Hydrol. Earth Syst. Sci. Discuss., 10, C6371–C6384, 2013
www.hydrol-earth-syst-sci-discuss.net/10/C6371/2013/

Hydrology and Earth System Sciences
Discussions

Interactive comment on "On the lack of robustness of hydrologic models regarding water balance simulation – a diagnostic approach on 20 mountainous catchments using three models of increasing complexity" *by* L. Coron et al.

L. Coron et al.

laurent.coron@irstea.fr

Received and published: 19 November 2013

Answer to Referee 2

NB. The referee comments have been repeated here and are written inside < < > > symbols.

———

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

< < Main point 1) Page 11345, line 4: The way year is defined in this study is not clear. There is a reference to 'hydrological year' but it is not clear to the Reviewer whether the Authors used water or calendar year [. . .] > >

We used hydrological years, which were in this study defined as starting on October 1st and ending on September 30th. As suggested by the reviewer, we added this information in the paper (see section 3.1). We also modified the x-axis in figures 4 and 5, to add the months.

———

< < Main point 2) Page 11349 Lines 9-10: when explaining the meaning of $\sigma[\omega_{\theta_{SP[i]}} - \omega_{\theta_{TP}}]$, the Authors state that 'only the shape similarities of the ω_θ curves are analysed and their vertical spacing is left out of consideration'. The Reviewer thinks this sentence is misleading because it can imply that two curves with identical shapes to $\omega_{\theta_{TP}}$ but at different distances from it have the same value for $\sigma[\omega_{\theta_{SP[i]}} - \omega_{\theta_{TP}}]$. This clearly cannot be true as the numerator of equation (5) (inside the sum), which measures the distance/vertical spacing, will be different. Moreover, this has implications for the meaning of ρ_i , which is presented as a measure of the 'degree of "parallelism" relative to the magnitude of bias variations'. Once again, this suggests that two perfectly parallel curves are identical in terms of this criterion. These performance measures are a crucial part of the study and therefore their meaning should be clarified. > >

The referee is correct about mentioning an issue here. This problem was also noted by referee 1 and the answer provided here can therefore be also found in our reply to referee 1.

The error lays in the formula written but not in the computations which are reported in the paper.

The computations reported in the paper correspond to the standard deviation, as it is mentioned in the text and shown using notations such as $\sigma[\omega_{\theta_{TP}}]$. However, the formula

HESSD

10, C6371–C6384, 2013

presented in equations (4) and (5) were wrong, as they do not reflect standard deviation computations (the ratio 1/p was misplaced and more importantly, "minus mean" component was missing as noted by referee 1).

This mistake explains why the Referee was misled. If the formula actually used were those given in equations (4) and (5), we would agree with Referee 2 when he/she says "[. . .] it can imply that two curves with identical shapes to $\omega_{\theta_{TP}}$ but at different distances from it have the same value [. . .] (which) cannot be true as the numerator of equation (5) [. . .] will be different. This has implications for the meaning of ρ_i [. . .]". The corrected version of the standard deviation can be written: $\sigma[u] = \sqrt{\frac{1}{p}\sum_{k=1}^{p}(u_k - \overline{u})^2}$

The "minus mean" component ensures the optimal situation at zero for all $\omega_{\theta_{TP}}$ curves. This is why two curves with exactly identical shapes but at different vertical positioning have the same value, and thus, why we are allowed to compare ρ_i values.

The corrected versions of the equations are the following:

$$\sigma\left[\omega_{\theta_{TP}}\right] = \sqrt{\left(\frac{1}{p}\sum_{k=1}^{p}(u_k)^2\right) - \left(\frac{1}{p}\sum_{k=1}^{p}(u_k)\right)^2} \quad ; \quad u_k = \frac{\left[\overline{\widehat{Q}_{SP[k]}}\right]_{\theta_{TP}}}{\overline{Q_{SP[k]}}} \tag{4}$$

$$\sigma\left[\omega_{\theta_{SP[i]}} - \omega_{\theta_{TP}}\right] = \sqrt{\left(\frac{1}{p}\sum_{k=1}^{p}(v_k)^2\right) - \left(\frac{1}{p}\sum_{k=1}^{p}(v_k)\right)^2} \quad ; \quad v_k = \frac{\left[\overline{\widehat{Q}_{SP[k]}}\right]_{\theta_{SP[i]}} - \left[\overline{\widehat{Q}_{SP[k]}}\right]_{\theta_{TP}}}{\overline{Q_{SP[k]}}} \tag{5}$$

< < Main point 3) What are the implications of ρ_i and ρ'_{M1M2} being smaller or larger in terms of parameter transferability? Given the focus of the manuscript, this should be explained in more detail when these measures are introduced in section 3.3. > >

There are different ways to get measurements of parameter transferability, which rely on the study of errors variations over time. In this work, we chose to focus on biases

HESSD
10, C6371–C6384, 2013

on mean volumes. Computing the standard deviation of the $\omega_{\theta_{SP}}$ or $\omega_{\theta_{TP}}$ curves is one way to assess the transferability of a given parameter set, with respect to volume errors. The smaller the standard deviation (i.e. the flatter the curve), the greater the temporal transferability of the parameter set.

