Response to Anonymous Referee #1

Manuscript Title: Climatology and trends of extreme precipitation in France: evaluation of an explicit-convection regional climate model

Authors: Nicolas Decoopman, Juliette Blanchet, Antoine Blanc, and Cecile Caillaud

Journal: Hydrology and Earth System Sciences (HESS)

We thank referee #1 for their constructive and detailed review of our manuscript. Below are our point-by-point responses to the comments, with line and figure numbers referring to the manuscript.

Main Comments

Comment 1: Annual Maxima Trends

Considering that you are not looking at very rare extremes (10-yr return level), I suggest to look also at trend in Annual Maxima, that is not affected by uncertainty in fitting a 3-parameters distribution like GEV, in order to compare/confirm your patterns.

- **Response:** We thank the reviewer for this constructive suggestion. We agree that comparing the trend of annual maxima with GEV-based return level trends is a useful check. However, a direct comparison between the trend of the annual maxima and that of the 10-year return level (RL_{10y}) is not fully consistent because they represent different quantiles and do not necessarily share the same trend. Instead, to perform a directly comparable analysis, we compare the trend of the annual maxima (obtained via a simple linear regression) with the trend of the GEV-estimated 2-year return level (RL_{2y}), since the average annual maximum corresponds approximately to a 2-year return period.

We have performed this consistency check, and the trends in annual maxima align very closely with those obtained from the GEV-estimated RL_{2y} in both sign and regional patterns. While we show the details of this comparison in this response (and will add a brief discussion of this consistency check in the revised text), we have chosen not to add a dedicated figure in the revised manuscript for the time being to keep the paper concise.

Comment 2: Redundancy of M^* Models for Hourly Data

Hourly data mostly starts in 1990. Thus, the M models are applied just to the daily case, right? Do you really need 6 non-stationary models? In how many cases the M^* models resulted better than the M ones?

- **Response:** We appreciate this comment. Indeed, for the hourly case (1990-2022), because the records begin after 1985, the breakpoint models (M^*) are mathematically equivalent to the standard linear models (M) since all observations lie after the breakpoint ($t \geq 1985$). In the revised methodology section, we will explicitly clarify that the model selection procedure defaults to the standard linear models for the hourly series. For the daily case (1959-2022), the M^* models represent a physical hypothesis of trend acceleration starting in the mid-1980s. To justify their inclusion, we will add statistics in the revised manuscript showing the percentage of stations andgrid points where an M^* model was selected over the standard linear (M) or stationary (M_0) models.

Comment 3: Rationale for the Two-Step Model Selection

Not clear to me why you use the 2-step procedure, why not simply choosing the model with the smallest p-value?

- **Response:** We agree that the rationale behind this selection procedure should be explained more clearly. The likelihood ratio test compares each non-stationary model M_j ($j \geq 1$) to the stationary model M_0. However, models M_3 and M_3^* (where both μ and σ vary over time) have 2 degrees of freedom, whereas models M_1/M_1^* and M_2/M_2^* have only 1 degree of freedom. Choosing the model with the absolute smallest p-value across different degrees of freedom can bias the selection toward simpler models or lead to overfitting. Our 2-step procedure is designed to first test if a trend in both location and scale (M_3 or M_3^*) is statistically justified. If it is not, we look at whether a simpler trend in either location only (M_1/M_1^*) or scale only (M_2/M_2^*) is justified. This prevents overfitting and prioritizes a physically consistent representation of changing extremes (i.e., both parameters changing) only when statistically supported. We will add a clarifying paragraph to Section 3.3.2 to explain this rationale.

Comment 4: Clarification of Evaluation Metrics

I suggest to illustrate in the method section which evaluation metrics you used (e.g. ME, r) and for which variables other than 10-yr RL (e.g. frequency of wet days)

- **Response:** We agree. The subsection defining the agreement metrics (Pearson correlation r, Mean Error ME, and relative bias in %) and introducing all the evaluated variables (mean number of wet days, mean annual total precipitation, daily and hourly

10-year return levels) was indeed commented out in the draft. In the revised manuscript, we will restore and expand this subsection (Section 3.4) to explicitly state which metrics and variables are used before they appear in the results section.

