We thank the reviewer for the time spent in reviewing our paper and making very helpful suggestions. We provided a pont-by-point response to the reviewer's comments.

Anonymous Referee #2

The article addresses the interesting issue of structural uncertainty in conceptual hydrological modelling. The authors test a large number of alternative structures on a catchment in the Andes in Chile and discuss their relative merits in a multi-objective framework.

Overall, I found the article interesting and well written. I think it could make a valuable contribution to HESS provided that a number of points are improved. I have two main concerns. First, the conclusions of this study do not appear so novel compared to existing works based either on multi-hypotheses or multi-objective frameworks. I think the authors should strengthen the last part (discussion/conclusion) of their paper to better demonstrate what was learnt from the quite complex testing scheme they set up and what is new compared to what was already shown in past studies.

Author's response

We thank the reviewer for this important remark. Our paper effectively draws upon the combination of a modular multiple-hypothesis approach with a multi-objective optimization scheme. We did not claim that these two modeling strategies were new (although modular approaches to multiple-hypothesis testing remain rare in comparison with multi-model approaches), but that the potential benefits of combining them within the same framework remain largely unexplored in current studies.

Perhaps more importantly, our study addresses a current lack of hydrological modeling effort in semi-arid Andes. As mentioned in the introduction part of the paper, "very few catchments in this region have been studied intensively enough to provide reliable model simulations, often with no estimation of the surrounding uncertainty". These two points are further detailed below in our answers to the following reviewer's comments. They have been emphasized in the new version of the manuscript.

Author's changes in manuscript

The discussion part of our paper has also been rewritten to better demonstrate what was learnt from the testing of a large number of model structures. In particular, note that the hypotheses made in Section 4.2 regarding possible links between model structures and the physical features of the catchment have been transferred to Section 5 ("Discussion and conclusion") in the updated manuscript.

Anonymous Referee #2

Second, their study would give more general conclusions if tests had been made on more than one catchment. Indeed, the conclusions may strongly depend on the characteristics of the selected catchment. It would be useful to test the approach to at least another catchment, to check whether similar conclusions are reached.

Author's response

In general, we agree that multiple-hypothesis frameworks should be tested on several catchments if one wishes to identify possible links between the physical features of a given catchment and some specific modeling decisions. This would also be very desirable considering the influence of data errors on the results obtained with any particular model structure or performance measure. While the need for comparative studies was only briefly mentioned in the paper (p. 12164, lines 9–11), it is further discussed in the updated manuscript (see Section 5, "Discussion and conclusion").

However, we would like to emphasize the fact that this study represents the first step of a larger research project, whose final aim is to assess the capacity to meet current and future irrigation water requirements in the Claro River catchment. Because considerable time and effort had to be devoted to gathering/interpolating the input data and implementing/testing the modeling framework, it was also necessary to limit the scope of our study to this particular catchment. Moreover, the main objective here was not to establish unambiguous relationships between the physical characteristics of Andean catchments and specific model requirements, but rather to assess the uncertainty associated with model non-uniqueness and structural inadequacy in the Claro River catchment. From this point of view, it should be stressed that the paper already provides a reliable framework by testing a total of 72 competing model structures in a region where catchment-scale conceptual models remain largely under-used. Adding other Andean catchments would be of particular interest to the objectives of the study if precipitation data on these catchments were available over the same 30-year period and could be considered more reliable. To our knowledge, this is not the case in the Andes in general. The dataset used in the paper actually includes several of the highest weather stations available at this time scale in the Chilean Andes.

Author’s changes in manuscript
We added a few comments to qualify our statements and insist on the need for comparative studies to confirm the hypotheses made in Section 4.2 regarding possible links between model structures and the physical features of the catchment. As mentioned above, note that these hypotheses have also been transferred to Section 5 (“Discussion and conclusion”) in the updated manuscript.

Anonymous Referee #2
I have also a number of detailed comments below. I think the paper could be reconsidered for publication after major revision.

