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 his/her 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 mean, since the mean of a GEV is an analytical function of its parameters (μ, σ, ξ) and can be compared directly to the linear regression.

We have performed this consistency check across all 1,558 daily stations (hydrological year, 1960–2022). For the GEV mean, we computed the mean at the start and end of the record using the selected best model parameters. For the annual maxima, we fitted a simple linear regression.

Both the spatial patterns (regional trends) and magnitude/sign of the trends show an extremely strong agreement across the stations, with some minor differences to note:

- **Spatial patterns**: The maps of observed annual maxima trend (Response Figure 1 a) and GEV-estimated mean trend (Response Figure 1 b) show identical regional patterns, with the strongest positive trends concentrated in the Southern Alps and the Rhone valley.
- **Quantitative agreement**: The Pearson correlation coefficient is $r = 0.86$ for absolute trends (mm/decade, Response Figure 2 a) and $r = 0.88$ for relative trends (%, Response Figure 2 b). The regression fit line (red) is very close to the 1:1 diagonal (grey dashed), confirming that the GEV-estimated trends are highly robust and not significantly affected by parameter fitting uncertainty.
- **Dampening / Smoothing**: The regression slope for the relative trends is 0.80 (Response Figure 2 b), indicating that GEV-estimated trends are on average slightly smaller (dampened by about 20%) than those obtained by simple linear regression. This difference is expected because non-stationary GEV parameters are estimated via maximum likelihood over the entire distribution of annual maxima, allowing extreme values to be modeled by the scale and shape parameters. This makes the GEV mean trend less sensitive to individual extreme outliers (especially at the boundaries of the series) than Ordinary Least Squares (OLS) linear regression, which minimizes squared residuals.

We have also extended this consistency check to the hourly scale across all 562 hourly stations (hydrological year, 1990–2022). Despite the much shorter record length (33 years vs 63 years) and the higher spatial and temporal noise typical of sub-daily convective extremes, the hourly GEV mean trends show a strong agreement with the raw annual maxima trends:

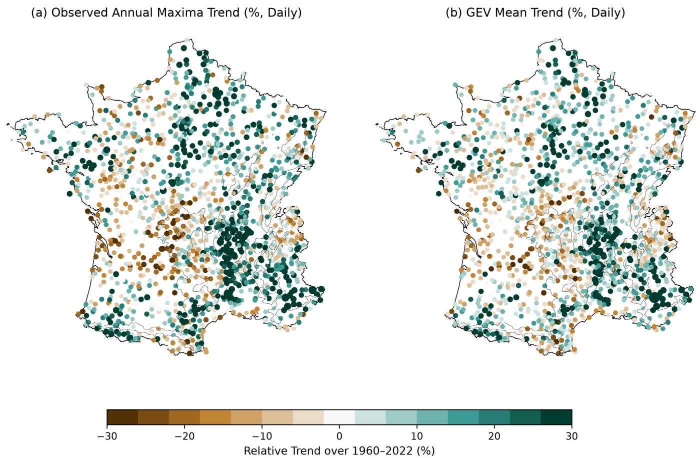
- **Spatial patterns**: The maps of observed hourly annual maxima trend (Response Figure 3 a) and GEV-estimated hourly mean trend (Response Figure 3 b) show highly consistent spatial patterns.
- **Quantitative agreement**: The Pearson correlation coefficient is $r = 0.72$ for absolute trends (mm/decade, Response Figure 4 a) and $r = 0.70$ for relative trends (%, Response Figure 4 b).
- **Dampening / Smoothing**: The regression slope is 0.71 (Response Figure 4 b), showing a moderate dampening (about 30%) for GEV-estimated trends. This is again expected due to the robustness of GEV maximum likelihood estimation against individual outlier years in short records.

Manuscript modifications: In the revised manuscript, we have added the following introductory paragraphs to `Section 4.2.1 (Daily)` and `Section 4.2.2 (Hourly)`:

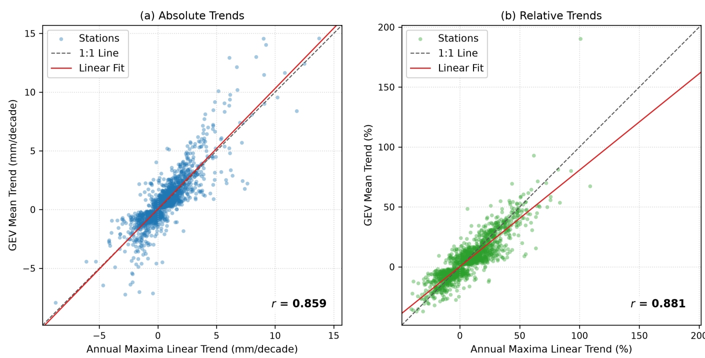
- In `Section 4.2.1 (Daily)` (at the very beginning): > *"Before analyzing trends in daily return levels, we verified the consistency of GEV-estimated trends in the mean against those obtained via a simple linear regression on observed annual maxima. This preliminary check shows a very strong spatial and quantitative agreement (Pearson correlation*

$r = 0.86$ *for absolute trends,* $r = 0.88$ *for relative trends), confirming the robustness of the statistical fits."*