Comment 5: Introduction of Relative Biases

I strongly suggest to add information on relative biases (also, in figures 4 and 5), because N mm/year could be small or big difference depending on the baseline!

- **Response:** We agree with the reviewer. Relative biases (expressed in %) are more informative for comparing regions with highly contrasted precipitation baselines, such as the dry Mediterranean coastal areas and wet mountainous regions. In the revised manuscript, we will calculate relative biases for annual cumulative precipitation and the daily/hourly 10-year return levels, discuss them in the text, and update Figures 4 and 5 (and the corresponding text in Section 4.1) to include relative bias information.

Comment 6: Presentation of Bias at the Hourly Scale

This seems mostly referring to spatial correlation; but also bias should be mentioned (e.g., underestimation at 1h)

- **Response:** We will revise this paragraph to explicitly mention the seasonal variations in bias, in addition to the spatial correlation. Specifically, we will highlight the systematic underestimation of hourly extremes (1h 10-year return level) by AROME, which is particularly pronounced during the convective seasons (spring and summer).

Comment 7: Climatology Maps Customization (Figure 4)

I suggest to substitute 3rd row or add a 4th row with relative biases. For all figures with maps: Use same size for dots (now, the highest values have both darker color and bigger size, hiding other; I think color is enough); maybe find a darker color for the 0 bios (not much visible as it is now).

- **Response:** We thank the reviewer for these suggestions, which will significantly improve the readability of the maps.

1. We will replace the absolute difference in the third row of the climatology maps (Figure 4) with the relative bias (%) to make the regional comparison clearer.
2. We will use a uniform dot size for all stations across all maps so that larger dots do not obscure adjacent stations.

3. We will modify the color scale so that a zero bias is represented by a clearly visible, distinct neutral color (like a dark grey or a white with a dark border) rather than a light color that blends into the background.

Comment 8: Seasonal Bias Synthesis (Figure 5)

I suggest a second panel summarizing %biases for the 4 variables.

- **Response:** We agree. We will add a second panel to Figure 5 showing the seasonal evolution of the relative bias (%) for the four precipitation indices. This will provide a concise and complete seasonal comparison of both spatial correlation and bias.

Comment 9: Restructuring the Monthly Trend Mapping (Figure 8 & 9)

I found this text a too qualitative, and all those maps in figure 8 not much relevant and their metrics not easily readable. My suggestion is to move this figure in supplementary, and to update figure 9 to give more complete summary information about monthly performance (not just r): trend from stations and model, bias, correlation for both daily and hourly extremes...

- **Response:** We agree with this suggestion. The monthly maps of hourly trends (Figure 8) are very dense and difficult to read.

1. We will move the monthly trend maps to the Supplementary Material.
2. We will redesign the summary figure (Figure 9) to show the monthly evolution of: the mean trend (stations vs. AROME), the relative bias, and the spatial correlation, for both daily and hourly extremes. This structured summary will make the monthly analysis much more quantitative and readable.

Comment 10: Rain Gauge Undercatch and Inhomogeneities

Link this to your results. Should we expect even higher underestimation at hourly scale by the model, if undercatch errors for stations will be corrected? What do you mean with "heterogeneity"?

- **Response:** We will clarify these points in the discussion (Section 5.1):

1. Since rain gauge measurements are affected by undercatch, correcting for these errors would increase the observed extreme values. Consequently, the model's underestimation of hourly extremes is likely even larger than what we report. We will explicitly state this in the text.

2. By “heterogeneity”, we actually meant “inhomogeneities” (non-climatic shifts in the time series caused by changes in station location, environment, or sensor types over decades). We will correct this term to “inhomogeneities” and mention its potential impact on local trend detection.

Comment 11: Introduction of the SMEV Approach

I think SMEV approach should also be mentioned (Marra et al. 2020 10.1029/2020gl090209) which has been also applied in both stationary mode for evaluation of short-run CPMs (Correa-Sanchez et al., 2025 https://doi.org/10.1016/j.jhydrol.2025.133324) and non-stationary mode on 90-yr CPM (Lompi et al., 2025 https://doi.org/10.1016/j.adwatres.2025.105071)

- **Response:** We thank the reviewer for pointing out these highly relevant references. The Simplified Metastasis Extreme Value (SMEV) distribution is indeed a very relevant alternative to the GEV for shorter records and sub-daily extremes. In the revised discussion, we will add a discussion of the SMEV approach, citing the recommended papers (Marra et al., 2020; Correa-Sanchez et al., 2025; Lompi et al., 2025), and highlight it as a promising framework for future studies evaluating sub-daily extremes from high-resolution climate runs.

