mean square error (numerator) or in the variance (denominator). However, the KGE' calculation is altered through the β ratio. Using the example of the average observed flow calculation, the conversion from cubic metres per second to litres per second gives the following:

$$\mu_{\log,o}[\mathrm{l} \cdot \mathrm{s}^{-1}] = \log(1000) + \mu_{\log,o}[\mathrm{m}^3 \cdot \mathrm{s}^{-1}] \tag{7}$$

Consequently, because the conversion term becomes additive when applying the logarithmic transformation, the β ratio value is modified. Similarly, the γ ratio is also altered. ~~–~~ Therefore, if the logarithmic transformation is used, the KGE' (and also the KGE) is no longer a dimensionless value. This can lead to interpretation problems.

3.3 Dependence on the constant added to avoid the zero-flow issue

When using a logarithmic (or an inverse) transformation, the case of null flows, which may exist in case of intermittent or ephemeral streams, prevents proper calculation. To avoid this, different techniques may be set up in the case of NSE:

- discarding the zero-flow values from the series, i.e. considering them as gaps (see e.g. Nguyen and Dietrich, 2018). The drawback is that parts of the hydrographs become neglected, though they can bring important information on the processes at play.
- ~~using a Box-Cox transformation to reproduce the effects of the logarithmic transformation without the zero-flow issue (Box and Cox, 1964; Hogue et al., 2000; Vázquez et al., 2008) .~~ adding a small constant to all flow values (Pushpalatha et al., 2012), typically a fraction of average flow. This option is widely used and Pushpalatha et al. (2012) showed that the NSE value has limited sensitivity to this constant with a logarithmic transformation as long as it is small enough compared to flow values. These authors advise a constant equal to one-hundredth of the mean observed flows. But the dependence of KGE' on this constant has not been investigated so far.
- using a Box-Cox transformation to reproduce the effects of the logarithmic transformation without the zero-flow issue (Box and Cox, 1964; Hogue et al., 2000; Vázquez et al., 2008) .

4 Testing methodology

To illustrate these numerical issues and their potential impacts, several tests were made on a wide range of catchments, using the GR4J rainfall-runoff model (Perrin et al., 2003).

4.1 Catchment set and data

A daily data set of 240 catchments across France (Fig. 1), set up by Ficchí et al. (2016), was used. The climate data of the SAFRAN daily reanalysis (Vidal et al., 2010) were used as input data. Precipitation and temperature were spatially aggregated on each catchment since the GR4J model is lumped. Potential evapotranspiration was calculated using a temperature-based

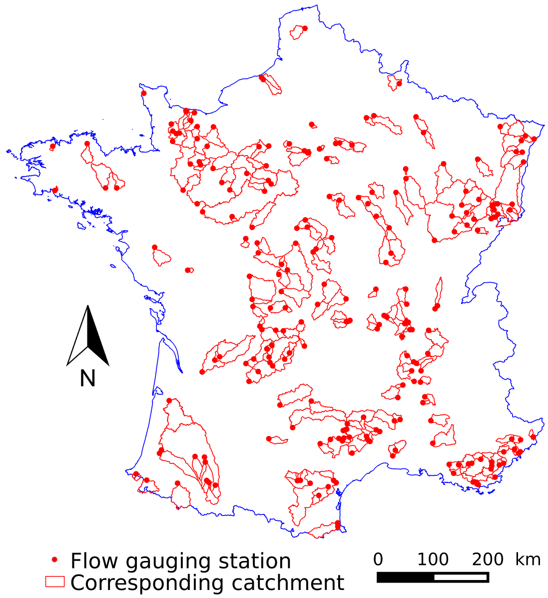

Figure 1. Location of the 240 flow gauging stations in France used for the tests and their associated catchments.

formula (Oudin et al., 2005). Full details on this data set are available in Ficchí et al. (2016). Observed flows were retrieved for each catchment outlet from the *Banque HYDRO* (http://www.hydro.eaufrance.fr/, Leleu et al., 2014). The availability of data covers the ~~2003-2013 period.~~ 2005-2013 period. The catchments were selected to have less than 10% of precipitation falling as snow, to avoid requiring a snow model.

