Reply to Referee 2 in black :

Authors : We thank the Referee for the careful reading of our manuscript and for the positive overall assessment of the modelling approach and its clarity. We appreciate the constructive suggestions provided, particularly regarding the quantitative evaluation of model performance and the improvement of result visualization. We agree that including additional performance metrics and adapting some figures to facilitate comparison between simulations will strengthen the robustness and readability of the manuscript. We propose to revise the manuscript accordingly and respond to each point in detail below.

In this reply, we have numbered each comment from the referee using the notation R2 (Referee 2) followed by the comment number. Our response begins with the label "Authors:".

R2.1: Overall, I found the models and methods were well described. The authors did a good job justifying the inclusion of snow in the karst discharge simulation under multiple climate-change temperature scenarios. Overall, I found the manuscript well written and clear. Thank you for letting me review this research.

Overall, there are a few major requests I have:

1. Greatest recommendation: the research needs to include RMSE, MAE, bias and/or other metrics between the simulations and in-situ data to calculate and quantify the amount of model error. These will also need associated visualization for the reader to ensure that the model statistically performed well.

2. I recommend aggregating the time series (e.g. figure 4-6) into bar charts and/or box and whisker, etc. so the reader can easily compare the simulation results.

Authors: These recommendations are taken into account and explained in responses to comments in the following document.

Please see my comments below:

Abstract:

R2.2: Line 18: At the end of this sentence, "Results show that accounting for snow processes is essential to reproduce the observed discharge dynamics" - I think it would be clearer to add "in karst environments"

Authors: We will add "in karst environment" to be more precise as you suggested.

Introduction:

R2.3: Line 57: The paragraph on hydrologic modelling as an efficient tool and the models that incorporate snow as a parameter come before the next paragraph that explains the importance of incorporating snow into karst models. I think these two paragraphs should be swapped – e.g. paragraph at line 57 should go at line 46.

Authors: Following your recommendation, we will switch these two paragraphs to provide a more logical progression in the introduction.

Methods:

Data

R2.4 : 137: It is unclear how "precipitation (mm/day) and mean temperature (°C)" factor into the sentence. Are you trying to say that precipitation (mm/day) and mean temperature (°C) also have the four grid meshes with 8x8 spatial resolution, etc.? If so – then this needs its own sentence – the format of this sentence leaves the reader confused. Please clarify.

Authors: We will split the sentence in two to make it clearer that both temperature and precipitation are computed from the average value of the four selected meshes.
"For each variable, the average value of the four selected meshes was computed and used as daily input to the KarstMod model from 2015 to 2019. The meteorological inputs consist of precipitation (mm/day) and mean air temperature (°C)."

Accuracy Assessment:

R2.5 Please include an RMSE, MAE, bias for evaluating the accuracy of the simulated results compared to the in-situ results. The KGEnp mentioned in Line 411 is one of the few mentions of model efficiency. I would like to see more on model accuracy and bias. Please include figures to visually demonstrate these results.

Authors: KarstMod provides several tools to assess and visualize model performance, including NSE, KGE, BE, VE, RPearson, and RSpearman. In response to your comment, we have expanded the evaluation of model performance by adding these additional criteria and calculated also RMSE, MAE and BIAS. We will report them in Table 1. Also we will update the figures in the article to show a range of results rather than only the result of the calibration with the maximum objective function. This was also a request from referee 1. We used KarstMod to obtain 5000 satisfactory results (for Wobj = KGEnp > 0.85), and we add to the figures the median value and the 5th and 95th percentile values."

Results:

R2.6 Figure 4: the pink/blue in Fig 4A and yellow/gray are hard to discern (e.g. the colors overlap). Can you try more high contrast line colors?

Authors: We will change the contrast in these colors to make de figure more discernable (Fig. 6 at the end of this document).

R2.7 The graphics don't line up with the text. For example, line 470 mentions Figure 8e, then the next figure is Figure 6. Please put the figures after the sections where they are first mentioned.

Authors: There was an error in figure numbering starting from line 430. We apologize for this typing error. This has now been corrected. The figure citations in the text were already correct, but the figure numbering itself was incorrect, which caused the apparent mismatch. We will fix this issue and ensure that all figures now appear in the correct order and after the sections where they are first mentioned.

