Dear authors,
Thank you for the excellent and comprehensive review. Included below are a few minor points that should be addressed before publication.
We thank the editor for the thorough and useful review. Please find the comments below, with our responses in red.

Both reviewers have requested a discussion of the sensitivity of the threshold choices. Thank you for the additional analyses. By including longer SPI time-scales, you are including other types of droughts. Please add a short discussion to this end (see for example information here: `https://drought.emergency.copernicus.eu/data/factsheets/factsheet_spi.pdf`). Thank you for the useful comment. Indeed, it is important to underline the different nature of the droughts that we obtain using different accumulation periods, and also their different predictability. We have added a short discussion on the topic at the end of Section 4 (line 326), as well as two sentences in the method part (line 122), specifically describing how the 3-, 6- and 12-months accumulations describe different types of droughts with different impacts, as well as possible effects that the accumulation period has on the predictability of this kind of events.

In the lead time dependent calculation of the SPI - do you see any changes in the underlying distributions? We have computed the mean and the standard deviation of the distributions for the calculation of SPI and SPEI (accumulated precipitation and water budget), for every lead month which was included in the study. We have then plotted the results for the initial (lead year 2) and final (lead year 5) leadyear. We then proceeded to subtract the values of lead year 5 from the values of lead year 2, to highlight the differences between the different lead times. We show below as an example, the result for the model HadGEM3-GC31-MM. In general, we find that the differences between lead year 2 and lead year 5 are both model- and month-dependent. For example, as it can be seen in the figures below, most of the models show a positive difference for equatorial South America, except CanESM5, which shows negative differences. It also appears that stronger differences, both in mean and standard deviations, appear between the months of January and June. These differences can be attributable to a remaining effect of the drift in decadal predictions, despite computing the distributions on a lead time level. However, all of the results show low differences compared to the absolute values of the mean and the standard deviations (not shown here), except in limited regions or specific gridpoints. In Section 2.1 (line 107), we have added a paragraph explaining the results of these analyses.