4.2 Model and calibration

The tests were performed with the daily lumped conceptual GR4J model (Perrin et al., 2003). The four parameters of the model are calibrated using the local search optimization algorithm used in Coron et al. (2017). The available records are split into a calibration (~~2003-2008~~from July 2005 to June 2009) and a validation (~~2008-2013~~from July 2009 to July 2013) period following a standard split-sample test procedure (Klemeš, 1986). The calibration procedure was run using the KGE' on untransformed flows as an objective function. The performance of the model is then evaluated during the validation period using KGE' on untransformed and log-transformed flows. The performance is also calculated using different transformations that can substitute the logarithmic transformation, namely the square-rooted flows, the inverted flows and the Box-Cox transformed flows. The NSE criterion is also calculated on log-transformed flows to be compared to KGE' using the same transformation. The zero flows were treated following the conclusions of Pushpalatha et al. (2012), i.e. by adding to flows a constant equal to one-hundredth of the mean observed flows. The parameter of the Box-Cox transformation is fixed at the value of 0.25, as Vázquez et al. (2008) argue that it is an usual value in hydrological studies.

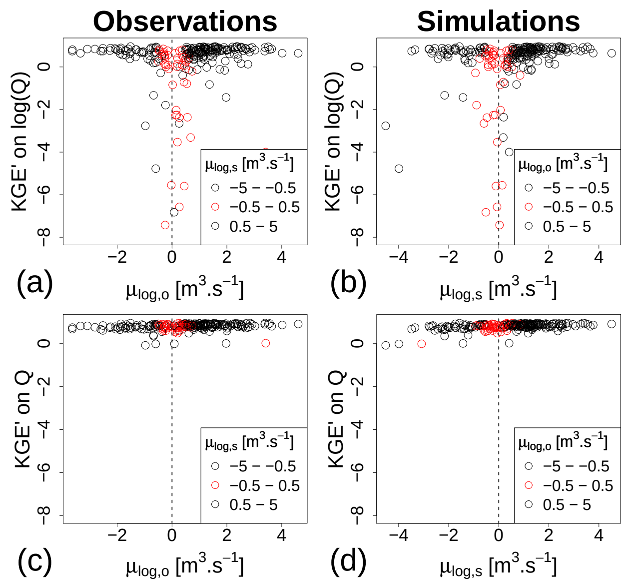

Figure 2. Values of KGE’ on ~~untransformed~~ log-transformed flows ((a) and (b)) versus the mean of the log-transformed observed ~~(~~and simulated flows compared. As a benchmark, the same plots are drawn with untransformed flows ((c) and ~~simulated~~ (~~b~~d))~~flows~~. Each dot represents the performance obtained in validation for one catchment after ~~calibrating~~ calibration with the KGE’ on untransformed flows as an objective function. ~~The red dots represent the catchments where~~ In plots (a) and (c), the ~~average of~~ axis values represent the observed log-transformed flow averages and the color represents the simulated ones while in plots (~~a~~b) ~~or observed~~ and (~~b~~d) ~~flows~~ it is ~~around 0.~~the opposite.

5 Results

5.1 Instability when the moments of log-transformed flows become close to zero

Figure 2 (a) and (b) analyses the stability of the KGE’ values with log-transformed flows obtained in the validation period. The KGE’ values were plotted against the mean of the log-transformed observed (a) and simulated (b) flows. When any of these means tends to be close to zero, the KGE’ criterion exhibits unusually low values. This plot illustrates the problem identified in section 3.1. These very negative values may alter model evaluation. When working on a large set of catchments, they may also bias the calculation of the mean performance over the catchment set, by heavily weighting these outlier values. ~~Figure 3~~ Figure 2 (c) and (d) shows that the catchments with negative KGE’ values in Fig. 2 (a) and (b) do not seem to exhibit any specific behaviour when evaluated with the KGE’ values on untransformed flows: the criterion values are not lower in these catchments than in other catchments. Furthermore, this result can be completed by making the same plot for other transformations giving more weight on low flows. Figure 3 shows that square rooted (Fig. 3 (a) and (b)) and inverse (Fig. 3 (c) and (d)) transformations do not encounter the same problems as with the logarithm for catchments that have an average log-transformed flow around zero.