Concerning now the ρ_i and ρ'_{M1M2} criteria: they have been designed to compare parameter sets with respect to their transferability. More precisely, they measure the degree of "parallelism" between ω_θ curves. While we could have simply studied the numerators from these criteria, we chose to normalise this "parallelism measurement" by the magnitude of bias variations. The ρ_i and ρ'_{M1M2} criteria can therefore be seen as "noise-to-signal" ratio where the "noise" = "parallelism imperfections" is compared to the "signal" = "variability of volume errors over time". Taken alone, they are not informative on the transferability level of a model or a parameter set. They only reveal whether different parameter sets (calibrated on different periods) face similar transferability issues.

In response to the referee's remark, we largely modified the section 3.4 where the numerical criteria are introduced to provide better and clearer explanations and therefore ease understanding.

< < Main point 4) Page 11353, lines 2-7: Are the differences statistically significant? > >

The limited number of points available to build the boxplots in Fig. 6 (20 points for each model) limits the possibilities to perform relevant statistical tests to confirm our qualitative assessments. Yet, we did perform some distribution comparisons using the non-parametric Kolmogorov–Smirnov test: The comparison between Mouelhi and GR4J-CemaNeige gives a K–S p-value of 0.571, which confirms that the distributions are similar. When results for Mouelhi and Cequeau are compared, we get a K–S p-value of 0.033, which tends to confirm that the differences are statistically significant.

HESSD
10, C6371–C6384, 2013

Finally, the comparison between GR4J-CemaNeige and Cequeau distributions gives a K–S p-value of 0.174, which does not allow to conclude on statistically significant differences, in spite of our qualitative interpretation of Fig. 6.

Because of the limited number of points used for these tests, we prefer not to provide p-values in the article. However, we modified the text to mention the fact that statements such as “are almost identical” and “differ from” are only qualitative assessment. We also make a reference to Table 2 (newly added), which provides simulation performances in accordance with the interpretations made from Fig. 6.

———

< < Minor point 1) American English and British English used interchangeably [. . .] > >

The language for the entire article has been checked to ensure coherence regarding this point (British English was chosen as convention).

———

< < Minor point 2) Throughout the text different terms are used with the same meaning, e.g. mean flow volume error, mean volume bias [. . .].Consistent terminology should be used throughout the paper. > >

Corrections have been made throughout the text and the term “mean flow volume error” was kept as the reference one.

———

< < Minor point 3) Page 11340, line 3-5: ‘(. . .) the blame for failure situations (. . .) seems to often be blamed (. . .)’, the Reviewer is of the opinion that this sentence should be rewritten, given its lack of clarity.. > >

Indeed, there was a mistake in this sentence which has been rewritten.

———

HESSD

10, C6371–C6384, 2013

< < Minor point 4) Page 11340, line 3-5: Reference needed > >

This comment refers to the sentence “Moreover, in conceptual modelling, failure situations of parameter transfer often seem be blamed on the overly simplistic model used or the inadequate calibration period chosen, without proper checking.”

The authors are unfortunately unable to provide any reference for this statement. It nonetheless expresses and summarises series of comments we have received from some colleagues (in paper reviews or during oral presentation at conferences) when we presented some of our work with parsimonious (and yet well performing) models such as GR4J, and tried to discuss the model uncertainties of robustness.

———

< < Minor point 5) Page 11340, line 11: The sentence should be rewritten to clarify this. > >

The sentence was modified.

———

< < Minor point 6) Page 11340, lines 13-14: [. . .]‘imagined’ should be replaced by ‘devised’ or ‘established’. > >

The sentence was modified.

———

< < Minor point 7) Page 11340, line 24: Attention must be paid to the punctuation. > >

The sentence was modified.