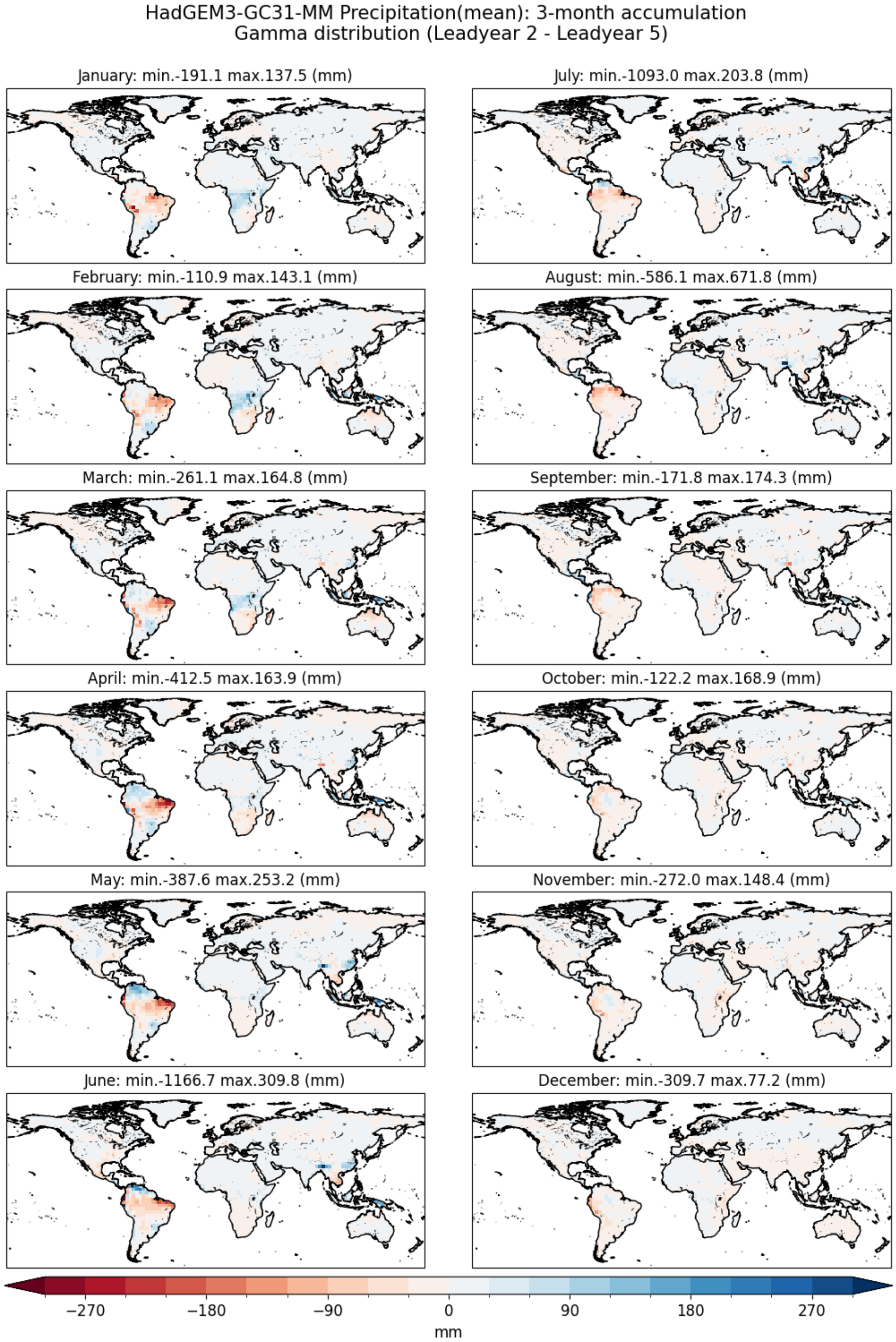

Figure 1: Difference between the means of the lead month distributions for lead year 2 and lead year 5 calculated from accumulated precipitation (SPI). Values in the titles indicate the minimum and the maximum values of these differences.