Detailed comments:

1. Anonymous Referee #2
   There are remaining typos that should be corrected. Consistency between references in the text and the list of references at the end of the manuscript should also be further checked.

   Author's response
   Agreed. We apologize for these typos which have been corrected in the updated manuscript.

   Author's changes in manuscript
   Please see the updated manuscript for the detailed corrections.

2. Anonymous Referee #2
   Page 12139, line 25: The authors may find interesting reflections on this issue in the book edited by Wainwright and Mulligan (2004).

   Author's response
   We thank the referee for this interesting suggestion which has been inserted in the updated manuscript.

   Author's changes in manuscript
   The following statement on page 12139, lines 27–28 of the discussion paper:

   > “… as ready-made engineering tools with little or no consideration for the specific features of each catchment (Savenije, 2009)”

   Has been replaced in the updated manuscript with the following statement:

   > “… as ready-made engineering tools with little or no consideration for the specific features of each catchment (Wainwright and Mulligan, 2004; Savenije, 2009)”

3. Anonymous Referee #2
   Page 12142, lines 1–10: I do not agree that the multi-model approach was mainly focused on small catchments. There are a number of studies in the literature that investigated larger ranges of catchment size.

   Author's response
   Here we think there may be a definitional issue. In our opinion, a substantial distinction should be made between current multi-model strategies and modular modeling frameworks (MMF). While both rely to some extent on the concept of multiple-hypothesis testing, it should be noted that modular approaches offer the additional opportunity to examine the effect of each individual hypothesis (i.e. each modeling decision) by modifying only one component or constitutive equation at a time and testing a wide range of alternative combinations between model components. By

contrast, multi-model strategies generally involve ready-made model structures borrowed from the literature (“off-the-shelf” models).

We think this distinction was made clear on Page 12141 (lines 25–27) but maybe not enough on the following page where the issue of catchment size was discussed. Generally speaking, multiple-hypothesis frameworks should not be completely identified with modular modeling frameworks (as we did on Page 12142 and later in the paper) since in reality the latter represents only one possible approach to multiple-hypothesis testing. The manuscript was therefore modified to maintain a distinction between the two. This is important because, to our knowledge, most conceptual modular frameworks currently used in hydrological modeling studies have been applied to relatively “small” catchments of, at most, a few hundreds of km^2. As argued by the reviewer, this is not the case of more traditional multi-model approaches, which indeed cover a much larger range of catchment sizes.

To our knowledge, there is only one example of a study making use of a modular framework on a semi-arid catchment of more than 2000 km^2 and that is the original paper of Clark et al. (2008) introducing the FUSE toolbox. We agree that this point was somewhat overlooked in our paper and modified this part of the introduction to balance our statements with other arguments.

<u>Author’s changes in the manuscript</u>

The following statements on page 12143, lines 17–18 of the discussion paper:

> “So far, however, this method has mostly been applied to small (<10 km2) experimental (well-monitored) catchments (e.g. Clark et al., 2008; Smith and Marshall, 2010; Buytaert and Beven, 2011; McMillan et al., 2012b; Fenicia et al., 2014), with less attention being given to larger scales of interest (100–400 km2) (e.g. Kavetski and Fenicia, 2011; Coxon et al., 2013) or long time periods. Therefore, the need remains to establish whether MHF can also be used to improve conceptual modeling on multi-decadal periods at operational scales of 1000 km^2 or more. The potential benefits of combining MHF with Pareto-based optimization schemes also remain largely unexplored in the current literature.”