Minor Comments

- **Line 11 (and 395). “added value” ... with respect to what?** *Response:* We will clarify that the “added value” is relative to traditional, coarser-resolution regional climate models (RCMs) with parameterized convection (typically 12 km resolution, such as those in EURO-CORDEX).
- **Line 19. What does it mean “under calm conditions”?** *Response:* We will rephrase “under calm conditions” in line 89 to “for non-extreme or average precipitation events” to be physically precise.
- **Line 54. “sensitivities”: maybe more clearly “temperature-scaling rate”** *Response:* We will change “sensitivities” to “temperature-scaling rates”.
- **Line 57. “hourly extremes”: what kind? Annual Maxima, percentiles, return levels?** *Response:* We will specify the exact metrics (e.g., annual maxima, high percentiles, or return levels)used in the cited papers.
- **Line 70. “over long period”: this is not generally true for CPM. This is way your study is valuable.** *Response*: We agree. We will rephrase this to emphasize that simulating long periods is a major computational challenge for CPMs, which makes our 63-year hindcast simulation particularly valuable.
- **Line 121. 50km? you mentioned 25-31 km at line 79** *Response*: We will correct this discrepancy. The native resolution of ERA5 is indeed ~31 km. The boundary conditions are interpolated, and we will ensure the text consistently states the resolution of the forcing data (~31 km).

- **Lines 126-128. I wonder if this should be placed in the discussion (and a smaller figure, considering the simplicity of its information)** *Response*: We agree. We will move the temperature trend comparison section and the accompanying Figure 2 to the Discussion section (under model limitations) as it is a secondary result explaining the underestimation of precipitation trends.
- **Line 133: Evaluation is made per year and season ... and month.** *Response*: We will add "and month" to this sentence.
- **Figure 3. Mention that is an example of M3* model with positive trend. Consider if moving to supplementary (I don't find it so relevant).** *Response:* We prefer to keep this illustration figure in the main manuscript because we believe it is very useful for readers who are less familiar with non-stationary GEV modeling. We will, however, clarify in the caption that it shows an example of the M_3^* model with a positive trend.
- **Line 253. "moderately" capture... in some seasons (highest r=0.4).** *Response*: We will tone down the text to use more cautious language (e.g., "moderately capture" or "partially capture") when discussing seasons with low to moderate correlations.
- **Lines 261-264. You mention twice "no trend" for seasons where about 1/3 of stations show significant trendI guess you intended something like "no clear spatial pattern" or similar. I suggest to clarify/correct** *Response*: Indeed, "no trend" is inaccurate since a significant number of stations do show trends, but they lack a coherent regional spatial pattern. We will rephrase this to "no clear spatial pattern" or "spatially incoherent trends".
- **Line 273. -84.7% in figure 7** *Response*: We will check the sign in the figure and correct it if necessary.
- **Line 291**. Also July and October null/very small Response: We will add that July and October also exhibit very small or null correlations.
- **Line 345. Where? Same regions where you find underestimation?** *Response*: We will specify that this timing/localization mismatch occurs primarily in regions dominated by summer convective activity (such as the Massif Central and the Alps) where the model underestimates peak hourly intensities.
- **Line 366. Is this really an issue at daily scale? How? I think it could lower the absolute values, not clearly the temporal trend. (same consideration for line 382-383)** *Response*: This is a very valid point. Spatial smoothing (comparing a 2.5 km grid box to point measurements) decreases the absolute magnitude of the extremes, but should not strongly affect the relative temporal trends (%) unless the spatial structure of storms changes over time. We will correct the text to distinguish between absolute values (affected by smoothing) and relative trends (primarily affected by the model's underestimated temperature trends). We will remove grid-box smoothing as an explanation for relative trend underestimation.