Hydrol. Earth Syst. Sci. Discuss., 10, C6358–C6370, 2013
www.hydrol-earth-syst-sci-discuss.net/10/C6358/2013/

Hydrology and Earth System Sciences
Open Access
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 3

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

———

< < Point 2. Although the paper does a good job in illustrating that a problem exists, it leaves several open questions, and as a result, it sounds somewhat incomplete. The paper shows that robustness problems in rainfall runoff modelling are common, but it does not answer the question of why these problems exist. The paper attributes this problem to the models themselves, but it could very well be a problem of the data, including discharge, precipitation, and potential evaporation estimates. > >

We agree with the reviewer that the paper leaves several open questions regarding the actual causes for the robustness problems observed during our tests. This is pointed out in the paper: in the abstract (submitted version: page 11338 lines 14-15; revised version: page 2 lines 13-14), in the discussion section (submitted version: page 11358 lines 2-4; revised version: page 32 lines 20-22), in the conclusion section (submitted version: page 11361 lines 17-23; revised version: page 37 lines 12-17).

The potential explanations for the robustness issues detected in our tests are discussed in the sections 5.1, 5.2 and 5.3. The potential role of incorrect input estimates on the modelling performances is specifically addressed in section 5.2.

However, in accordance with the reviewer's remark, we acknowledge that one of the statements made in the conclusion was too strong because it was incomplete. This statement was: "[. . .] this reveals that the lack of robustness identified for some catchments on 10-yr-mean flows is not caused by a poor choice of calibration period but rather stems from the models' overall inability to reproduce water balances simultaneously on different sub-periods" (submitted version: page 11361 lines 11-14). We modified this statement to balance the model's responsibilities and we now say: "this reveals that [. . .] is not caused by [. . .] but rather stems from the models' overall inability to reproduce water balances simultaneously on different sub-periods (considering their usage conditions: structural choices, inputs data, PE formula, etc.)."

Overall, we acknowledge that the paper may sound somewhat incomplete as a result of this lack of answer on the causes for the robustness problems. This acknowledgment was made in the discussion section (submitted version: page 11358 lines 5-8; revised version: page 32 lines 23-25). In spite of this lack of answer, we believe this work represents an interesting contribution in that it shows strong similarities between simulation errors from different models, used in (almost) similar conditions. The extent of the "parallelism effect" was of great interest for us and we do not know previous work showing this in the literature. We think that this type of analysis can be a sound basis to help further diagnosing the causes of modelling failures and assess model transposability.

< < Point 3. Some additional analyses could be helpful to investigate this in more detail. For example, I see from fig. 5 that precipitation and discharge time series are quite well correlated. The difference between them is the actual evaporation. How does the actual evaporation correlate with the potential? And how does the actual evaporation from different catchments correlate with each other? There should probably be some correlation, which could give some indication of potential uncertainties in the data, and particularly in the evaporation data.

Point 4. Similarly, how do precipitation and discharge estimates from the different catchments correlate? Can this be used to say something about data uncertainty?

Point 5. How do the anomalies observed in different catchments correlate? Are they all different, or are they similar? > >

As the reviewer rightly points out, many analyses can be done to investigate if the input data are responsible for the robustness issues observed.

When we started our work on this topic, we had true hopes for understanding the causes of transferability issues. We thought that studying the climate forcings temporal variations and their corresponding impacts on the model state variables and output errors would permit such understanding. However, we faced a rather complex scientific

problem, where the potential causes are multiple and may combine. As a result, we are unable to provide clear answers to the reviewer's comments #3, 4 and 5. Some elements of discussion regarding these comments are provided in the next paragraphs (a-b-c).

a) During our numerous tests on the temporal robustness of hydrological models, we searched for correlations between climate forcings and model robustness issues over the large catchment sets considered. During this search, changes in various climatic indicators were considered, among which: changes in mean air temperature, mean total and solid precipitation volumes, mean aridity index, as well as inter-annual variability and seasonal variability of these climatic indicators.

Changes in mean air temperature and mean total precipitation volumes showed the highest correlation levels with the model robustness issues. Corresponding results on an Australian and a French dataset can be found in Coron et al. (2012, WRR) and Coron (2013, PhD thesis (in French), available at: http://pastel.archives-ouvertes.fr/pastel-00879090/), respectively. Complementary tests were even made on the French dataset to assess the risk of misinterpreting the correlation between model errors and temperature changes (see Fig 1 at the end of this document for more information).