Have been replaced with the following statements:

> “So far, however, this method has mostly been applied to relatively small (<500 km^2) and humid catchments of the Northern Hemisphere (Krueger et al., 2010; Smith and Marshall, 2010; Staudinger et al., 2011; Kavetski and Fenicia, 2011; McMillan et al., 2012b; Coxon et al., 2013), with less attention being given to larger scales of interest (>1000 km^2) and semi-arid regions (Clark et al., 2008). Moreover, several of these studies have insisted on the need for multiple criteria related to different aspects of the system’s behavior in order to improve the usefulness of MMF. Yet, most of the time these additional criteria or signatures were not used to guide model development or constrain calibration but rather as posterior diagnostics in validation (see e.g. Kavetski and Fenicia, 2011). Thus, the potential benefits of using the concept of Pareto-efficiency to constrain model development and help differentiate between a large number of competing hypotheses remain largely unexplored in the current literature devoted to MMF. Also, very few studies have included alternative conceptual representations of snow processes in their modular frameworks (e.g. Smith and Marshall, 2010), even though snowmelt may have played a significant role in several cases (Clark et al., 2008; Staudinger et al., 2011).”

<u>Anonymous Referee #2</u>

Besides, what makes the application of such approaches to larger catchments essentially different given the lumped approach used? I found that the argument of scale to explain the novelty of the study not really convincing here.

<u>Author's response</u>

We do agree that the argument of scale may not be the most relevant of all to explain the novelty of our study. More fundamentally, we would like to insist on the fact that very little is currently known about the adequacy of commonly used conceptual models in semi-arid Andean catchments. Most modular multiple-hypothesis frameworks such as FUSE, SUPERFLEX and RRMT have been applied to humid or subhumid catchments characterized by a limited role of snowmelt. Moreover, as mentioned in the introduction, there is no strong evidence in our opinion that lumped conceptual models designed on small catchments remain adequate at larger spatial scales.

Author’s changes in the manuscript
As mentioned above, the argument of scale has been further qualified with other arguments in the updated manuscript.

4. Anonymous Referee #2
Section 2: As explained above, I found that adding another case study (possibly under similar or different conditions) would make conclusions more general. Here the catchment is quite specific in the sense that there seems to be a huge uncertainty in precipitation estimates. Adding another catchment with better known precipitation would provide a comparative reference to balance the results presented here.

Author's response
Please see our answer on the first page of this document.

5. Anonymous Referee #2
Page 12143, line 26: The location of gauges could be shown in Fig. 1.

Author's response
Agreed. This comment was also made by the other anonymous referee. Note however that all the weather stations cannot be shown on Figure 1 because many of them are actually located outside the catchment.

Author’s changes in the manuscript
Figure 1 was modified to include those precipitation and temperature stations which belong to the catchment.

6. Anonymous Referee #2
Page 12144, lines 17–22: I did not understand why the Oudin’s PE formula was adjusted to the Penman-Monteith’s one. Why not directly using the latter if it is found more adapted to the study site?

Author's response
We did not find the physically-based Penman-Monteith approach more adapted to the objective of our paper. Oudin et al. (2005) showed that this approach may actually be less advantageous than more empirical formulas when used in daily conceptual models. This is why we chose to use the Oudin’s formula in this study. The reason why we chose to adapt its parameters K_1 and K_2 to the local conditions is because our study was part of a larger project which involved the assessment of current and future irrigation water requirements in the catchment. In this project, the Penman-Monteith equation appeared more suited to simulating crop water needs in the valleys, but, because it required meteorological data that were only available for the last three or four years (relative humidity, wind speed, solar irradiance), it was decided to rely on a modified version of Oudin’s temperature-based formula, in which the values of K_1 and K_2 were determined by selecting those giving the best fit to the available Penman-Monteith estimates of PE. In fact, these modified values of K_1 and K_2 were very close to those found by Oudin et al. (2005) and a sensitivity analysis showed that such modifications had no impact on the performance of the hydrological models used in the present paper. As a consequence, we kept these modified values to remain consistent with the other part of the project.

Author’s changes in manuscript
In order to simplify our statement and avoid any misunderstanding, we removed these details on the estimation of PE from the updated manuscript. Instead, the reader is referred to Hublart et al. (2014) for more details on the values of K_1 and K_2.

