Referee comment: Impact of seasonal snow on the recharge of a mountain karst aquifer under climate change: the Dévoluy case study (Southern Alps, France)

Rispal et al. present a modelling approach for a karstic catchment with a special focus on the inclusion of snow dynamics. It is based on KarstMod and is a conceptual model. Additionally, the impact of rising temperatures on the snow-discharge dynamics is assessed by using projected temperature data as input data and analyse the changes in discharge. This was done by applying the model to a catchment of the Dévoluy massif.

While I believe that it is important to conduct research on the hydrological behaviour of karst dominated catchments, especially under climate change impact, the presented study could be a bit more ambitious. I feel like an expansion of the study objective while improving its focus would increase its value to the hydrological community.

Major comments:

As already mentioned, the current study design seems to be a bit shallow. While I agree that the study of temperature changes and their impacts on catchment dynamics is relevant, this seems to be done a bit superficial right now. The use of a very conceptual model does not really allow to expand the study to analyse model system states or related aspects. However, what I would think could be interesting in this context is an uncertainty assessment of the model (results). As described in line 286, 10,000 parameter sets yield simulation results that satisfy the objective function criteria. These could be used to conduct the analysis regarding the effect of temperature changes. This would result in a range of potential model reaction behaviour under changing temperatures, letting us make more robust interpretations and deductions from the results. I think this would greatly enhance the study. In the same way, different model structures and their impact on simulated discharge under temperature shift could be included in the analysis — as you mentioned that there are different structures that show equally good model performance.

Also, I dont really see the sense behind the split in model 1 and model 2 in the current study design. If model 1 is without snow routine, meaning that all precipitation occurs as rainfall, then why is a model 2 necessary with snow routine deactivated? From my understanding, the study only needs one model without snow routine and one model with snow routine (both including potential parameter sets / model structures that show satisfactory model performance) to assess the impact of snow inclusion on model performance. This would enhance clarity regarding the methodology.

Lines 74 to 77: I think this is more methodology rather than being relevant for the introduction.

Line 180: I dont think this figure is necessary in the current form. Subfigures a and c really dont offer anything that needs to be depicted as a figures. Subfigures b and d are debatable, if they are really necessary, as the snow depth is shown later within comparisons as well as the discharge. The only unique information here is the discharge for the Souloise river, which is not used later if I am correct?

Lines 204 to 208: Are [L], [L/T], or [/T] supposed to be the units for the corresponding values? Are those something like m^3, litres, km^2?

Line 222: I would suggest writing things like Wobj or anything like that as some sort of variable: W_{obj}, $W_{obj,min}$, E_{min}, P_{sr}.

Line 234: I think to say that you only using one single evaluation crtieria is similar to a multi-objective criteria calibration is difficult, because it's not true. You still only rely on the discharge, so no other aspects like snow height or ETa are explicitly evaluated, which would be a real multi-criteria evaluation.

Lines 237 to 242: These "soft-criteria" mentioned here are surely important, but the calibration (and evaluation) of model parameterisations is only done on the discharge by the KGEnp. Why are those aspects mentioned here not incorporated into the calibration/evaluation?

Lines 247 to 265: Why is the snow routine introduced after model optimization? I think it would make more sense if the order is following a more logic way. First the study area, then the data, then the model and how it works, then how its is calibrated and then how the further study is conducted.

Lines 269: I feel like the figure would benefit from a legend that shortly points out what each abbreviation stands for. Currently it is in the caption, but that makes it less easy to grasp.

Lines 271 to 275: I think this paragraph can be deleted.

Lines 276 to 277: This is more suited for the introduction. No need here.

Lines 286 to 288: Why is only the "best" model retained if there are several parameterisations that seem to be fulfilling the criterion — and therefore show sufficiently good model behaviour? Also, what does "Cross calibration-validation tests did not improve performance" mean here? From my understanding and experience, the sense of a cross-calibration would here be to assess, if model performance is highly infleunced (dominated) by a single year, for example.

Lines 320 to 324: kinda a repetition of what is mentioned in lines 310 to 313.

Lines 324 to 325: Expalantion of positive and negative shift makes no sense in this formulation. It should be formulated more as a relative value that is negative for heights with greater values than the reference and positive when heights are lower than the reference.

Lines 332 to 338: This can be combined with the explanations before to shorten on this, as it seems trivial.

Lines 339 to 342: You said those alternative structures did not *improve* model performance. Does that mean, those are equally feasible? Why are those then not incorporated into evaluation? Model structure uncertainty could be well assessed here and would help increase the robustness of the model result's interpretation.

Line 356: Isn't the comparison of the model with and without snow module the relevant method to assess the impact of snow incorporation on model performance, rather than the inclusion of different temperature changes?

Lines 369 to 373: This is more suited for model description.

Lines 374 to 377: Delete.

Line 380: The figure should be updated accordingly with the changes made to the methodology.

Line 389: The table can probably be moved to the supplementary.

Line 416: Figure numbers are wrong through the whole manuscript from here. Please check.

Line 425: Please check sentence syntax.

Line 428: The model can not provide a more robust performance criteria. Semantically not logic.

Line 443: What sense has a model here that is not validated?

Lines 512 to 513: If these temperature variant simulations are not to be understood as predictive forecasts, why is it framed like that through the manuscript?

Lines 590 to 592: Does the fact, that the scaling parameter for the catchment had to be expanded to 191 km² instead of the actual 150 km² not imply that underground exchanges or a differing aquifer catchment size are very likely possible?

Minor comments:

Line 126: main instead of "mains"

Line 171: are rare and make

Line 193: hydrological

Line 207: the reference unit length is fixed

Lines 249 to 251: Units in []-brackets, multiply dots wrongly formatted.

Line 268: Input liquid precipitation stands for rainfall?

Lines 308 to 310: Formal language.

Line 330: Range

Line 331: unit formatting

Line 359: "climate models already show warming"

Line 364: ambiguous formulation regarding the temperature definition. Do you mean you use the temperature recommendations to define the simulated temperature increases?

Line 486: comma missing