However, for other parts of the catchment sets considered, no significant correlation could be established between the model errors and changes in climatic conditions (among which some of the catchments used for the work reported in this paper). In spite of several attempts, we were unable to determine the differences between catchments where high correlation were establish from those (sometimes neighbouring catchments) where no obvious explanation of robustness issues could be found.

b) During our attempts to better understand the results of our tests, we looked at how model anomalies from different catchments correlate (as suggested by the reviewer). An example of such investigation is illustrated in the Figs 2-3 (at the end of this document). This example shows GR4J-CemaNeige volume error variations on the Allier at Vieille-Brioude catchment (included in the 20-catchment set used in the paper) and some upstream nested catchments.

This example illustrates that drawing strong conclusions is complex. In average, some correlations do exist between the anomalies observed on neighbouring catchments or on nested catchments. This could advocate for a model overall inability to reproduce some long term hydro-climatic variations and/or for inaccurate regional estimates for the model's input data. However, dissimilarities can also be found between model anomalies for neighbouring catchments or catchments with similar characteristics. Such dissimilarities may be the sign that model errors are related to errors on discharge measurements or site specific characteristics which were not taken explicitly into account in the model (e.g. local geology or possible land cover changes).

c) As shown in this simple example, all these configurations mix in practice. This makes it extremely complicated to draw generic conclusions, valid over a sufficient number of cases. That is the reason why we decided not to report in the paper our tests regarding these aspects and only make simple statements to limits the risk of overwhelming the readers with information and make sure our main finding (i.e. the "parallelism effect") would stay out as the main message.

Concerning now the discussion of these aspects in the paper, we stated in the section 4.1 of the submitted version that "we observe different behaviours depending on the catchment considered. [. . .]Explaining why these errors occur is complex. Some causal links may be inferred from these examples, related to changes in climate forcings (e.g. changes in mean air temperature for the Lot River). Our recent investigations on this topic, however, showed that these correlations are not systematic and that their significance greatly varies from one catchment to another (Coron, 2013)."

In the revised version, we complete this statement with: "[...] However, our recent investigations on this topic showed that if such correlations can be established in nu-

merous cases, they are not systematic and their significance greatly varies from one catchment to another (Coron, 2013). To date, we therefore remain unable to draw general conclusions regarding the spatial similarities in model volume error variations and can only acknowledge the need to further investigate this complex question."

Moreover, we added the following paragraph to the revised discussion section 5.2: "Finally, investigations on the spatial similarities of model volume errors over the catchment set can help assess the role of data quality on models robustness issues. Indeed, strong dissimilarities between the volume error curves of different catchments, in spite of their common characteristics, may be caused by time-variable errors in discharge measurements. Conversely, similarly shaped volume error curves of neighbouring catchments may be obtained as a result of inaccurate regional estimates of the model's input forcings (e.g. if the bias in precipitation estimates evolves in time or if the method used for PE computation is inappropriate)."

< < Point 6. The Authors could show the temporal dynamics of parameter estimates. These could give indications on how model structures compensate for data inaccuracy. I think this is quite interesting and could be a good complement to the paper. > >

This point was investigated during some previous work we did on the large French dataset with four hydrological models (Coron 2013, PhD thesis). We performed a large number of split sample tests using the GSST methodology introduced in Coron et al. (2012, WRR). Then we tried to relate the model error variations with the changes in climatic conditions between the calibration and validation periods.

From these investigations, we were able to establish some correlations between temporal changes in optimal value for some parameters and model changes in model volume errors. However, the existence of these correlations and the type of parameters concerned were highly dependent on the model structure.

Not all the parameters showed such temporal variations. We found that the ones that were primarily concerned had a significant role in closing the water balance between precipitation, actual evapotranspiration and river discharge. Depending on the model, this could be a parameter influencing the amount of water to evaporate from conceptual reservoirs, a correction factor for precipitation bias or some control on groundwater flow exchange with surrounding catchments.

For less parsimonious models, such as Cequeau, it was however impossible to deduce clear tendencies on any parameter, likely due to the multiple possibilities the model has to adjust its water balance and the possible model overparameterisation.