7. Anonymous Referee #2
Page 12144, lines 22–25: This statement is a bit vague. Could the authors give more details on this and explain to which extent the naturalization process may introduce uncertainty in the evaluation of models?

Author's response

Agreed. This comment was also made by the other anonymous referee. We admit that this point was not made clear in the paper and this was mainly due to space limitations. As explained in Section 2.1., vineyards and orchards cover most of the valley floors and lower hill slopes, where they benefit from a unique combination of clear skies, high temperatures and overall dry conditions throughout the growing season. Most of the annual precipitation, however, occurs as snow during the winter months, leading to an entire dependence on surface-water resources to satisfy crop water needs during the summer. Irrigation water abstractions occur at multiple locations along the river's course depending on both historical water rights and water availability. Because these abstractions are likely to influence the hydrological behavior of the catchment, especially during low-flow periods, they were added back to the observed stream flows before calibrating the models. This inevitably adds some uncertainty to the modeling of daily stream flows because a significant part of surface-water abstractions actually return to the river system within a few days. In general, ignoring these return flows will lead to overestimating natural stream flows on a daily basis. In this paper, however, the actual water withdrawals were not known with precision but only as percentages of the nominal water rights (these percentages are fixed on a monthly basis by the authorities depending on water availability), so the overall effects of streamflow naturalization on model uncertainty remained unknown.

Author's changes in manuscript

The following statements on page 12143, lines 17–18 of the discussion paper:

> "…but account for less than 1% of the total catchment area (INE, 2009; CIREN, 2011). By contrast, natural vegetation outside the valleys is extremely sparse…"

Has been replaced in the updated manuscript with the following statements:

> "…but account for less than 1% of the total catchment area (INE, 2009; CIREN, 2011). Most of the annual precipitation, however, occurs as snow during the winter months, leading to an entire dependence on surface-water resources to satisfy crop water needs during the summer. Irrigation water abstractions occur at multiple locations along the river's course depending on both historical water rights and water availability. By contrast, natural vegetation outside the valleys is extremely sparse…"

The following statement on page 12144, lines 22–25 of the discussion paper:

> "Naturalized streamflow time series were estimated using information provided by the Chilean *Dirección General de Aguas*, mainly streamflow measurements at the gauging station of Rivadavia and historical surface-water diversion data."

Has been replaced in the updated manuscript with the following statements:

> "Water abstractions for irrigation were estimated using information on historical water allocations provided by the Chilean authorities. Because these abstractions are likely to influence the hydrological behavior of the catchment during recession and low-flow periods, they were added back to the gauged streamflow in Rivadavia before calibrating the models."

The following statements on page 12164, lines 21–27 of the discussion paper:

> "It was also possible to highlight some errors in the streamflow data. Part of these errors might be associated with uncertainties in the estimation of natural streamflow. Further research is therefore required to better integrate the effect of water abstractions in the hydrological modeling process. From a multiple-hypothesis perspective, the modeling of irrigation water withdrawals should be regarded as a testable model component in its own right."

Have been replaced in the updated manuscript with the following statements:

> "It was also possible to highlight some errors in the streamflow data. The observed streamflow was 'naturalized' by simply adding back the estimated historical water abstractions (Sect. 2.2). When applied on a daily basis, this process inevitably adds some uncertainty to streamflow values because a

significant part of surface-water abstractions actually return to the river system within a few days due to conveyance and field losses. In general, ignoring these return flows would lead to overestimating daily natural flows. In this paper, however, the actual water withdrawals were not known with precision but only as percentages of the nominal water rights – these percentages being fixed on a monthly basis by the authorities to account for variations in water availability. The combined impact of streamflow and precipitation errors on the assessment of structural uncertainty thus remained unknown. Further research is currently underway to integrate the effects of water abstractions and crop water-use in the hydrological modeling process (Hublart et al., 2015; see also Kiptala et al., 2014 for another approach). From a multiple-hypothesis perspective, the modeling of irrigation water water-use should be regarded as a testable model component in its own right."