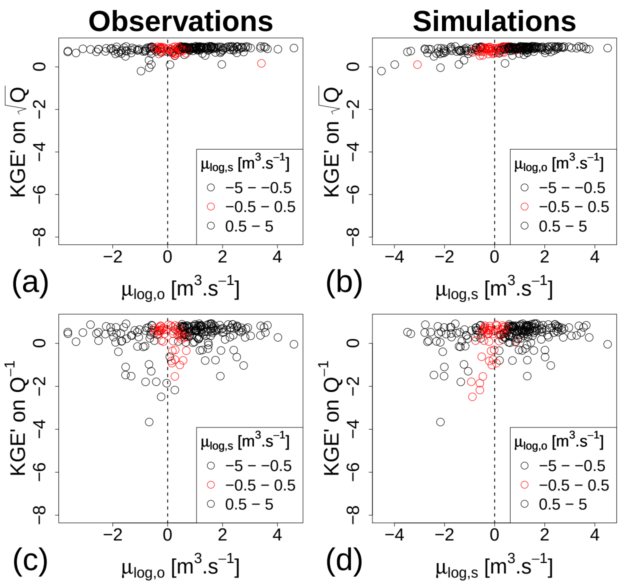

Figure 3. Values of KGE' on square root ((a) and (b)) and inverse ((c) and (d)) transformed flows versus the mean of the log-transformed observed and simulated flows. Each dot represents the performance obtained in validation for one catchment after calibration with the KGE' on untransformed flows as an objective function. ~~The red dots represent the catchments where the average of the log-transformed simulated (~~In plots (a) and (c)~~) or~~ , the axis values represent the observed ~~(~~log-transformed flow averages and the color represents the simulated ones while in plots (b) and (d) ~~) flows~~ it is ~~around 0.~~the opposite.

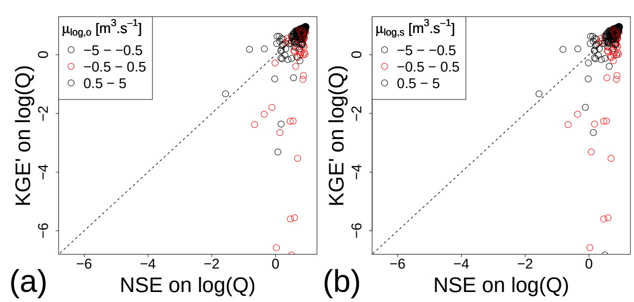

Figure 4. Comparison between KGE' and NSE values on the validation period using a calibration with KGE' on untransformed flows as an objective function. The red dots represent the catchments where the average of log-transformed observed (a) or simulated (b) flows is around 0.

The KGE' on log-transformed flows can also be compared to the NSE using the same transformation. Figure 4 shows that, when KGE' is significantly lower than NSE, the average of log-transformed flows (observed or simulated) is around zero (red dots in the figure). This tends to confirm that the strongly negative KGE' values stem more from a numerical issue than an actual problem in simulated values because the NSE values in these catchments remain ~~correct~~positive or around zero.

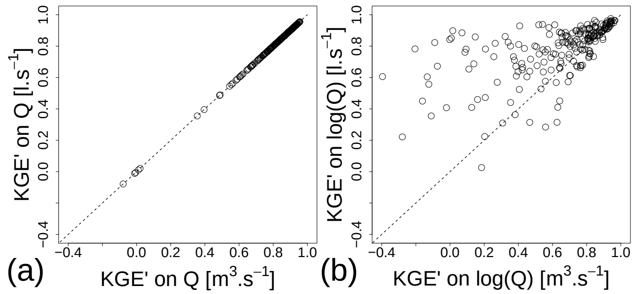

Figure 5. Dependence on flow units of the KGE' using untransformed flows (a) and log-transformed flows (b) on the 240 catchments. The parameters used for simulation evaluation were obtained by calibrating GR4J using KGE' on untransformed flows.

In this technical note, the impact of a near-zero standard deviation of log-transformed flows is not presented because it is rarer than near-zero mean values. The standard deviations of flows on the catchments studied are indeed all significantly higher than zero.

5.2 Dependence on the flow unit chosen

The dependence of KGE' on log-transformed flows on the chosen flow units can easily be shown by plotting the KGE' on log-transformed flows in cubic metres per second versus the KGE' on log-transformed flows in litres per second. Figure 5 (b) shows that, for the catchments tested, the values of KGE' on log-transformed flows clearly depend on the flow unit used. A more optimistic evaluation of model performance will generally be obtained with the flows in $\mathrm{l \cdot s^{-1}}$. As a comparison, Fig. 5 (a) shows that the KGE' with untransformed flows is not affected by the flow unit change. This dimension dependence makes the KGE' values based on log-transformed flows very difficult to interpret.