The existence of such trends on some parameters results from a combination of two aspects: a) the objective function commonly used are relatively demanding with respect to volume bias, and b) the models tested show an overall inability to reproduce observed water balances simultaneously on different sub-periods (considering their usage conditions: input data sets, PE formula, etc.).

Because the objective function is demanding regarding mean volume errors, the calibration algorithm tends to select parameter values ensuring low volume errors (the parameter concerned depending on the model structure). Then, since models face difficulties to reproduce mean flow volumes on different periods simultaneously, the optimal values for the parameters allowing the water balance adjustment will vary in time. Finally, because these two aspects (a & b) are linked, correlation are found between the variations of optimal values for these parameters and the transferability issues detected via split-sample test procedures.

Coming back to the work reported in the paper, we did consider showing the variations of optimal parameter values for the three case studies used in Fig. 5 from the paper. The Fig 4 at the end of this document illustrates the type of figures we could have shown. However, we preferred not to do so for the following reasons:

A) Since no clear correlation was visible for the Cequeau parameters, it was difficult

to make an arbitrary choice regarding which of the 19 parameters to show and we could not provide the variations for all of them since their number was too large, B) It is true that some relevant correlations could be seen for the 1st or 2nd parameter of GR4J-CemanNeige and could have been interpreted, with respect to our water balance adjustments issues. In the same time, large parameter variations were observed for the 5th and 6th parameters, which correspond to the snow module. They should not be interpreted in relation with our robustness issues since they simply results from the absence of sufficient snow pack on the latest years for proper calibration. They have a very limited impact on 10-yr-mean flow volume errors during transfer and we thought showing them might mislead the readers understanding. C) We have had difficulties to find relatively simple tools to explain and illustrate the “parallelism effect” discussed in the paper. We think adding too much information could harm the clarity of the general message we want to give here, especially since this type of additional information was hard to interpret and discuss.

< < Point 7. These analyses could in place of paragraph 4.4, and the resulting figures, which do not add much to the paper. > >

We agree with the referee that section 4.4 and the associated Figs. 9 and A3 do not present new findings compared with the elements discussed in section 4.2 and illustrated with Fig. 7. However, the way the “parallelism-effect” is emphasised and graphically illustrated is quite different between these two sections.

The numerical criteria proposed in the section 3.3 (now numbered 3.4) are quite powerful to assess the extent of the “parallelism” we found during our testing, but the associated figure are not entirely self-explanatory and might be difficult to interpreted without reading the paper in detail. On the other hand, Fig. 9 and associated Fig. A3 are rather simple to understand and illustrate well how the relative changes in 10-yr-mean flow remain stable when different parameter sets are considered.

We therefore see section 4.4 as an additional help for the reader to understand the basic/practical impact of the “parallelism-effect” previously described in a more detailed/numerical way.

———

< < Point 8. A minor comment on Figure 4 panel c. The lines and corresponding dots on the x-axis could be of different colour. > >

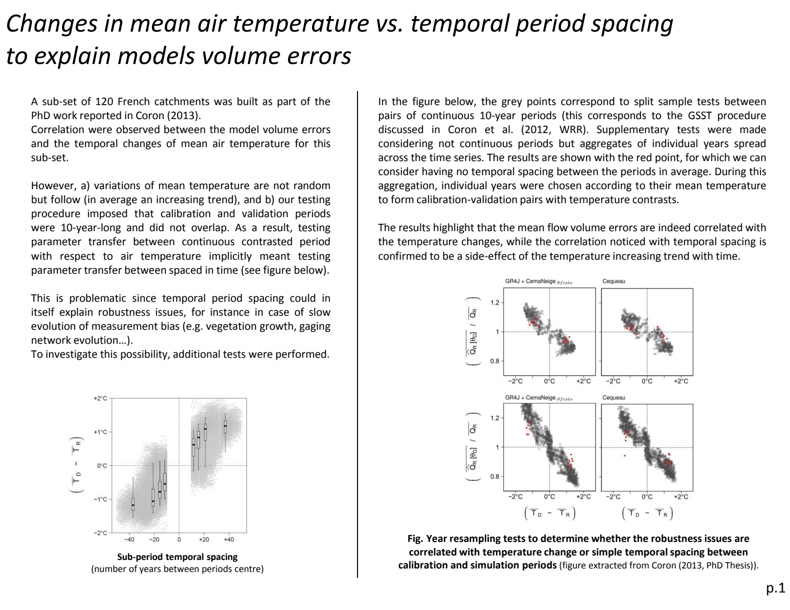
Figure 4 has been modified to add colours.