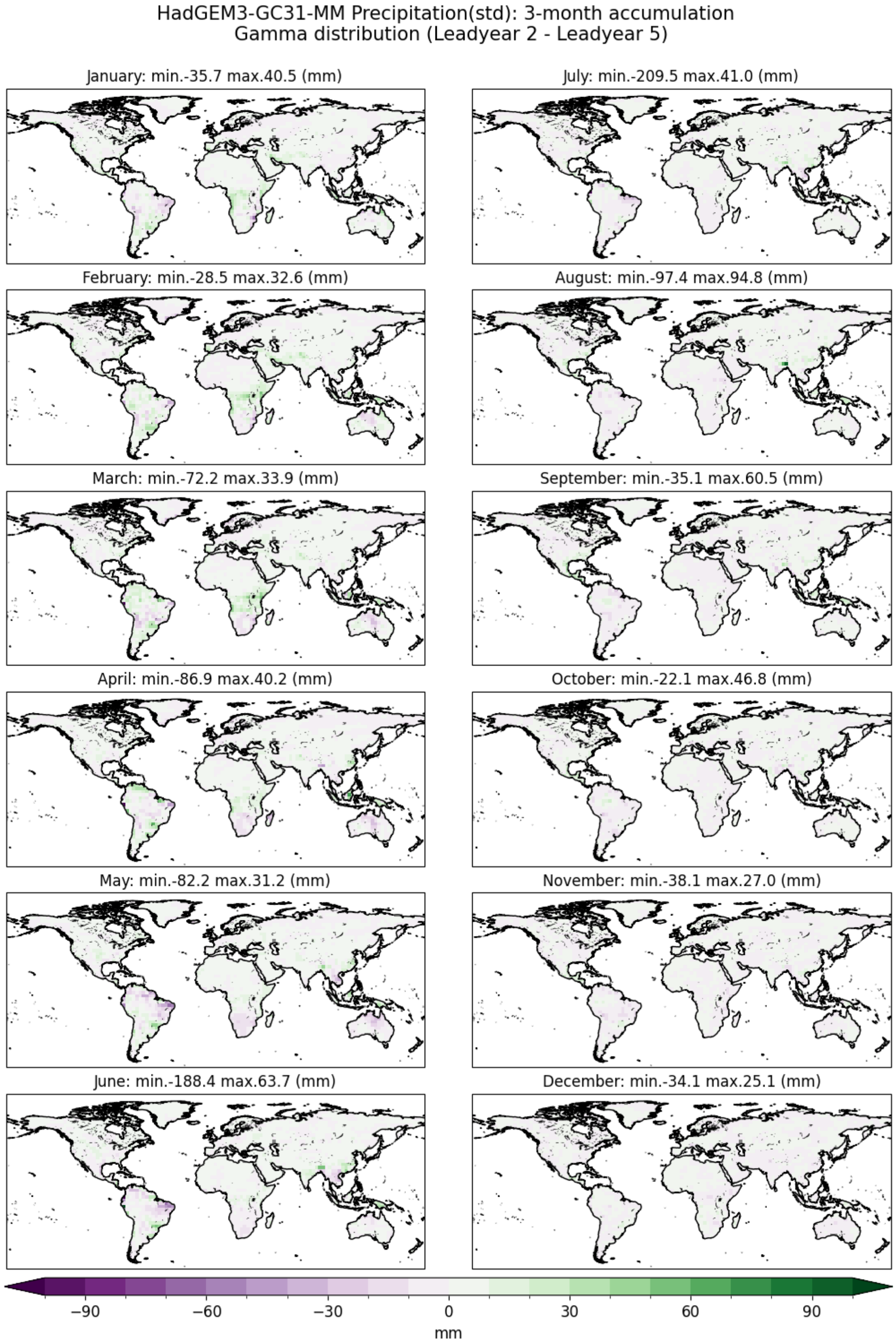

Figure 2. Difference between the precipitation (SPI) standard deviations of the lead month distributions for lead year 2 and lead year 5 calculated from accumulated precipitation (SPI). Values in the titles indicate the minimum and the maximum values of these differences.