The ~~more optimistic results~~ higher model performance when using $\mathrm{l \cdot s^{-1}}$ than when using $\mathrm{m^3 \cdot s^{-1}}$ can be explained analytically. Considering Eq. 7, ~~a value of log(1000) is added to the~~ the formula of the bias ratio in $\mathrm{l \cdot s^{-1}}$ regarding the averages in $\mathrm{m^3 \cdot s^{-1}}$ ~~average.~~ is:

$$\beta_{\log}[\mathrm{l \cdot s^{-1}}] = \frac{\log(1000) + \mu_{\log,s}[\mathrm{m^3 \cdot s^{-1}}]}{\log(1000) + \mu_{\log,o}[\mathrm{m^3 \cdot s^{-1}}]} \qquad (8)$$

Because $\log(1000)$ is not negligible compared to the averages, adding ~~it~~ this constant term would artificially improve β ~~and γ~~ and, by extension, the KGE' value. The γ ratio is also affected and, due to the interactions between the standard deviation and the averages, modify differently the KGE' value.

5.3 Dependence on the value added to avoid the zero-flow issue

Pushpalatha et al. (2012) showed that the sensitivity of the NSE criterion on log-transformed flows to the small added constant declines when this constant decreases (from one-tenth to one-hundredth of the mean observed flow) and becomes limited for

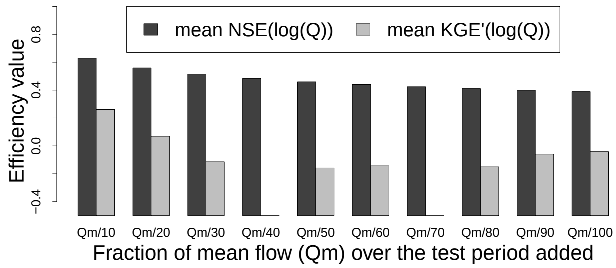

Figure 6. Sensitivity of NSE and KGE' to the fraction of average flows that is added to flows to avoid zero flows in the logarithmic transformation for 240 catchments over the validation period. This graph is inspired by Fig. 9 in Pushpalatha et al. (2012).

very small values (see Fig. 9 in Pushpalatha et al., 2012). We performed the same test with the KGE' criterion and we obtained a very different result (Fig. 6). The impact on performance is erratic for different values added to flows and does not show any trend. This may be due to the numerical issues shown in Sec. 5.1. For these reasons, the impact of added values can be major and may alter the model evaluation.

5.4 The case of the Box-Cox transformation

As presented in Sect. 3.3, instead of adding a small value to flows, a Box-Cox transformation can be applied to flows to mimic the logarithm transformation without the zero-flow problem. However, even though it removes the dependence of the KGE' value to the value added to avoid zero flows, the other issues presented in the previous sections exist as for the logarithm. For catchments in which the log-transformed flows' average is close to zero, the Box-Cox transformed flows exhibit the same behaviour as with the logarithm (Fig. 7). This result is logical because the Box-Cox transformation of 1 is equal to 0, as for the logarithmic transformation.

The Box-Cox transformation is also dependent on the units (Fig. 8 (a)). However, for this last issue, a slight modification of the Box-Cox formula allows one to address this problem. The classical Box-Cox transformation can be written as:

$$\mathrm{f_{BC}}(Q) = \frac{Q^{\lambda} - 1}{\lambda} \tag{9}$$

in which λ is an exponent to be chosen by the user, Q is the flow value for any unit and $\mathrm{f_{BC}}$ is the Box-Cox function.

Using this equation, the KGE' on transformed flows will be unit-dependent because of the additive term 1 in the numerator. To avoid this, we can slightly modify the formula, by replacing the term 1 by a constant with a unit dependence (here we propose the hundredth of the mean flow) and by putting it to the power λ:

$$\mathrm{f'_{BC}}(Q) = \frac{Q^{\lambda} - (0.01\mu_{\mathrm{o}})^{\lambda}}{\lambda} \tag{10}$$

Using Eq. 10, the KGE' criterion remains dimensionless using the Box-Cox transformation (Fig. 8 (b)).