———

< < Minor point 8) Page 11341, line 23: Explicitly state which empirical formula was used to calculate PE. > >

HESSD

10, C6371–C6384, 2013

PE data series are inputs for the Mouelhi formula and the GR4J-CemaNeige model. For these models, daily values were computed using the formula from Oudin et al. (2005). They were used to feed GR4J-CemaNeige, while an aggregation was performed at the annual time-step to feed the Mouelhi formula. The situation was different for the Cequeau model whose internal code computed its own PE. This was done using the formula from Thornthwaite (1948) and the model therefore took daily series of air temperature as input.

The paper was modified to clarify this, both in the data description section (2.1.1) and the model description section (2.2.1, 2.2.2, 2.2.3).

———

< < Minor point 9) Page 11342, line 17: [. . .] it would be clearer to say 'the two simpler models'. > >

The sentence was modified.

———

< < Minor point 10) Page 11343, lines 4, 7, 8: Is the input rainfall or precipitation? > >

The inputs for all three models are indeed total precipitation. The terms were modified in the paper.

———

< < Minor point 11) Page 11343, line 17: By production part do the Authors mean soil moisture accounting part? > >

Yes, this was changed in the paper.

———

< < Minor point 12) Page 11343, line 6: ËĘQ and Q have not been defined before. > >

Indeed. This has been clarified.

HESSD

10, C6371–C6384, 2013

———

< < Minor point 13) Page 11344: Gradient descent is used to find a local minimum, while to find a local maximum of a function the procedure is known as gradient ascent. Given that the Authors say on line 2 that 'Model parameters were calibrated by maximising KGE' in line 10 where it reads 'descent' it should read 'ascent'. > >

That is correct. Historically, we used an RMSE-type objective criterion which needed to be minimised. But we forgot to adapt the terms when we switched to a KGE criteria. "descent" has now been replaced by "ascent".

———

< < Minor point 14) Page 11344, lines 2122 –Page 11345, line 1: GSST procedure is introduced and then the Authors state that the procedure adopted in this paper is different. At no point the Authors explain why or how the procedure adopted here is different. Therefore, it is unclear to the Reviewer why the GSST procedure is introduced. If the GSST procedure is of relevance, it should be clearly stated how the procedure adopted differs and the reasoning for this. > >

The paragraph was modified to add some information about the GGST procedure, and the simpler/modified version used in this study.

———

< < Minor point 15) Page 11345, lines 5-7: Given the lack of clarity of the sentence beginning 'The length of this sliding window (. . .)', the Reviewer is of the opinion that this text should be rewritten. > >

The paragraph was modified and should be easier now to read and understand.

———

< < Minor point 16) Page 11345, line 11: For 10-yr long calibration period and 40 years of data, there will be 31 sub-periods (as the Authors state). For 10-yr long calibration

HESSD

10, C6371–C6384, 2013

period and 62 years of data, there will be 53 sub-periods (and not 52 as the Authors state). > >

This is correct. This typo has been corrected.

———

< < Minor point 17) Page 11348, equation 3: p in equation (3) has not been defined. > >

This is correct. The section 3.4 has been largely rewritten. Missing notations were added, equations (4) and (5) were corrected (since they were wrong in the submitted version), and some rewriting and reorganising have been made with the aim of easing the readers understanding.

———

< < Minor point 18) Page 11348, lines 11-12: The reason why the curve calibrated over the total period is used as a reference should be made clear. > >

This curve reflects the mean flow volume errors under calibration conditions, when both the calibration and simulation periods correspond to the total period. Because volume errors are an important component of the KGE calibration criterion, we expected this reference curve to be the flattest of all the ω_θ curves. This is why we chose to consider it as a reference in our comparison scheme.

The corresponding sentence in the article was modified to add this information.

———

< < Minor point 19) Page 11349, line 5: The optimal situation with no errors corresponds to. In that case, equation (4) would give $\sigma[\omega_{\theta_{TP}}] = 1/p$ instead of 0 (as stated by the Authors). Either '0' in line 5 is wrong or the equation used to calculate $\sigma[\omega_{\theta_{TP}}]$ is different from what is shown by equation (4). > >

HESSD

10, C6371–C6384, 2013

Equation (4) was wrong. Please see the authors answer to minor point 17, and the new version of section 3.4 in the paper.

———

< < Minor point 20) Page 11349, line 7: where it reads '0 (situation where the $\omega_{\theta_{TP}}$ curves are rigorously identical' should be rephrased to be consistent with the rest of the text. [. . .] > >

The sentence was modified to add consistency.

———

< < Minor point 21) Page 11350, line 6: The sentence 'the smaller the value, the stronger the similarities between the ω_θ curves' would be clearer if written as 'the smaller the value, the stronger the similarities between the $\omega_{\theta_{SP}}$ and $\omega_{\theta_{TP}}$ curves'. > >

The sentence was modified to ease the readers understanding.