R2.8 I find the time series difficult to interpret. I think it would be easier if you included aggregated final results for the readers such as:

Authors: This figure, in its current form, allows the visual comparison of the Model 1 and Model 2 curves, showing the fact that Model 1 does not correctly represent discharge during key periods influenced by seasonal snow, and that Model 2, with its snow routine, allows a better result to be obtained during these periods. Using box plots based on the calibration/validation period and mixing all discharge values together would not show this mismatch observed on the curves, as the model nonetheless achieves a good overall Wobj criterion, meaning that over the four-year periods simulated, volumes are quite close to observed ones. We therefore choose to keep Figure 4 (now Figure 6) with the curves, adding the modifications suggested by Referee 1. However, following your recommendation to include a figure summarizing the results, we have used a violin chart to illustrate the results by month (see Fig.10 at the end of this document - this figure will be included in the revised paper).

R2.9 Figure 5: I recommend a similar aggregated visual that makes it easy for the readers to compare. Also, make the caption clearer to define was E-S, M-S, and C-S mean (e.g. explain it – the reader needs more than a copy/paste from the visual).

Figure 5 is not intended to compare different models but to illustrate internal variables of the model 2, such as water levels and discharge within each compartment. Therefore, an aggregated representation similar to Figure 7 is not straightforward or necessarily meaningful for these time series. However, we propose to clarify the caption and now explicitly define the meaning of E-S, M-S, and C-S to improve readability. Please note that the corrected number of this figure will now be Figure 7.

R2.10 Figure 6 and 8 – same comment – aggerate so the reader can easily compare.

Authors: The curves shown in Figure 7 (previously Figure 6) have been aggregated on a monthly basis in Figure 9 (previously Figure 8). The corresponding results have been updated to address Referee 1's suggestion regarding the assessment of model results uncertainty. We chose to keep Figure 7 because it allows a visual inspection of discharge evolution across the four modeled years for the different scenarios. Following Referee 1's recommendation, we decided to remove model2_SRoff, the version without the snow routine, from these comparisons. Please find the modified figure at the end of this document.

For example, Figure 7 was easily digestible to me because I could make side-by-side comparisons.

R2.11 I think you should reference Figure 4 in the Figure 8 caption so readers can easily reference the different model simulations.

Authors: We propose to update the caption of Figure 8 to reference Figure 4, allowing readers to easily relate the different model simulations as suggested.

R2.12 Figure 6A: I find this unimportant – a reader could just as easily find that in the text – or you could put it in the caption, but the bar at the top seems out of place and unnecessary.

Authors: We agree that Figure 6A is not essential, we propose to remove the top bar to simplify the figure.

Authors: We have added the accuracy assessment values to support the conclusion that including snow improves the model results.

Modified figures (to be included in the revised paper):

Please find below the modified figure and the caption for Figure 6; Figure 8 and Figure 10. Please note that. Here the 5000 simulations satisfying W_{obj_min} = 0.85 in Model2 were kept and used for the temperature scenarios.
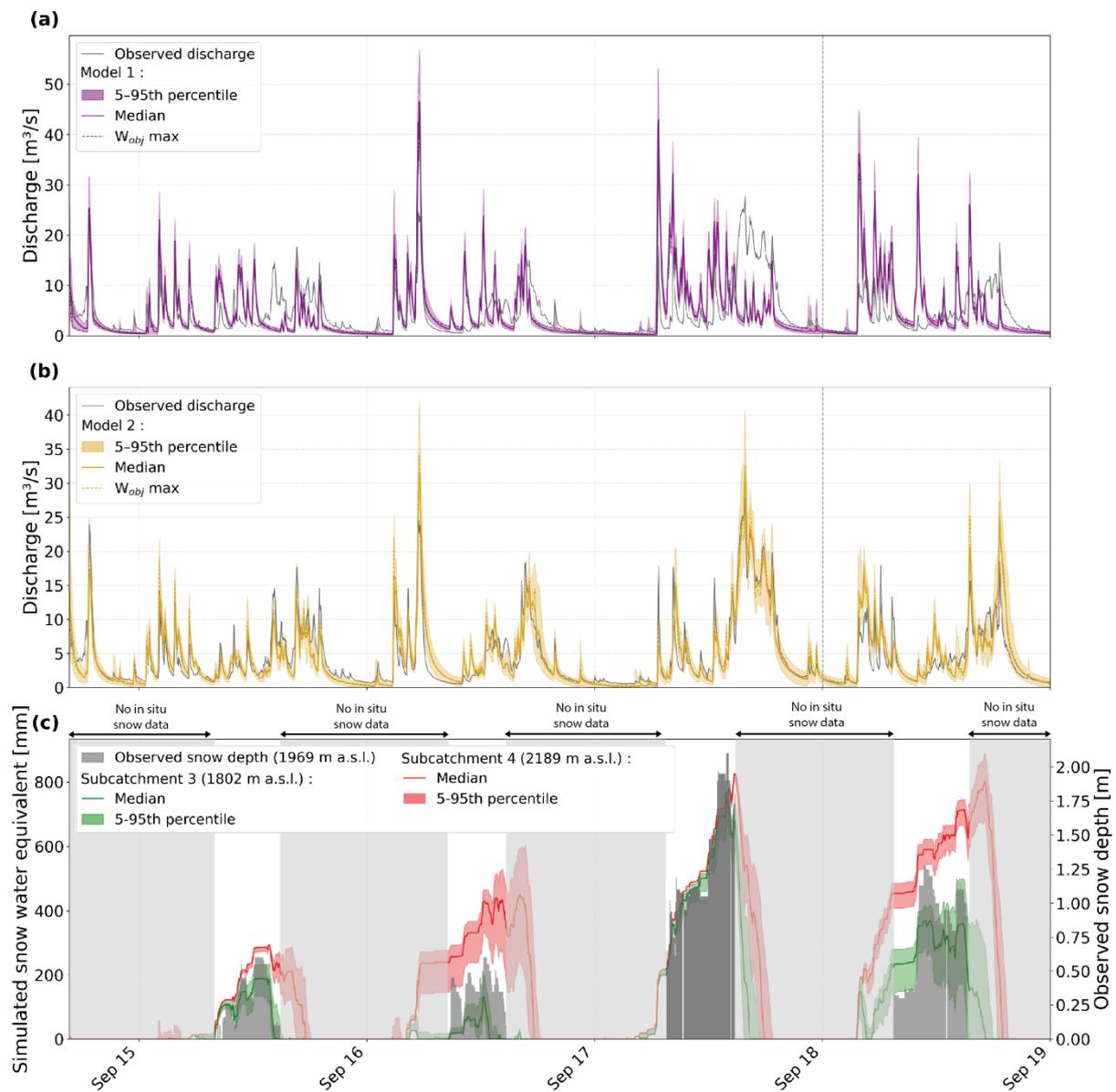