Furthermore, because the zero of the modified Box-Cox function is not 1 any more, this transformation would reduce the issue of strongly negative values when $\mu_{\mathrm{log,o}}$ or $\mu_{\mathrm{log,s}}$ are around zero. However, there still is an issue if the average of simulated

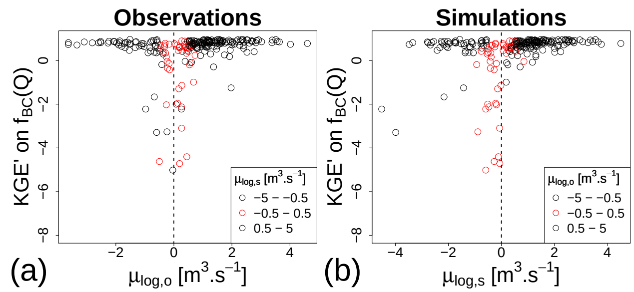

Figure 7. Values of KGE' on Box-Cox transformed flows versus the mean of the log-transformed observed (a) and simulated (b) flows. Each dot represents the performance obtained in validation for one catchment after calibration with the KGE' on untransformed flows as an objective function. ~~The red dots represent the catchments where the average of the log-transformed simulated~~ In plot (a)~~or~~ , the axis values represent the observed log-transformed flow averages and the color represents the simulated ones while in plot (b) ~~flows~~ it is ~~around 0.~~the opposite.

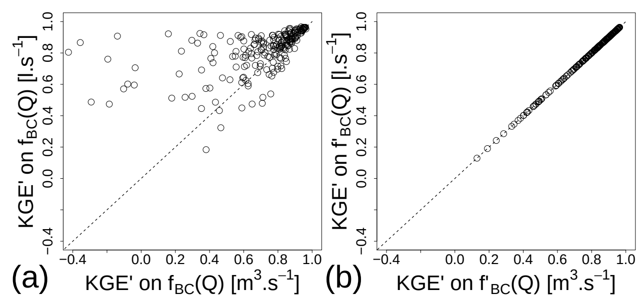

Figure 8. Dependence on flow units of the KGE' using Box-Cox transformed flows without adaptation ((a), Eq. 9) and with adaptation ((b), Eq. 10) on the 240 catchments. The parameters used for simulation evaluation were obtained by calibrating GR4J using KGE' on untransformed flows.

flows is around the zero of the modified Box-Cox function (i.e. if $\cancel{\mu_s = (0.01 * \mu_o)^\lambda}$ $\mu_s = (0.01\mu_o)^\lambda$, Fig. 9). This instability occurs more rarely than for the logarithm transformation but can be more frequent if bigger percentage of the average of observed flow or different λ value are used. Because this instability is due to μ_s (which is only in the denominator of the γ ratio in Eq. 6), it will only affect the KGE'~~(not the KGE~~. The KGE is not affected because an α ratio is used instead of the γ ratio (Eq. 1 and 5).

The modified Box-Cox transformation (Eq. 10) allows to avoid unit dependence and to reduce the instability issues due to the values of average flows (especially when using the KGE). The behaviour of this modified transformation also remains similar to the one of the initial Box-Cox transformation except when $\mu_{log,o}$ or $\mu_{log,s}$ are around zero (Fig. 10).

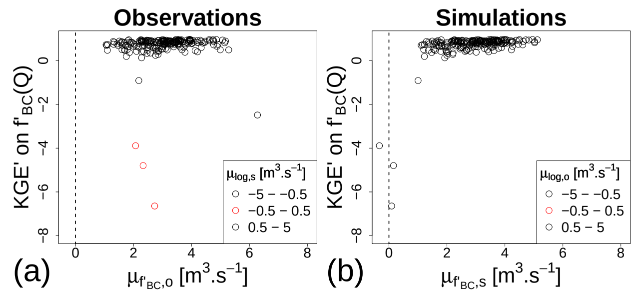

Figure 9. Values of KGE’ on modified Box-Cox transformed flows (Eq. 10) versus the mean of this transformed observed (a) and simulated (b) flows. Each dot represents the performance obtained in validation for one catchment after calibration with the KGE’ on untransformed flows as an objective function. ~~The red dots represent the catchments where the average of the transformed simulated~~ In plots (a)~~or~~, the axis values represent the observed transformed flow averages and the color represents the simulated ones while in plot (b) ~~flows~~ it is ~~around 0.~~ the opposite.