8. Anonymous Referee #2
Page 12146, lines 10–15: Is not there any seasonality in these processes?

Author's response
The referee is correct to raise the question of seasonal variations in sublimation processes. However, snow cover in the catchment is only present during the winter months and it seemed reasonable, as a first approximation, to assume that sublimation rates remain constant during this period.

9. Anonymous Referee #2
Page 12146, line 22: Do the authors mean that the geological boundaries may be different from the topographic ones?

Author's response
No, that is not what we meant. We hope that the following changes in the updated manuscript will clarify our statement.

Author's change in manuscript
The following statement on page 12146, line 22 of the discussion paper:

"… or a greater contribution of groundwater to surface flow…"

Has been replaced in the updated manuscript with the following statements:

"… or a delayed contribution of groundwater to surface flow from one year to another…"

10. Anonymous Referee #2
Page 12151, line 25: Do the authors wish to refer to Section 2.3.1 instead?

Author's response
Agreed. We apologize for this typo which has been corrected in the updated manuscript.

11. Anonymous Referee #2
Page 12152, line 21: It is unclear how the SCA was modelled given the lumped approach followed here.

Author's response
There seems to be a misunderstanding about how SCA data were used in our study. What was modelled by the snow-accounting options is the total snow water equivalent (SWE) stored in the catchment. As explained on page 12152 of the paper, the SCA data were used "to quantify the error made in simulating the seasonal dynamics" of snow processes "in terms of snow presence or absence" at the catchment scale. The snow error criterion described in Figure 3 corresponds to the number of days when SCA observations and SWE simulations disagree as to whether snow is present in the catchment (no matter 'how much' snow is present). It relies on an indirect comparison between these two quantities.

Author's changes in manuscript
To make this statement clearer, the following sentence on page 12152, lines 22–23:

> "… were used to evaluate the consistency of snow-accounting modeling options in terms of snow presence or absence in the basin"

Has been changed into:

> "… were used to evaluate the consistency of snow-accounting modeling options in terms of snow presence or absence at the catchment scale"

12. Anonymous Referee #2

Page 12154, lines 4–12: I found this choice questionable. Uncertainty bounds should refer to actual nominal values. For example, if one seeks to build 90% confidence intervals, then one should expect that the uncertainty bands contain 90% of the observations, not the maximum of observations. Does it mean here that the authors wish to build 100% confidence intervals? If one wishes to use other confidence intervals, how the approach should be applied?

Author's response

Here it seems necessary to clarify some choices. This paper aimed at assessing model inadequacy and non-uniqueness using a combination of two *non-probabilistic* approaches: a modular modeling framework and a multi-objective optimization scheme. More precisely, we aimed at assessing to which extent structural uncertainty could be reduced by identifying which minimal set of best-performing models maximized the number of observations covered by the ensemble of Pareto-envelopes. This strategy relies on the expectation that a maximum of observations should lie within the overall envelope, provided that structural uncertainty is adequately represented by the ensemble of Pareto-envelopes and that additional sources of uncertainty are negligible. The success or failure of this objective merely indicates to which extent the aforementioned assumptions are correct. Where the objective is to reach X% of the observations (with X<100), one can modify the fourth step of the algorithm detailed on page 12154–12155 by considering only a fraction X/100 of $N_{obs}(N_{max})$ and changing the equality sign on Page 12155, line 4 into a greater-than (or equality) sign.

However, given the non-probabilistic nature of this approach, we have some serious reservations as to whether the resulting simulation bounds can be interpreted in terms of "confidence intervals" or "confidence bands". By comparison with probabilistic methods, multi-objective schemes based on the concept of Pareto-efficiency do not provide any estimate of the residual error variance. The envelopes derived from the sets of Pareto-optimal solutions quantify only the uncertainty arising from the trade-offs between competing criteria and do not have a predefined statistical meaning. This is of course a major drawback of non-probabilistic approaches to uncertainty. However, more probabilistic methods based on the statistical description of model residuals also have their disadvantages. In our opinion, what this comment actually reflects is a common issue in hydrological modeling regarding the definition and assessment of structural uncertainty in probabilistic or non-probabilistic terms. We admit that investigating this issue was far beyond the scope of our study.

