The authors evaluate the perfomance of a 12-km RCM and a 2.5-km CPRCM with respect to reproduction of primarily observed precipitation (rainfall) and discharge extremes in Quebec. Is is found (claimed) that the CPRCM better reproduces both types of extremes. While the topic is relevant and interesting, the methods appear overall well selected and applied, the presentation is neat, I cannot recommend publication of the manuscript in its present form. In the following I will explain why.

General comments

- **The paper tries to argue that the CPRCM does a much better job than the RCM, but this is, in my view, not (well) supported by the results. Sometimes it is clear that the CPRCM greatly overestimates the extreme rainfall (Fig. 3a, Fig. 4a), but this is basically neglected. Sometimes argumentation is based on visual inspection (Figs. 6-9) or a few numbers (Table 3) but without testing whether differences are statistically significant (which is really important, especially as you sometimes work with very small data sets). OK, it may be that the CPRCM does a (significantly) better job than the RCM, but the current results do not prove that.**

We thank the reviewer for this comment. Our response has two parts: Part 1 addresses the need for additional quantitative assessment, and Part 2 provides explanations regarding the apparent overestimation of precipitation by the CPRCM relative to station observations.

Part 1:

We agree that additional quantitative assessments and statistical tests are necessary to better demonstrate the significance of the differences between CPRCM and other RCM simulations. To address this point, we conducted additional analyses focusing on extreme precipitation (99.99th percentile) at station locations. First, we compared model outputs with station observations using standard performance metrics, RMSE, mean bias, and correlation, to quantify how well each model reproduces extreme precipitation during the summer and fall seasons. Station-based estimates were compared with the corresponding values extracted at the same grid points from the model simulations (CRCM6/GEM5 at 12 km and 2.5 km resolution) and from the reanalysis dataset (RDRSv2.1). The RMSE was computed using values from all stations, while the mean bias was calculated as the average of station-wise biases for each dataset. Correlation coefficients were also computed using values from all stations.

Overall, CRCM6/GEM5 at 2.5 km resolution shows the best performance across all metrics (see Tables below). RMSE values are lowest for CRCM6/GEM5–2.5 km, even lower than those obtained from the reanalysis dataset. In addition, CRCM6/GEM5–2.5 km exhibits a smaller mean bias relative to station-based extreme precipitation estimates. Furthermore, results show a higher correlation with station data during summer. In fall, all datasets exhibit low correlation.

	RMSE (mm)	
	Summer	Fall
CRCM6/GEM5-2.5KM	**5.63**	**14.28**
CRCM6/GEM5-12KM	13.42	15.72
RDRSv2.1	8.74	14.87

	averaged bias	
	Summer	Fall
CRCM6/GEM5-2.5KM	**0.06**	**0.11**
CRCM6/GEM5-12KM	-0.47	-0.27
RDRSv2.1	-0.24	-0.16

	Correlation	
	Summer	Fall
CRCM6/GEM5-2.5KM	**0.37**	-0.01
CRCM6/GEM5-12KM	0.32	0.01
RDRSv2.1	0.25	0.06

Considering previous assessments, we formally evaluate whether the differences between the two model configurations are statistically significant by applying the

Wilcoxon signed-rank test to the station-based 99.99th percentile of hourly precipitation, comparing the two simulations. These additional analyses allow for a more rigorous assessment of the added value of the CPRCM relative to the coarser-resolution RCM and strengthen the interpretation of the results. The table below presents the results of the Wilcoxon signed-rank test (significance level of 5%, p-value = 0.05). The null hypothesis states that the median of the differences between the 99.99th percentile hourly precipitation simulated by CRCM6/GEM5 at 2.5 km and 12 km resolution is zero, implying no systematic difference between the two resolutions. The test results (H = 1) indicate rejection of the null hypothesis. Specifically, the null hypothesis is rejected for both summer and fall, demonstrating that the differences between the two simulations are statistically significant in both seasons.

	p-value	H
Summer	0.00	1
Fall	0.00	1

Part 2:

Nevertheless, we do not fully agree that CPRCM greatly overestimates summer values, as suggested by Figs. 3a and 4a. In Fig. 3a, the CPRCM histogram matches the station data reasonably well. It is true that the vertical line representing the 99.99th percentile is substantially higher than the observed one; however, this likely results from estimating an empirical 99.99th percentile, which is highly sensitive to a small number of extreme values in the tail of the distribution. In addition, the station shown in the paper is only one example. When considering other stations (with histograms provided in the supplementary material), the difference between the station-based and CPRCM-based 99.99th percentiles is not consistently large. Furthermore, the RMSE and mean bias of the 99.99th percentile indicate that CPRCM has a lower RMSE (i.e., better performance) than the other datasets, as well as a smaller bias. Regarding Fig. 4a, some overestimation is also present, but its magnitude remains limited.

- **Some very important aspects are not taken into account, notably the impact of spatial resolution and the impact of climate model bias, both having a huge impact on both rainfall and discharge (extremes). They are briefly mentioned but without being investigated, which makes the significance of the results virtually impossible to judge, in my opinion.**

We thank the reviewer for this important comment. We fully agree that both spatial resolution and climate model bias can strongly affect simulated precipitation and discharge, particularly for extremes.

The impact of spatial resolution is investigated by comparing simulations performed with CRCM6 at 2.5 km and 12 km, driven by the same large-scale forcing and using consistent physical parameterizations. Differences between the two simulations therefore primarily reflect the effect of spatial resolution and the representation of convective processes.