———

END

Please also note the supplement to this comment:
http://www.hydrol-earth-syst-sci-discuss.net/10/C6358/2013/hessd-10-C6358-2013-supplement.pdf

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



Changes in mean air temperature vs. temporal period spacing to explain models volume errors

A sub-set of 120 French catchments was built as part of the PhD work reported in Coron (2013).
Correlation were observed between the model volume errors and the temporal changes of mean air temperature for this sub-set.

However, a) variations of mean temperature are not random but follow (in average an increasing trend), and b) our testing procedure imposed that calibration and validation periods were 10-year-long and did not overlap. As a result, testing parameter transfer between continuous contrasted period with respect to air temperature implicitly meant testing parameter transfer between spaced in time (see figure below).

This is problematic since temporal period spacing could in itself explain robustness issues, for instance in case of slow evolution of measurement bias (e.g. vegetation growth, gaging network evolution…).
To investigate this possibility, additional tests were performed.

Sub-period temporal spacing
(number of years between periods centre)

In the figure below, the grey points correspond to split sample tests between pairs of continuous 10-year periods (this corresponds to the GSST procedure discussed in Coron et al. (2012, WRR). Supplementary tests were made considering not continuous periods but aggregates of individual years spread across the time series. The results are shown with the red point, for which we can consider having no temporal spacing between the periods in average. During this aggregation, individual years were chosen according to their mean temperature to form calibration-validation pairs with temperature contrasts.

The results highlight that the mean flow volume errors are indeed correlated with the temperature changes, while the correlation noticed with temporal spacing is confirmed to be a side-effect of the temperature increasing trend with time.

Fig. Year resampling tests to determine whether the robustness issues are correlated with temperature change or simple temporal spacing between calibration and simulation periods (figure extracted from Coron (2013, PhD Thesis)).

p.1

Fig. 1.

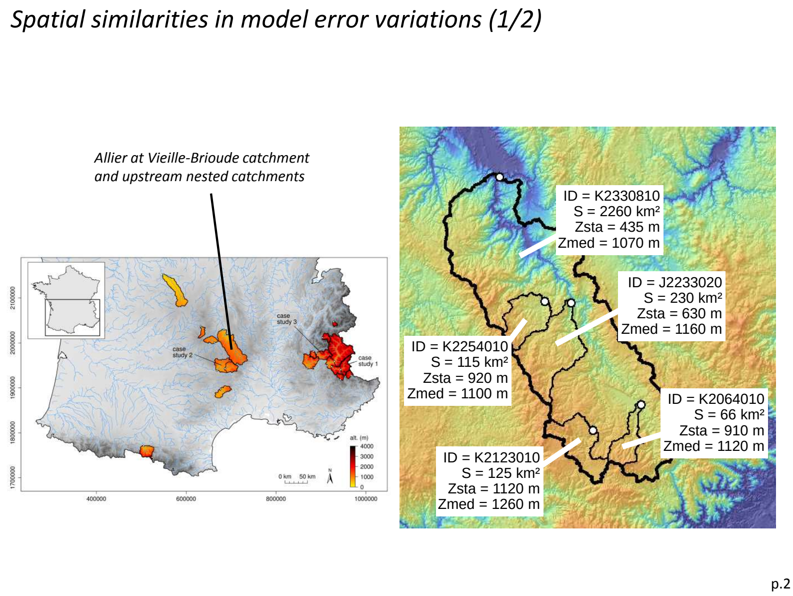

Spatial similarities in model error variations (1/2)
Allier at Vieille-Brioude catchment
and upstream nested catchments
case study 3
case study 2
case study 1
alt. (m)
ID = K2330810
S = 2260 km²
Zsta = 435 m
Zmed = 1070 m
ID = J2233020
S = 230 km²
Zsta = 630 m
Zmed = 1160 m
ID = K2254010
S = 115 km²
Zsta = 920 m
Zmed = 1100 m
ID = K2064010
S = 66 km²
Zsta = 910 m
Zmed = 1120 m
ID = K2123010
S = 125 km²
Zsta = 1120 m
Zmed = 1260 m
p.2

Fig. 2.