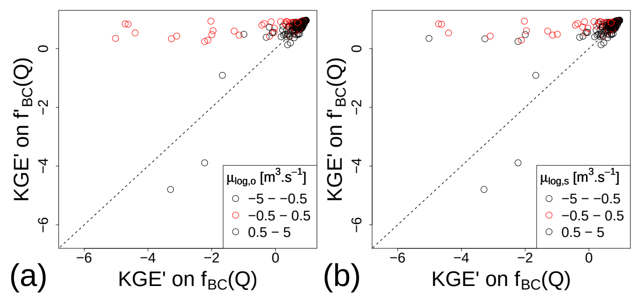

Figure 10. Comparison between KGE’ values on Box-Cox and modified Box-Cox transformed flows on the validation period using a calibration with KGE’ on untransformed flows as an objective function. The red dots represent the catchments where the average of log-transformed observed (a) or simulated (b) flows is around 0.

6 Summary

6.1 Log transformation should not be used in the KGE or KGE’ criterion

Given the previous results, we can argue that using log-transformed flows to calculate the KGE or the KGE’ criterion can lead to difficulties in the interpretation of criterion values. The criterion does not remain dimensionless like NSE with a prior logarithmic transformation. It also becomes overly sensitive when the log-transformed flows’ average becomes close to zero, yielding potentially very negative values, or when a small constant is added to flows prior to logarithmic transformation to cope with zero flows. Because of all these issues, logarithmic transformation should be avoided when using KGE’.

Table 1. Pros (+) and cons (-) of different flow transformations to improve consideration of low flows in KGE'. In the second column, the number of (+) represents the intensity of low-flow weight increase. There are parentheses around the last + for inverted root and Box-Cox transformations because the low-flow weight depends on parameters.

Flow transformation	~~Decrease high-flow weight~~	Increase low flow weight	No issue with zero flows	Dimensionless	No issue when flows average around 1
Square root	~~+~~	~~-~~ +	+	+	
Inverse	~~+~~	+++	-	+	+
Inverted root		++(+)	-	+	+
Logarithm	~~+~~	++	-	-	-
Box-Cox	~~+~~	+ (+)	+	+ (if using Eq. 10)	+ (if using Eq. 10)

6.2 Alternatives

Instead of KGE' on log-transformed flows, several transformations can be used to calculate KGE'. The pros and cons for several transformations are summarised in Table 1. The inverted root is an example of used transformation that is not tested in the article but leads to increase the weight of low flows (Chapman, 1964; Ishihara and Takagi, 1965; Ding, 1966) . It can be parametrised with the value of the power in the root ($Q^{-\frac{1}{N}}$). Depending on the value of N, there will be more or less weight on low flows. The higher N is and the less the weight on low flows is. This N value can also be determined with the recession curves of observed flows. Regarding this table, the modified Box-Cox transformation (Eq. 10) seems to be the best solution but it still faces instabilities for some flow average values (for the KGE'). Thus, there is no ideal solution to avoid all problems. Modellers have to make a choice depending on their specific applications. According to the intensity of low-flow weight increase that is needed, the choice of transformation has to be adapted. Garcia et al. (2016), for example, recommend averaging two KGE' criteria computed on untransformed and inverted flows, into a composite criterion.

Note that many studies use NSE on log-transformed flows (see for example Lyon et al., 2017; Nguyen and Dietrich, 2018). Fortunately, the mathematical formulation of NSE avoids all the problematic aspects identified for KGE with the logarithmic transformation. However, this may not be a sufficient argument to continue to use NSE given the issues presented by Gupta et al. (2009) and Schaefli and Gupta (2007):

- the underestimation of variability,
- the low weight of water balance errors for catchments with highly variable flows,
- the poor benchmark represented by the mean flows for catchments with highly variable flows.

6.3 Final remarks

Two additional remarks should be taken into account on this topic. First, as noted by H. Kling in a personal communication, prior transformations on flows in KGE (or in NSE) lead to a misinterpretation in the estimation of the water balance. The other components of the KGE also lose their initial physical meaning. KGE on transformed flows can give more information on low flows, but the physical interpretation of the criterion is not as simple as in the case of untransformed flows.