———

< < Minor point 22) Page 11350, lines 21-22: The Reviewer suggests to remove the expression 'meet the objective seeking to' to improve the clarity of the sentence. > >

The sentence was modified accordingly.

———

< < Minor point 23) Page 11351, lines 14-15: It is not clear to the Reviewer what is 'obviously not the case for the catchments considered here'. The Reviewer agrees that when the parameter set is optimised making use of the full record that is an "interpolation case" and so $\omega_{\theta_{TP}}$ curves fall in this category. Moreover, and as the Authors highlight, 'this curve is placed so that the mean volume bias of the entire period remains close to 1'. Therefore, it is unclear what 'obviously not the case' refers to. The

HESSD

10, C6371–C6384, 2013

last part of the sentence should be rewritten to clarify this. > >

The reviewer is correct that this sentence was unclear, as referee 1 also pointed out.

As a result, we chose to almost fully remove this sentence, since it did not add much. The previous sentences were adapted to improve clarity.

———

< < Minor point 24) Page11352, lines 9-11: Do the Authors mean here that the most spaced out curves are the ones calibrated based on the first and last 10-yr periods? Please clarify. > >

No, this is not what we meant. Therefore we modified the sentence to ease the readers understanding (and we provided an example using Fig. 4).

———

< < Minor point 25) Page 11353, lines 9-10: Did the Authors notice the same difference when using Cequeau's model with PE calculated using the same methodology as for the other two models? [. . .] > >

Unfortunately the comparison of the Mouelhi formula, GR4J-CemaNeige and Cequeau using exactly the same PE formula was not performed. The reason for this is that we could not modify one of the models' code to implement different formula. Therefore we simply followed the recommendations of model developers in terms of common application conditions. However, we fully agree that the comparison must be made to progress on the understanding of such differences. We made tests replacing the Oudin's PE with a Penman-Monteith PE when feeding Mouelhi and GR4J-CemaNeige. However, we did not obtain easily interpretable results, which could help our understanding (this is mentioned in the paper in section 5.2). We plan further comparison work especially focusing on this issue and will report our findings then.

Regarding the current paper, we modified the sentence to express the need for further

HESSD

10, C6371–C6384, 2013

tests.

< < Minor point 26) Page 11354, lines 7-12: The questions on the volume error similarities between sub-period and total period calibration for each model over different catchments are the first and third (instead of the first two). > >

This is perfectly correct. Thank you for pointing that out. We updated the items order in the section 3.3.4.

< < Minor point 27) Page 11354, lines 18-19: The Authors state that either the simplest or the most complex of the three models is used as M1, but in Figure 8, M1 can be any of the three models. In Figure 8, on the x-axis M1 and M2 should be swapped, as M1 are the models used as a reference. > >

Again, this is perfectly correct. There was a typo there which has been corrected.

---

28) Page 11355, lines 1-8: The interoperability of the manuscript could be improved by the Authors clearly stating which metric from Section 3.3 they are referring to when they say 'volume bias variations', 'relative variations of these biases'. > >

As stated in the answer to minor point 2, modifications were made throughout the text to reduce the multiplicity of terms used in the paper when discussing mean flow volume error. The modifications made also concern the use of terms such as "bias" or "relative bias".

< < Minor point 29) Page 11358, lines 15: It is unclear what the Authors mean by 'structural deficit'. > >

HESSD

10, C6371–C6384, 2013

With these terms we meant “the issues linked with inappropriate model structure”.

For better clarity, the terms were simply removed (the previous sentences being already clear enough regarding the topic discussed in this paragraph).

———

< < Minor point 30) Page 11359, line 23: The Reviewer does not agree with the use of the expression ‘particularly suspected’ and suggests this should be replaced. > >

The terms were replaced by “particular attention should be given to the estimation of . . .”

———

< < Minor point 31) Page 11373, Figure 2: What is the meaning of ‘m’ in the caption of Figure 2? > >

This was a typo which has been corrected: “m” was replaced by “parameters”.

———

< < Minor point 32) Page 11377, Figure 6. Y-axis label should use the same symbols (i.e. $\sigma[\omega_{\theta_{TP}}]$ contained in the text (as in Figure 7). > >

This modification was made.

———

< < Minor point 33) Page 11382, Figure A2: Same as previous comment. > >

This modification was made.

———

END

Please also note the supplement to this comment:

HESSD

10, C6371–C6384, 2013

http://www.hydrol-earth-syst-sci-discuss.net/10/C6371/2013/hessd-10-C6371-2013-supplement.pdf

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 10, 11337, 2013.