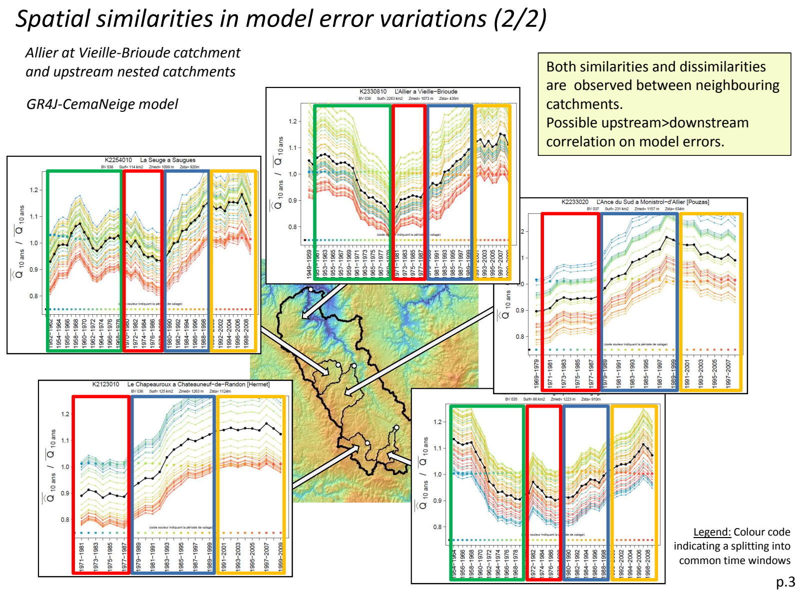

Spatial similarities in model error variations (2/2)
Allier at Vieille-Brioude catchment
and upstream nested catchments
GR4J-CemaNeige model
Both similarities and dissimilarities are observed between neighbouring catchments.
Possible upstream>downstream correlation on model errors.
K2330810 L'Allier à Vieille-Brioude
K2254010 La Seuge à Saugues
K2233020 L'Ance du Sud à Monistrol-d'Allier [Pouzas]
K2123010 Le Chapeauroux à Chateauneuf-de-Randon [Hermet]
Legend: Colour code indicating a splitting into common time windows
p.3

Fig. 3.

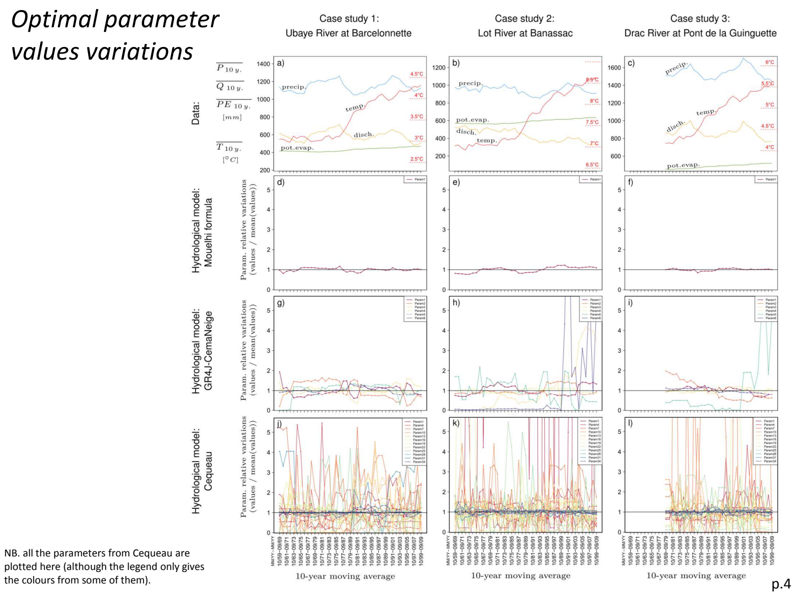

Optimal parameter values variations
Case study 1: Ubaye River at Barcelonnette
Case study 2: Lot River at Banassac
Case study 3: Drac River at Pont de la Guinguette
Data:
Hydrological model: Mouelhi formula
Hydrological model: GR4J-CemaNeige
Hydrological model: Cequeau
Param. relative variations (values / mean(values))
precip.
temp.
disch.
pot.evap.
10-year moving average
NB. all the parameters from Cequeau are plotted here (although the legend only gives the colours from some of them).
p.4

Fig. 4.