To make this point clearer, we will revise Section 5 to explicitly discuss the role of spatial resolution and its expected influence on extreme precipitation intensities and spatial coherence.

We agree that climate model biases play a crucial role in the simulation of precipitation and discharge extremes. Numerous studies have shown that increasing spatial resolution, particularly toward convection-permitting scales, tends to increase the intensity of short-duration precipitation extremes due to an improved representation of convective processes and reduced reliance on convective parameterization (e.g., Prein et al., 2015; Kendon et al., 2017; Ban et al., 2021). This increase in extreme precipitation intensity is often associated with a reduction of the well-documented underestimation of sub-daily extremes in coarser-resolution regional climate models, although biases are not fully eliminated. In this study, we therefore interpret differences between the 2.5 km and 12 km simulations primarily as resolution-related effects, while acknowledging that residual model biases remain.

The second paragraph (line 546) of section 5.1 will be replaced by following paragraph:

'*The comparison of precipitation intensity distributions and spatial patterns (Figs. 3 and 4) highlights a key limitation of coarser-resolution climate models, namely their tendency to underestimate short-duration, high-intensity rainfall due to an inadequate representation of fine-scale convective processes. In contrast, CRCM6/GEM5-2.5 km better captures both the intensity and localized nature of* *extreme precipitation, particularly in summer when convective storms dominate, consistent with previous studies showing that finer spatial resolution improves the simulation of convective systems (e.g., Ban et al., 2014; Prein et al., 2015, 2020; Kendon et al., 2014; Ban et al., 2021). The larger extremes simulated at 2.5 km should not be interpreted as the absence of model bias; rather, convection-permitting models are known to produce higher short-duration precipitation intensities partly due to reduced spatial smoothing and improved process representation, which often alleviates, but does not eliminate, the underestimation of sub-daily extremes in coarser models. Because both simulations use the same large-scale forcing and model framework, the observed differences are mainly due to resolution effects rather than independent model biases. However, residual biases in the absolute magnitude of extremes may still affect hydrological impact assessments. Consequently, while the 2.5 km configuration demonstrates clear added value for applications such as flood risk analysis, future work using bias-corrected simulations or multi-model ensembles would help further disentangle the respective roles of spatial resolution and model bias.'*

- **The text is overall well written, but with far too much of superfluous information (things that one must assume is known to HESS readers) and repetition. To me it has the feel of a (good) student essay, but not on the level of a scientific paper in HESS. Some examples:**
  - **The description of GCMs and RCMs in the introduction (40-55) is known to readers.**
  - **Readers know how KGE works (248-250), what a CDF is (296-299), how box plots should be interpreted (447-451), and other similar examples.**

Thank you for these fair points. The text will be revised to reduce repetition by shortening some explanations and making the discussion more concise and straightforward.

  - **Section 5 (Discussion) is in my view mainly a rather long summary of the study, mainly consisting of repetitions and with few real conclusions, other than general statements.**

We thank the reviewer for this suggestion. We agree that the original version of Section 5 contains a substantial summary of the paper. To improve clarity and focus, lines 568-571 will be replaced by following paragraph:

'*The use of a lumped hydrological model in this study was a deliberate and conservative choice aimed at maintaining the methodology simple and enhancing the reproducibility. By design, lumped models do not explicitly take advantage of the spatial organization of precipitation, and therefore the fine-scale rainfall structures simulated by convection-permitting climate models (CPRCMs) are not directly translated into the runoff generation process. Consequently, the hydrological response primarily reflects differences in precipitation intensity and temporal variability, while spatially localized convective extremes are averaged at the catchment scale when used in a lumped model. This aggregation may attenuate or even mask potential CPRCM added value, and likely contributes to cases where the 12 km simulation yields higher KGE values than the 2.5 km simulation. Moreover, the fact that hydrological improvements are nevertheless observed using CPRCM outputs within a lumped modeling framework indicates that these results represent a lower-bound estimate of the added value of high-resolution climate simulations. This suggests that greater gains in the simulation of extreme flows could be achieved in future studies through the use of distributed hydrological models, such as Hydrotel (Fortin et al., 2001), which are better suited to exploit spatial rainfall heterogeneity. Overall, these findings underscore that hydrological added value depends not only on atmospheric model resolution, but also on the consistency between climate and hydrological model structures, and that the limitations of the lumped approach should be explicitly considered when interpreting improvements in flood simulations*.'

Specific comments:

- **164: Is this supposed to be a separate section?**

Thanks for your attention, yes, it will be corrected.

- **270: 18 annual maxima is not much to work with, consider peak-over-threshold instead.**

.This is a valid point. To increase the sample size for the analysis, a peak-over-threshold (POT) approach will be adopted. In the POT method, a minimum separation of three days between consecutive peak events will be imposed to ensure the independence of the extracted events.

- **315: Should be (Fig. 3a, 3c, 3f), I think.**

This is correctly mentioned, since it was supposed to refer to the summer season plots for both CRCM6/GEM5 configurations.

- **358-363: There is no need to repeat the figure caption in text, same happens also elsewhere. And how to visually interpret the match bewteen histograms is trivial. In addition, you do not need to explain the colour legends in the captions; we understand from the panels.**

This is a fair point, it will be corrected in revised version

- **454: One “generally” too much here.**

Thanks, the first “generally” will be removed