

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.

Anonymous Referee #2

Received and published: 9 October 2013

Overview: The aim of the paper is very clear: evaluate the temporal robustness of conceptual hydrological models to changing conditions. In other words, the Authors examine if parameters can simulate flows adequately on a different period from that on which they were calibrated (transference of parameters in time). In the literature, the lack of transferability frequently found is commonly attributed to the simplistic model structure used without adequate investigation of the causes (as noted by the Authors in the Introduction). In this study the Authors try to understand if the lack of robustness is a consequence of the limited transferability in time of the water balance adjustments

C5497

made through parameter optimisation, or whether it stems from the models' overall inability to reproduce water balances simultaneously on different sub-periods. This is a very pertinent question and the Authors address it in a systematic way. Overall the paper is well-structured and well-referenced. However, it is advisable that the Authors rephrase some sentences throughout the text. Specific examples are listed in the minor comments section, but the list is not exhaustive. The Reviewer would like to encourage the Authors to revise the manuscript in order to improve the overall standard of the paper.

Main Points: 1) The way year is defined in this study is not clear. There is a reference to 'hydrological year' on page 11345, line 4, but it is not clear to the Reviewer whether the Authors used water or calendar year. Given the prevalence of snow in some of the catchments it would be advisable to use hydrologic year rather than calendar year to avoid possible impacts of carry-over water storage from one calendar year to the next one (Sawicz et al, 2013). 2) On page 11349, lines 9-10, when explaining the meaning of $\sigma[w_{\theta_{SP[i]}} - w_{\theta_{TP}}]$, the Authors state that 'only the shape similarities of the w_{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 $w_{\theta_{TP}}$ but at different distances from it have the same value for $\sigma[w_{\theta_{SP[i]}} - w_{\theta_{TP}}]$. 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. 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. 4) Page 11353, lines 2-7: Are the differences statistically significant?

C5498

Minor points: 1) American English and British English used interchangeably. Some examples (among many others) include: on page 11458, line 3, 'generalize'; on page 11351, line 12, 'optimized', on page 11338, line 21, 'modelling', on page 11345, line 20, 'behaviours'. 2) Throughout the text different terms are used with the same meaning, e.g. mean flow volume error, mean volume bias, mean volume error, mean flow bias, relative bias, etc. Consistent terminology should be used throughout the paper. 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. 4) Page 11340, line 3-5: Reference needed. 5) Page 11340, line 11: The Authors say that 'DMIP, MOPEX and HEPEX are good examples of use for these testing schemes'. The Reviewer assumed that the Authors mean 'are good examples of the use of these testing schemes'. The sentence should be rewritten to clarify this. 6) Page 11340, lines 13-14: The Reviewer does not agree with the use of the expression 'imagined' and therefore it should be replaced, e.g. 'devised' or 'established'. 7) Page 11340, line 24: Attention must be paid to the punctuation. 8) Page 11341, line 23: Explicitly state which empirical formula was used to calculate PE. 9) Page 11342, line 17: Instead of 'the other two models', it would be clearer to say 'the two simpler models'. 10) Page 11343, lines 4, 7, 8: Is the input rainfall or precipitation? 11) Page 11343, line 17: By production part do the Authors mean soil moisture accounting part? 12) Page 11343, line 6: \hat{Q} and Q have not been defined before. 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'. 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. 15) Page

C5499

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. 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 period and 62 years of data, there will be 53 sub-periods (and not 52 as the Authors state). 17) Page 11348, equation 3: p in equation (3) has not been defined. 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. 19) Page 11349, line 5: The optimal situation with no errors corresponds to $[\hat{Q}_{SP[k]}]_{\theta_{TP}} = \hat{Q}_{SP[k]}$. In that case, equation (4) would give $\sigma[w_{\theta_{TP}}] = 1/p$ instead of 0 (as stated by the Authors). Either '0' in line 5 is wrong or the equation used to calculate $\sigma[w_{\theta_{TP}}]$ is different from what is shown by equation (4). 20) Page 11349, line 7: where it reads '0 (situation where the $w_{\theta_{TP}}$ curves are rigorously identical' should be rephrased to be consistent with the rest of the text, e.g. '0 (situation where the $w_{\theta_{TP}}$ and $w_{\theta_{SP[i]}}$ curves are rigorously identical' (as in page 11348, line 9) or '0 (situation where the $w_{\theta_{TP}}$ and $w_{\theta_{SP}}$ curves are rigorously identical' (as in page 11351, line 10) or '0 (situation where the w_{θ} curves are rigorously identical' (as in line 10). 21) Page 11350, line 6: The sentence 'the smaller the value, the stronger the similarities between the w_{θ} curves' would be clearer if written as 'the smaller the value, the stronger the similarities between the $w_{\theta_{SP}}$ and $w_{\theta_{TP}}$ curves'. 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. 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 $w_{\theta_{TP}}$ 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 last part of the sentence should be rewritten to clarify this. 24) Page 11352, 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. 25) Page

C5500

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? This would provide a better understanding of whether the Cequeau's better results are related to its greater structural complexity. Without that comparison, it is not possible to isolate the potential effects of improved input quality on the model performance (as pointed by the Authors). 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). 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. 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'. 29) Page 11358, lines 15: It is unclear what the Authors mean by 'structural deficit'. 30) Page 11359, line 23: The Reviewer does not agree with the use of the expression 'particularly suspected' and suggests this should be replaced. 31) Page 11373, Figure 2: What is the meaning of 'm' in the caption of Figure 2? 32) Page 11377, Figure 6. Y-axis label should use the same symbols (i.e. $\sigma[w_{\theta_{TP}}]$) contained in the text (as in Figure 7). 33) Page 11382, Figure A2: Same as previous comment.

References: Sawicz, K. A., Kelleher, C., Wagener, T., Troch, P., Sivapalan, M. Carrillo, G. (2013). Technical Note: Characterizing hydrologic change through catchment classification. *Hydrol. Earth Syst. Sci. Discuss.*, 10 (5), 6599-6627.

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