Secondly, even if it did not occur in our experiment, the issue described in this technical note may lead to problems during the calibration process. Indeed, it can create a strongly negative zone in the objective function hyperspace, which may negatively impact the performance of local calibration algorithms.

Author contributions. L. Santos made the technical development and the analysis. The manuscript was written by him, G. Thirel and C. Perrin.

Competing interests. The authors declare that they have no conflicts of interest.

Acknowledgements. The authors thank Météo France for providing the data used in this work. We also wish to thank Alban De Lavenne, Laure Lebecherel, Maria-Helena Ramos and Cedric Rebolho for the discussions on the different aspects of the issues using the logarithmic transformation with KGE. We thank Andrea Ficchí for his work on the database and Linda Northrup for her correction of the English language of the manuscript. Finally, we extend our thanks to Harald Kling for discussions on this issue.

We thank the topical editor, Bettina Schaefli, for her careful reading of the manuscript, her suggestion on the modified Box-Cox transformation and to follow up the discussions. We also thank the two reviewers, Lieke Melsen and Björn Guse, for taking the time to read our manuscript and for their remarks that helped us to make the manuscript and the figures more understandable. We thank Sivarajah Mylevaganam for the discussions that helped us to be more precise in the KGE and KGE' description. Finally, we particularly want to thank John Ding for his suggestion to add the RoSR transformation (that we did not know about before) to the article and for the fruitful discussions that followed.

References

Beck, H. E., van Dijk, A. I. J. M., de Roo, A., Miralles, D. G., McVicar, T. R., Schellekens, J., and Bruijnzeel, L. A.: Global-scale regionalization of hydrologic model parameters, Water Resour. Res., 52, 3599–3622, doi:10.1002/2015WR018247, 2016.

Box, G. E. P. and Cox, D. R.: An Analysis of Transformations, Journal of the Royal Statistical Society. Series B (Methodological), 26, 211–252, http://www.jstor.org/stable/2984418, 1964.

Chapman, T. G.: Effects of groud-water storage and flow on the water balance, in: Proceedings of "Water resources, use and management", pp. 291–301, Australian Academy of Science, Melbourne Univ. Press, 1964.

Coron, L., Thirel, G., Delaigue, O., Perrin, C., and Andréassian, V.: The suite of lumped GR hydrological models in an R package, Environmental Modelling & Software, 94, 166–177, doi:10.1016/j.envsoft.2017.05.002, 2017.

De Vos, N. J. and Rientjes, T. H. M.: Multi-objective performance comparison of an artificial neural network and a conceptual rainfall-runoff model, Hydrological Sciences Journal, 52, 397–413, doi:10.1623/hysj.52.3.397, 2010.

Ding, J. Y.: Discussion of "Inflow hhydrograph from large unconfined aaquifers" by Ibrahim, H. A. and Brutsaert, W. J., J. Irrig. Drain. Am. Soc. Civ. Eng., 92, 104–107, 1966.

Ficchí, A., Perrin, C., and Andréassian, V.: Impact of temporal resolution of inputs on hydrological model performance: An analysis based on 2400 flood events, Journal of Hydrology, 538, 454–470, doi:10.1016/j.jhydrol.2016.04.016, 2016.

Garcia, F., Folton, N., and Oudin, L.: Which objective function to calibrate rainfall–runoff models for low-flow index simulations?, Hydrological Sciences Journal, 62, 1149–1166, doi:10.1080/02626667.2017.1308511, 2016.

Gupta, H. V., Kling, H., Yilmaz, K. K., and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, Journal of Hydrology, 377, 80–91, doi:10.1016/j.jhydrol.2009.08.003, 2009.

Hogue, T. S., Sorooshian, S., Gupta, H., Holz, A., and Braatz, D.: A Multistep Automatic Calibration Scheme for River Forecasting Models, Journal of Hydrometeorology, 1, 524–542, doi:10.1175/1525-7541(2000)001<0524:AMACSF>2.0.CO;2, 2000.

Ishihara, T. and Takagi, F.: A study on the variation of low flow, Bulletin of the Disaster Prevention Research Institute, 15(2), 75–98, http://hdl.handle.net/2433/124698, 1965.