Author's changes in manuscript

Because we agree that these assumptions and choices made in defining structural uncertainty bounds may be questionable, we provided more information on their limitations in the discussion part (Section 5) of the updated manuscript. The following statements were inserted:

> "Eventually, the number of models used to represent structural uncertainty was reduced by searching for the minimal set of best-performing structures which maximized the number of observations covered by the ensemble of Pareto-envelopes. It is important to make clear that model inadequacy and non-uniqueness were evaluated here in non-probabilistic terms. In particular, the Pareto-envelopes derived for each model structure quantify only the uncertainty arising from the trade-offs between competing criteria and do not have a predefined statistical meaning (Engeland et al., 2006). Consequently, the overall simulation bounds shown in Figure 8 cannot be easily interpreted as 'confidence bands'. Although discussing the adequacy of non-probabilistic approaches to structural uncertainty was far beyond the scope of this study, it is interesting to analyze the reasons why between 15 and 20% of the observations remained outside the overall simulated envelope in both calibration and validation. To a large extent, this lack of performance can be attributed either to uncertainties in

> the precipitation and streamflow data that were overlooked in this study or to an insufficient coverage of the hypothesis and objective spaces."

Also, we modified the following statements on page 12154, lines 4–12 of the discussion paper:

> "The overall uncertainty envelope should be wide enough to include most of the observed discharge but not so wide that its representation of the various aspects of the hydrograph (rising limb, peak discharge, falling limb, baseflow) becomes meaningless. In general, one will seek to reduce as much as possible the width of the envelope while maximizing the number of observations enclosed within the bounds. In this study, priority was given to maintaining at its lowest value the number of outlying observations before searching for the best combination of models which minimized the envelope area."

These statements have been replaced with the following ones:

> "The overall uncertainty envelope should be wide enough to include a large proportion of the observed discharge but not so wide that its representation of the various aspects of the hydrograph (rising limb, peak discharge, falling limb, baseflow) becomes meaningless. In this study, priority was given to maintaining at its lowest value the number of outlying observations before searching for the best combination of models which minimized the envelope area."

<u>Anonymous Referee #2</u>
I understand that the authors rightly distinguish reliability and sharpness as two expected qualities of the uncertainty estimates, but there are many criteria proposed in the literature to evaluate these qualities. Maybe the authors should use the commonly applied criteria to strengthen the evaluation of uncertainty bounds.

<u>Author's response</u>
We are aware that many other criteria exist in the literature, in particular regarding sharpness. However, choosing one of these over the others seemed quite arbitrary in the absence of preliminary sensitivity analyses, for which we lacked time.

13. <u>Anonymous Referee #2</u>
Page 12156, lines 14–16: It is a bit difficult to see at first glance the structural differences between these three models. The reader has to reconstruct the structures from table 4 and figure 2. Could the authors help the reader here by detailing these differences?

<u>Author's response</u>
We thank the reviewer for his remark, which allowed us to clarify this point in the updated manuscript.

<u>Author's changes in manuscript</u>
The following sentence on page 12156, lines 14–16 of the discussion paper:

> "Models no. 22, 46 and 54, for instance, yield very similar values of the high-flow criterion (Crit1), despite huge differences in their modeling options."

Have been replaced with the following sentence:

> "Models no. 22 (A1–B1a–C3–D2–E1–F2b), 46 (A1–B1b–C3–D2–E1–F2b) and 54 (A1–B1c–C1–D3–E2–F1b), for instance, yield very similar values of the high-flow criterion (Crit1), despite some differences in their modeling options."