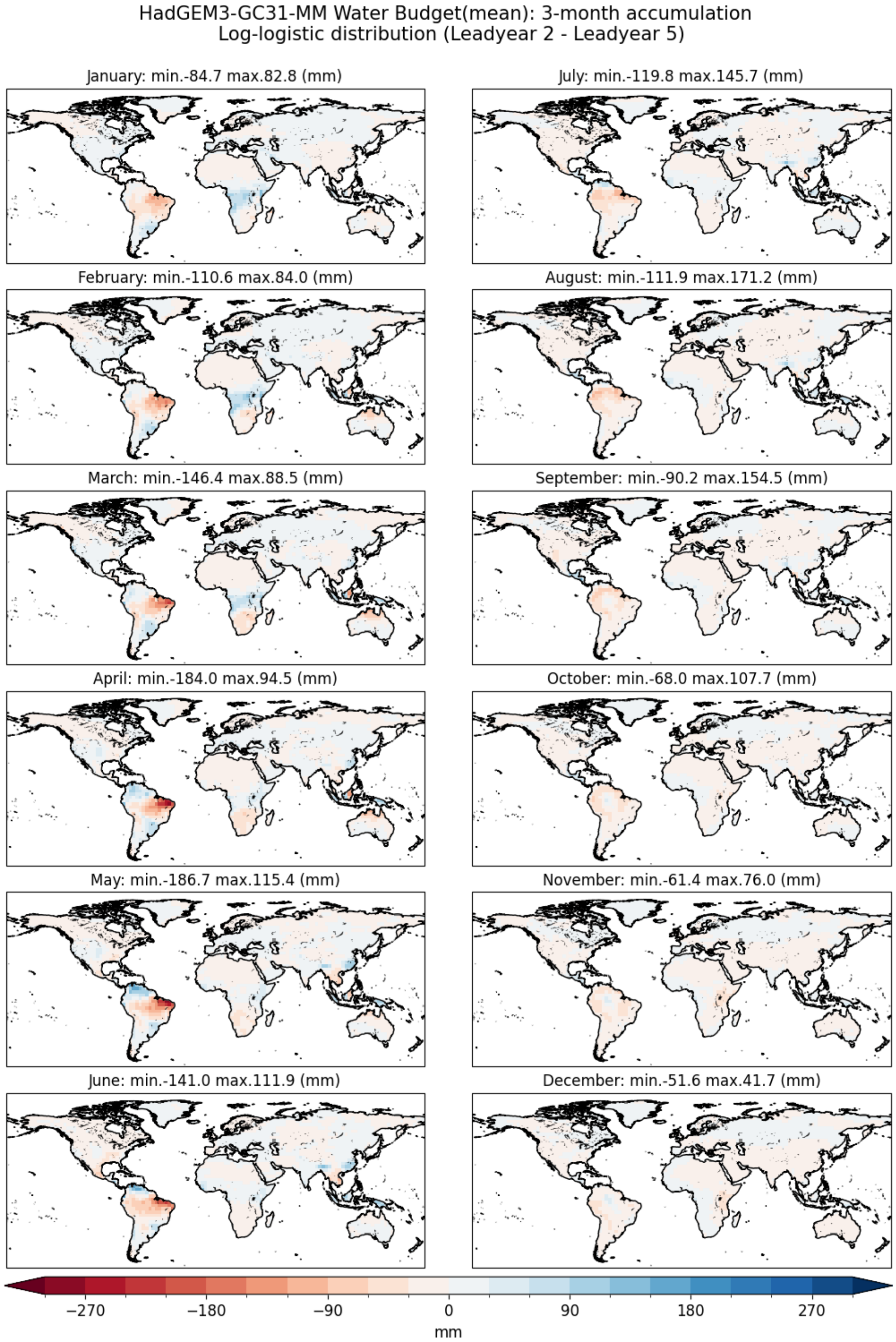

Figure 3. Same as Figure 1, but for the water budget (SPEI).