Figure 6: Observed and simulated discharge for the Dévoluy karst system (2015-2019) for (a) Model 1 and (b) Model 2. Shaded areas show the 5th-95th percentile of the 5000-parameter set, solid lines the ensemble median, and dashed lines the W_{obj_max}

simulation. (c) Model 2 simulated snow water equivalent for subcatchments 3 and 4 (shaded: 5th-95th percentile) and observed snow depth at the Super-Dévoluy station (grey bars); gaps indicate periods with no available measurements.
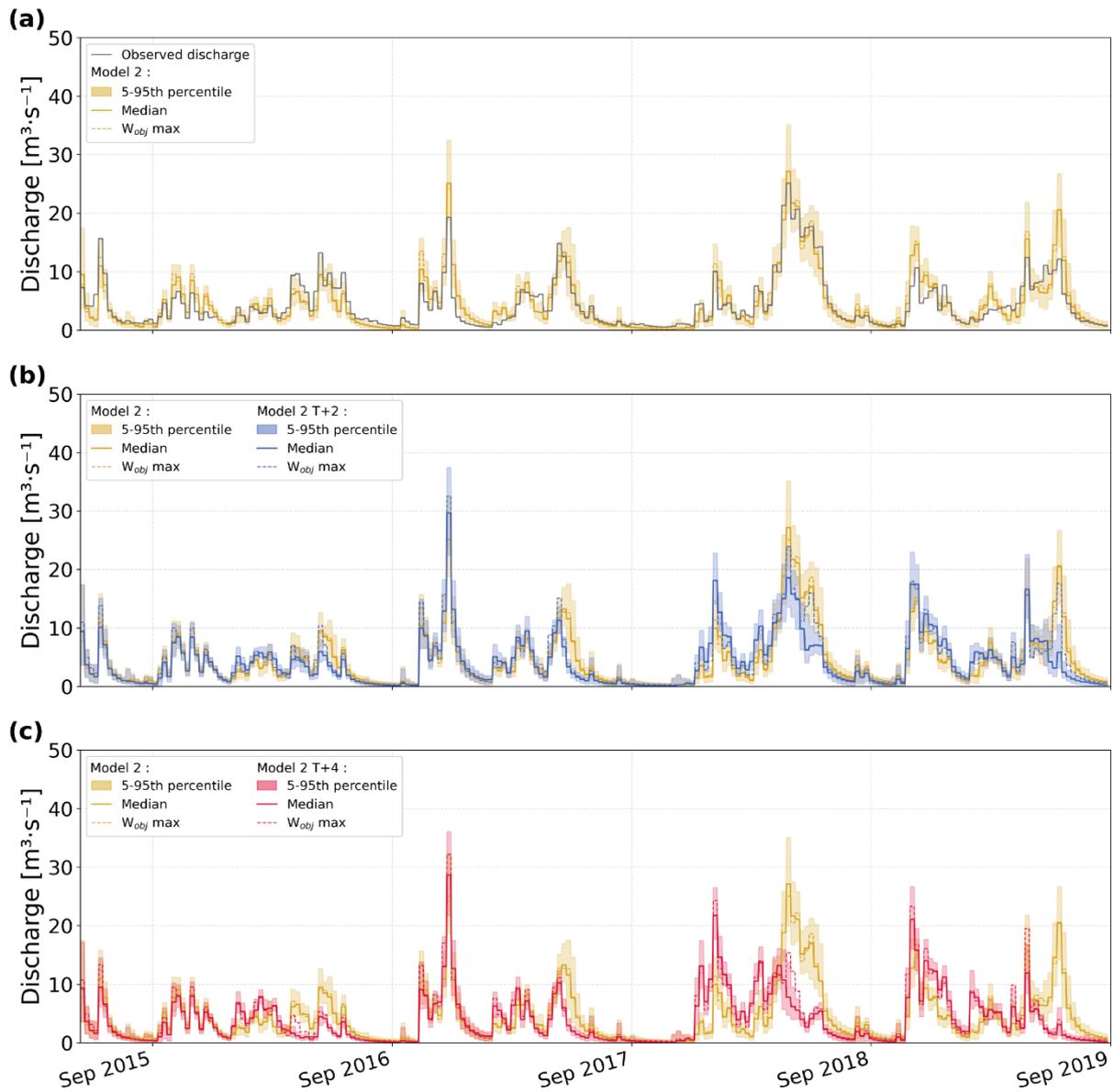