14. <u>Anonymous Referee #2</u>
Page 12162, lines 5–6: Was this actually demonstrated here, given there remains similarly performing structures? Besides, I think the usefulness of multi-model frameworks was already demonstrated by past studies. So maybe this should be seen more like a confirmation of existing results.

Author’s response
We agree that “demonstrated” sounds a bit excessive here. It has been removed from the updated manuscript. Regarding the usefulness of multi-model frameworks, please see our answer on Page 1 of this document as well as the modifications provided in the updated manuscript.

15. Anonymous Referee #2
Page 12162, lines 16–22: Can 9-parameter models be considered as parsimonious? The difference between 9 and 13 parameters is not so large, since many modellers may consider 9-parameter models already overparameterized. Maybe this discussion could further refer to past works discussing parsimony in conceptual modelling.

Author's response
We thank the reviewer for helping us to further clarify the important issue of model parsimony in a multi-objective context. Many authors rightly consider that a maximum of 5 to 6 free parameters should be accepted in calibration when using a single objective function. Efstratiadis and Koutsoyiannis (2010) extended this empirical rule to the case of multi-objective schemes by allowing “a ratio of about 1:5 to 1:6 between the number of criteria and the number of parameters to optimize”. For a multi-objective scheme based on four criteria, this would lead to consider 20 to 24-parameter models as still being parsimonious, which, of course, would seem highly unlikely to many modelers. This is because in most cases, as Efstratiadis and Koutsoyiannis (2010) also pointed out, the various criteria used are not independent of each other. In our case, for instance, the information added by the low-flow criterion does not appear so different from that already introduced by the high-flow criterion. By contrast, the snow error criterion really adds new information on some specific snow-accounting parameters. Thus, 9-parameter models should not be regarded as being ‘parsimonious’ in general but only with respect to the number and quality of the criteria used in calibration.

Author’s changes in manuscript
These reflections were included in the updated manuscript to clarify our statement (see Section 5).

16. Anonymous Referee #2
Page 12164, line 1: Would groundwater data be actually helpful in the case of this catchment, given the large uncertainties in precipitation estimates?

Author's response
The reviewer is correct to question the usefulness of additional groundwater data in our case. Additional information on precipitation would be probably far more relevant to improve the reliability of model predictions. This point was made clearer in the updated manuscript.

Author's changes in manuscript
The following words on page 12164, line 1 of the discussion paper:

> “e.g. groundwater levels”

Have been replaced with:

> “e.g. observed snow heights, irrigation water-use”

17. Anonymous Referee #2
Table 1: I do not understand the first equation for snow, which seems larger than P. Maybe remind the option type in the table.

Author's response
We apologize for this typo which has been corrected in the updated manuscript.

Author's changes in manuscript
Please see the modifications made to Table 1 in the updated manuscript.

18. Anonymous Referee #2

Table 2: Where does the range for Kc come from? The ranges given for K3 seem dependent on the option but are the same in the table.

Author's response

The range of values tested for this parameter stem from the following assumption:

$$\mathrm{AE} = K_{\mathrm{veg}}\mathrm{Area}_{\mathrm{veg}}\mathrm{PE} = K_{\mathrm{C}}\mathrm{PE}$$

where K_{veg} is a coefficient which varies between 0 and 1, and $\mathrm{Area}_{\mathrm{veg}}$ is the fraction of land covered with vegetation, which we limited to a maximum of 0.5 given the extreme aridity of the Claro River catchment.

Author’s changes in manuscript

Initially, this explanation was not included in the discussion paper for brevity's sake. In the updated manuscript, we inserted a brief explanation on this point in the caption of Table 2.

Anonymous Referee #2

The ranges given for K3 seem dependent on the option but are the same in the table.

Author's response

We apologize for this typo which has been corrected in the updated manuscript.

Author's changes in manuscript

Please see the modifications made to Table 2 in the updated manuscript.