Klemeš, V.: Operational testing of hydrological simulation models, Hydrological Sciences Journal, 31, 13–24, doi:10.1080/02626668609491024, 1986.

Kling, H., Fuchs, M., and Paulin, M.: Runoff conditions in the upper Danube basin under ensemble of climate change scenarios, Journal of Hydrology, 424–425, 264–277, doi:10.1016/j.jhydrol.2012.01.011, 2012.

Krause, P., Boyle, D. P., and Bäse, F.: Comparison of different efficiency criteria for hydrological model assessment, Advances in Geosciences, 5, 89–97, doi:10.5194/adgeo-5-89-2005, 2005.

Leleu, I., Tonnelier, I., Puechberty, R., Gouin, P., Viquendi, I., Cobos, L., Foray, A., Baillon, M., and Ndima, P.-O.: Re-founding the national information system designed to manage and give access to hydrometric data, La Houille Blanche, 1, 25–32, doi:10.1051/lhb/2014004, in French, 2014.

Lyon, S. W., King, K., Polpanich, O., and Lacombe, G.: Assessing hydrologic changes across the Lower Mekong Basin, Journal of Hydrology: Reginal Studies, 12, 303–314, doi:10.1016/j.ejrh.2017.06.007, 2017.

Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models. Part I - A discussion of principles, Journal of Hydrology, 10, 282–290, doi:10.1016/0022-1694(70)90255-6, 1970.

Nguyen, V. T. and Dietrich, J.: Modification of the SWAT model to simulate regional groundwater flow using a multicell aquifer, Hydrological Processes, 32, 939–953, doi:10.1002/hyp.11466, 2018.

Oudin, L., Hervieu, F., Michel, C., Perrin, C., Andréassian, V., Anctil, F., and Loumagne, C.: Which potential evapotranspiration input for a lumped rainfall–runoff model?, Journal of Hydrology, 303, 290–306, doi:10.1016/j.jhydrol.2004.08.026, 2005.

Oudin, L., Andréassian, V., Mathevet, T., Perrin, C., and Michel, C.: Dynamic averaging of rainfall-runoff model simulations from complementary model parameterizations, Water Resour. Res., 42, doi:10.1029/2005wr004636, 2006.

Pechlivanidis, I. G., Jackson, B., McMillan, H., and Gupta, H.: Use of an entropy-based metric in multiobjective calibration to improve model performance, Water Resour. Res., 50, 8066–8083, doi:10.1002/2013WR014537, 2014.

Perrin, C., Michel, C., and Andréassian, V.: Improvement of a parsimonious model for streamflow simulation, Journal of Hydrology, 279, 275–289, doi:10.1016/s0022-1694(03)00225-7, 2003.

Pushpalatha, R., Perrin, C., Moine, N. L., and Andréassian, V.: A review of efficiency criteria suitable for evaluating low-flow simulations, Journal of Hydrology, 420-421, 171–182, doi:10.1016/j.jhydrol.2011.11.055, 2012.

Quesada-Montano, B., Westerberg, I. K., Fuentes-Andino, D., Hidalgo, H. G., and Halldin, S.: Can climate variability information constrain a hydrological model for an ungauged Costa Rican catchment?, Hydrological Processes, 32, 830–846, doi:10.1002/hyp.11460, 2018.

Schaefli, B. and Gupta, H. V.: Do Nash and values have value?, Hydrological Processes, 21, 2075–2080, doi:10.1002/hyp.6825, 2007.

Seeger, S. and Weiler, M.: Reevaluation of transit time distributions and mean transit times and their relation to catchment topography, Hydrol. Earth Syst. Sci., 18, 4751–4771, doi:10.5194/hess-18-4751-2014, 2014.

Vázquez, R. F., Willems, P., and Feyen, J.: Improving the predictions of a MIKE SHE catchment-scale application by using a multi-criteria approach, Hydrological Processes, 22, 2159–2179, doi:10.1002/hyp.6815, 2008.

Vidal, J.-P., Martin, E., Franchisteguy, L., Baillon, M., and Soubeyroux, J.-M.: A 50-year and high-resolution atmospheric reanalysis over and France with the Safran system, International Journal of Climatology, 30, 1627–1644, doi:10.1002/joc.2003, 2010.