Figure 8: Seven-day average simulated and observed discharge for the Dévoluy catchment over the 2015–2019 period. (a) Model 2 and observed discharge. (b) Comparison of Model 2 and Model 2 T+2°C simulations. (c) Comparison of Model 2 and Model 2 T+4°C simulations. Shaded areas represent the 5th-95th percentile envelope of the 5000-parameter sets. Solid lines show the ensemble median and dashed lines the simulation with the parameter set corresponding to the maximum objective function value (W_{obj_max}).
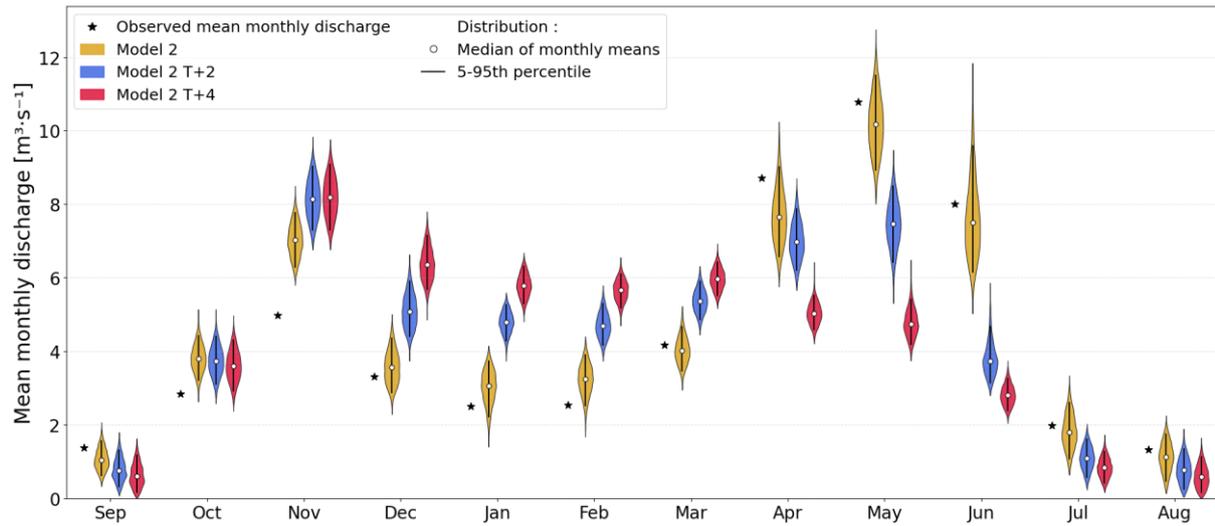
Figure 10: Mean monthly discharge distributions for the Dévoluy karst system over the 2015-2019 period, shown for each calendar month (September to August). For each month, the star symbol indicates the observed mean monthly discharge. Violin plots show the distribution of mean monthly discharge across the 5000-parameter ensemble for Model 2 (gold), Model 2 T+2°C (blue), and Model 2 T+4°C (red). The vertical bar and white dot indicate the 5th-95th percentile range and the median of the 5000-parameter ensemble.