The Authors have addressed all comments from the previous review round and adopted some of the suggestions. I kindly ask them to address the comments below, most of which are based on the replies to the previous review round:

We thank the reviewer for their comments. We have added a discussion to the "study limitations" section, please refer to lines 875-901 of the revised manuscript. The reviewer's comments are in black*, our responses are in blue italic fonts, and the added manuscript text are* in blue plain text.

1) Thank you for addressing this comment.

2) The authors have included in their reply of this comment the difference between forecast and projection. However, both weather-forecast and climate-projections should be accompanied by a probability of occurrence, since they both include expectations (or in other words, a statistical estimation of the expected/mean value of an envelope of events). In high-complex systems (such as climate dynamics), this can be achieved by a sensitivity analysis of the input parameters in order to create the envelope of different possible projections to see "what might happen if the world warms by +4-5 oC", as for example, for the RCP8.5 scenario that the authors use (e.g., see, for example, recent discussions in Schwalm et al., 2020; Hausfather and Peters, 2020). Therefore, a sensitivity analysis of the input parameters is required to create such an envelope of events. For example, the comparison between the observed and simulated river stages, groundwater levels and snow cover at the 3 stations indicated in Figures R1a-d, illustrate that (with the exception of snow cover) there is a large difference between them. However, if the Authors perform a sensitivity analysis on the several input parameters of their model, they would see that the cloud of simulations is expected to cover the observed records and to also assign a probability of occurrence to the end of century expectations.
Moreover, the high-dependence of the simulations on the initial conditions is a known issue in atmospheric dynamics and thermodynamics. One of the first studies (there are many since then) is by Lorenz (1963), who discovered the existence of the phenomenon of chaos in the atmospheric dynamics and the high-dependence on the initial conditions, which increases with time. Please note that only the statistical estimations, as for example the expectation/mean of an envelope of events, is expected to be independent to initial conditions, but this requires a sensitivity analysis and not the application of a deterministic model with fixed values of the input parameters and initial conditions.
I understand it is difficult for the authors to perform such a huge sensitivity analysis at this stage, and so, this is why I kindly ask to at least discuss the issue of uncertainty and variability on the input parameters and initial conditions, and thus, on the simulated output.

We agree with the reviewer that due to the uncertainties associated with the numerous input parameters required by both the climate and hydrological models, a sensitivity analysis is ideal. However, these models are highly computationally expensive which makes the sensitivity analysis difficult to undertake. Indeed, it takes more than 24 clock hours to perform a single year simulation with the integrated hydrologic model using 320 cores on a supercomputer. Because the model requires a lot of parameters, a full exploration of the parameter space to perform a rigorous sensitivity analysis may require thousands runs of the model making such analysis *infeasible with the current available computational resources. We acknowledge this limitation in the revised manuscript by adding the following statement:*

"In this study, we relied on deterministic models to represent both the atmospheric (VR-CESM) and hydrologic (ParFlow-CLM) dynamics. These models are very sensitive to the initial conditions and input parameters (La Follette et al., 2021; Lehner et al., 2020; Song et al., 2015) which are uncertain given the lack of data characterizing the above and below-ground environment, including its hydrological response. Thus, while it is important to assess the sensitivity of the model outputs to these uncertain parameters, these models are computationally expensive and require many parameters. For example, a complete sensitivity analysis of the hydrologic model requires running it thousands of times to explore the full parameter space (which has a dimension of over 29). Such an approach is not feasible with the currently available computational resources because it takes longer than one wall-clock day to simulate a single water year for a single model parameterization, even in a high-performance computing environment. Future work could employ reduced order models based on a subset of the physics-based model runs to explore parameter space further (e.g., Maina et al., 2022)."

In addition, the sensitivity of the model outputs to the initial conditions could be reduced by performing a spin-up which consists of running the model multiple times repetitively until an equilibrium is reached. We have used this approach to develop our model.

3) The so-called Long-Range Dependence (LRD; or else known as long-term persistence or the Hurst phenomenon; Hurst, 1951) was the first study that gave a justification of why the so-called clustering of events (i.e., the tendency of wet/dry years to occur together in a non-predictive manner forming clusters) may occur in natural processes. Earlier, and independently, Kolmogorov (1940) introduced the so-called fractional-Gaussian-noise (fGn) model that was able to simulate this clustering behaviour. Since then, there are many developments towards the simulation of the LRD in natural processes and atmospheric dynamics (the literature is huge; see, for example, studies by Mandelbrot Wallis, 1968; Klemes, 1974; Tsonis, 1999; Dimitriadis et al., 2021, etc., which all include many references on this issue and identify the LRD in key hydrological-cycle processes - like streamflow, precipitation, air temperature, specific humidity, atmospheric pressure, wind speed, solar radiation, evapotranspiration, etc.- from analysis of thousands of stations and billions of records). The impact of the LRD behaviour on the atmospheric dynamics extends to the climatic scale (over 30-years), and so, it is not enough to show a validation of a 5 years with the model but rather to validate the model in an over-30-years window. I understand that it may be difficult to perform such analyses at this stage, however, I kindly ask the authors to at least discuss the impact and need for preservation of the LRD behaviour in the atmospheric dynamics model. This is in line with the previous comment, since the LRD is a stochastic attribute that arises from the intrinsic uncertainty of the chaotic behaviour apparent in atmospheric dynamics and thermodynamics models.