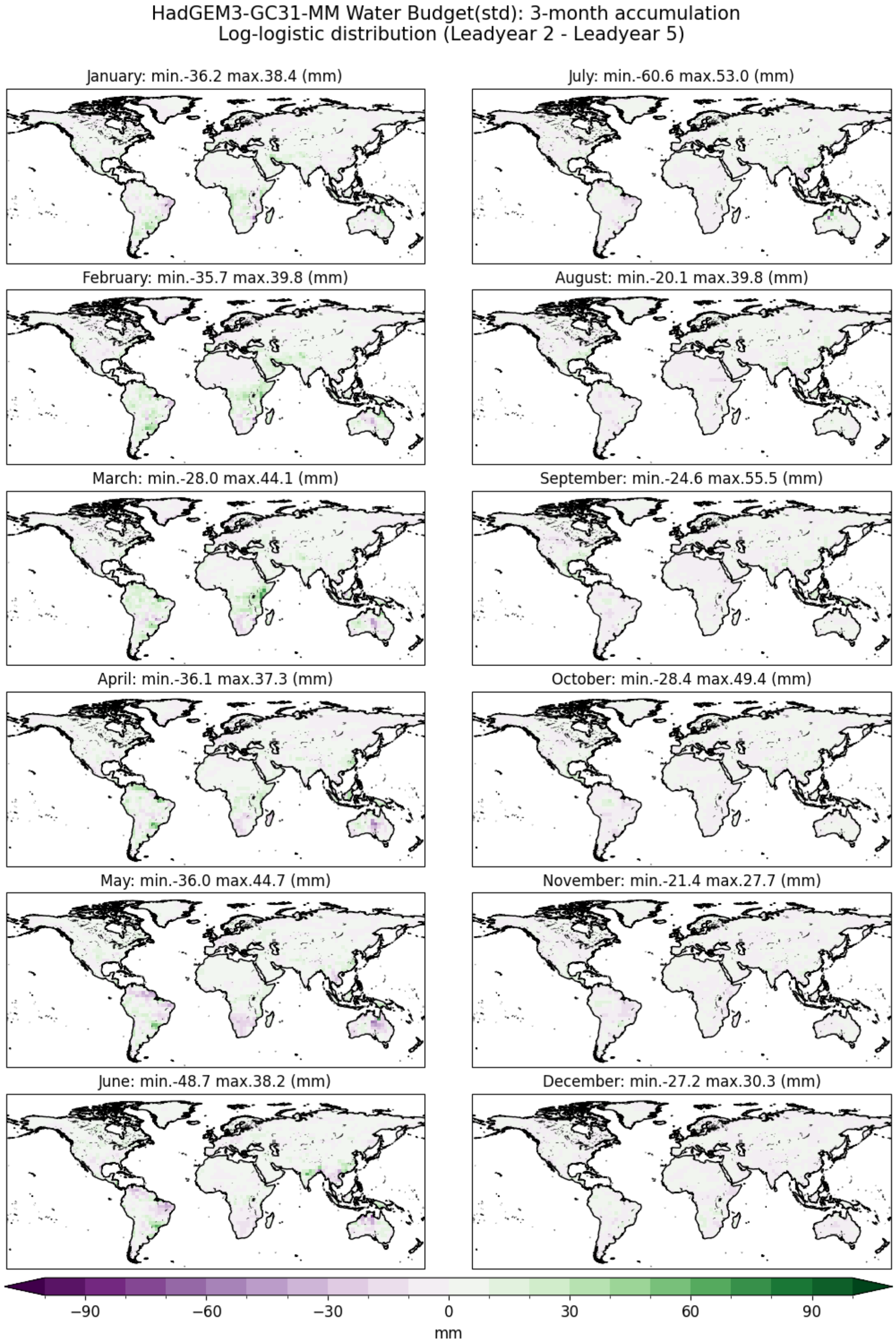

Figure 3. Same as Figure 2, but for the water budget (SPEI).

One reviewer asks about the focus on years 2 - 5. What are examples of the preventive

measures that decision makers can take on these lead times? Who will be using information with a 2 to 5 year lead-time? If you have some insight, it would be greatly appreciated.
We thank the reviewer for the comment. Climate information on the frequency and intensity of hot-dry compound extremes at multi-annual time scales can support decision-making processes in several sectors, such as agriculture, water management, energy and infrastructure. We have added the following sentences at the end of the second paragraph in Section 1 (line 33):
"For instance, multi-annual predictions of compound hot-dry extremes can inform strategic and preventive decisions across sectors. For instance, they can support agricultural planning such as irrigation investments, the multiplication of drought-resilient crop varieties, post-harvest management, and enhance the preparedness against pests and diseases (Delgado-Torres et al., 2025). Multi-annual climate information can also guide infrastructure and energy planning (Dunstone et al., 2022), including the enhancement of energy storage capacity or reinforcement of urban green areas to mitigate heat and dry stress, and anticipate impacts on society and environment."

Request 135 is not really answered. The reviewer is seeking guidance from you. No changes in the manuscript are needed, but a more informative response to the reviewer is needed if you can provide it.
Thank you for the comment. We have modified the answer to the reviewer,which can be found above in this document. We have tried to highlight how, especially before the production step, it is important to test the hindcast against several reference datasets, since both of them show different strengths and limitations. For this reason, we do not feel in the position of recommending one over the other.

Model overview table: if feasible, please summarize the most pertinent information on the model initialization procedures there.
Thank you for the comment. We have added a column to the Model overview table titled "Initialization scheme". There, when we were able to find the information, we summarized the initialization process of every experiment and, if available, the main datasets used in the assimilation.

Reply to the comment on ensemble averaging: please add your reply and a reference to the manuscript.
Thank you for the comment. We have added in section 2.2 (line 128) a summary of the reply provided to the reviewer. Specifically, we stated that, while the computation of the indicators was done at an ensemble-member level, we used the ensemble mean in the forecast quality assessment as it is interpreted as the representation of the predictable signal. In addition, adding a couple of references on the topic (Eade et al., 2014; Smith et al., 2019).

Remark on the use of high-impact events: I recommend a reformulation along the lines that some of these events are high-impact events, but not all of them.
Thank you for the comment. Indeed, we agree that it is an important difference. In addition, the focus of this manuscript is not impacts, but rather the predictability of the hazards. For this reason, in section 1 (line 15), we have substituted "high impact weather events" with "extreme weather events".

Please increase the figure panel size as requested by the reviewer. The figure details are unreadable for not-so-young eyes.
The figures have been enlarged as suggested in the comment. More specifically, Figures 1,2 and 3 in the main text, as well as A1, A2 and A3 in the Supplementary Materials have been modified to have all the panels the same size. In addition, for figures A4 and A5 in the Supplementary Material, the figures have been reorganised vertically to enlarge the size of the panels.