We have now included in the "study limitations" section, these limitations of our approach based on deterministic models. Such approaches are embedded with uncertainties that must be assessed; however, given the computational demand of these models this is not currently feasible. Please see below the added paragraph.

"In addition, because of the behavior of hydrological processes, the climate variability, and the uncertainties of deterministic models, model validation should ideally be performed over a long period to account for different changes and variabilities. In this study, model validation was limited to a period of 5 years due to computational constraints. Although this period encompasses the wettest and driest years on record in the region, we acknowledge that it may not be sufficient to capture the full range of hydrological variability. Another limitation of using deterministic models is that the temporal variations of hydrological processes tend to follow a stochastic behavior in accordance with the so-called Hurst phenomenon (Hurst, 1951; Koutsoyiannis, 2003). As a result, the use of deterministic models such as the ones employed in this study could intensify the impacts of hydrological extremes and climate change."

4) Thank you for addressing this comment. From the provided Tables, it can be observed that there are in total 29 input parameters (8 for the hydrodynamic properties based on the geology, 12 for the surface roughness and crop properties based on land use, and 9 for the numerical model set-up) and 5 output variables.

Yes. Because the model requires a lot of parameters, exploring the parameter space with 29 dimensions is not feasible with the current available computational resources. We now describe this in the study limitations section of the revised manuscript.

5) Please consider further discussing the effect of groundwater and present both scientific opinions. It is mentioned by the Authors that the simulations without pumping do not significantly change the observed dynamics of the system; however, the overexploitation of groundwater is considered an important anthropogenic impact at the local and global climate and the hydrological cycle with a visible effect on sea level rise (see, for example, a recent analysis by Koutsoyiannis, 2020).

We have added the following statement in the revised manuscript to discuss the effects of pumping:

"Finally, it has also been demonstrated that while the changes in water balance exhibit greater variability on climatic scales, the most important changes in hydrologic processes remain the overexploitation of groundwater (Ferguson and Maxwell, 2010) which has an impact on the rise in sea level (Koutsoyiannis, 2020). In addition to projecting the use of groundwater by the end of the century, future studies could compare the two approaches (deterministic and stochastic) to better assess the limitations and the uncertainties associated with them.".

References

Dimitriadis, P., D. Koutsoyiannis, T. Iliopoulou, and P. Papanicolaou, A global-scale investigation of stochastic similarities in marginal distribution and dependence structure of key hydrological-cycle processes, Hydrology, 8 (2), 59, doi:10.3390/hydrology8020059, 2021.

Hausfather, Z., and G.P. Peters, RCP8.5 is a problematic scenario for near-term emissions, Proc. Natl. Acad. Sci. USA, 117, 27791–27792, 2020.

Hurst, Long-Term Storage Capacity of Reservoirs, Trans. Am. Soc. Civ. Eng., 116, 770–799, 1951.

Klemes, V., The Hurst phenomenon: A puzzle?, Water Resour. Res., 10 (4) 675-688, 1974.
Kolmogorov, A.N., Wiener spirals and some other interesting curves in a Hilbert space, Dokl. Akad. Nauk SSSR, 26, 115–118, 1940.

Koutsoyiannis, D., Revisiting the global hydrological cycle: is it intensifying?, Hydrology and Earth System Sciences, 24, 3899–3932, doi:10.5194/hess-24-3899-2020, 2020.
Lorenz, E.N., Deterministic nonperiodic flow. Journal of the Atmospheric Sciences, 20 (2), 130–141, 1963.

Mandelbrot, B.B. and J.R. Wallis, Noah, Joseph and operational hydrology, Water Resour. Res., 4, 909–918, 1968.

Schwalm, C.R., S. Glendon, and P.B. Duffy, RCP8.5 tracks cumulative CO 2 emissions, Proc. Natl. Acad. Sci. U.S.A., 117, 19656–19657, 2020.
Tsonis A.A., P.J. Roebber, and J.B. Elsner, Long-range correlations in the extratropical atmospheric circulation: origins and implications, J. Clim, 12, 1534–41, 1999.

Dimitriadis, P., D. Koutsoyiannis, T. Iliopoulou, and P. Papanicolaou, A global-scale investigation of stochastic similarities in marginal distribution and dependence structure of key hydrological-cycle processes, Hydrology, 8 (2), 59, doi:10.3390/hydrology8020059, 2021.