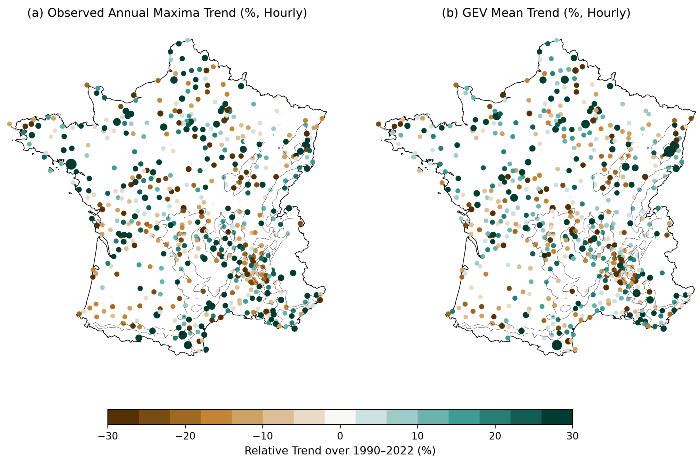
- In `Section 4.2.2 (Hourly)` (at the very beginning): > *"Similarly to the daily scale, we verified the consistency of GEV hourly trends by comparing GEV-estimated trends in the mean to linear regression trends of observed annual maxima. Despite the shorter record length and the higher noise level of sub-daily extremes, we find a robust agreement (Pearson correlation* $r = 0.72$ *for absolute trends,* $r = 0.70$ *for relative trends)."*



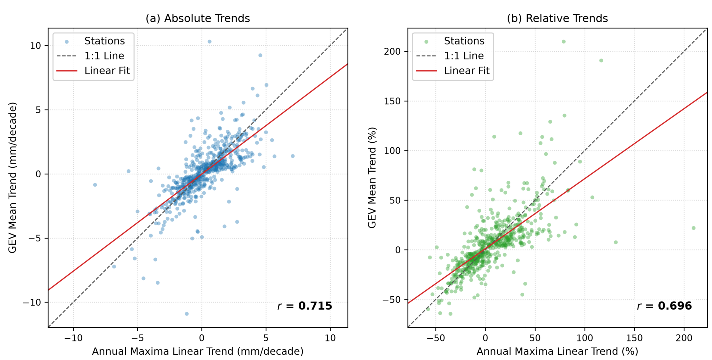
Response Figure 1: Comparison of the spatial distribution of relative daily trends (%) over 1960–2022 across all 1,558 daily stations: (a) observed annual maxima trend obtained by simple linear regression; (b) GEV-estimated mean trend.



Response Figure 2: Scatter plots of GEV-estimated daily mean trend vs. observed daily annual maxima trend across all 1,558 daily stations: (a) absolute trends in mm/decade; (b) relative trends in %. The 1:1 line is shown in dashed grey, and the linear regression fit is in red, with Pearson correlation coefficients (r) indicated.



Response Figure 3: Comparison of the spatial distribution of relative hourly trends (%) over 1990–2022 across all 562 hourly stations: (a) observed annual maxima trend obtained by simple linear regression; (b) GEV-estimated mean trend.



Response Figure 4: Scatter plots of GEV-estimated hourly mean trend vs. observed hourly annual maxima trend across all 562 hourly stations: (a) absolute trends in mm/decade; (b) relative trends in %. The 1:1 line is shown in dashed grey, and the linear regression fit is in red, with Pearson correlation coefficients (r) indicated.

Response Table 1: Percentage of locations where each model category is selected as the best fit for daily precipitation extremes (M_0 representing no trend, M a linear trend, and M^* a breakpoint trend starting in 1985). All percentages sum to 100% for each subset.

Period	**Stations (Obs.)**			**Grid Points (AROME)**		
	M_0	M	M^*	M_0	M	M^*
Annual (YEAR)	66.1%	17.8%	16.0%	75.5%	11.8%	12.7%
Autumn (OND)	62.8%	16.2%	21.0%	73.5%	7.9%	18.6%
Winter (JFM)	64.5%	22.6%	12.9%	65.6%	22.5%	11.9%
Spring (AMJ)	62.1%	16.6%	21.3%	75.0%	11.7%	13.3%
Summer (JAS)	68.8%	14.2%	17.1%	65.6%	10.5%	24.0%

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 ($t \geq 1985$). For the daily case (1959-2022), the M^* models represent a physical hypothesis of a trend emerging in the mid-1980s.

To address this comment, we present the table below (Response Table 1) showing the full breakdown of selected models: the stationary model (M_0, representing no significant trend at the 10% significance level), standard linear trend models (M), and breakpoint models (M^*). In this way, the percentages for each subset sum to 100%.

For the grid point percentages (AROME), the statistics are calculated over the entire domain of mainland France. To ensure this does not introduce spatial representation biases, we also computed these statistics restricted only to the grid points closest to the daily stations ($N = 7,486$). The results are highly consistent: for the Annual period, selection rates at matched grid points are 75.6% for M_0, 11.3% for M, and 13.0% for M^* (compared to 75.5%, 11.8%, and 12.7% over all mainland grid points). We therefore present the full-grid statistics in the table to reflect the complete spatial support of the model.

Manuscript modifications:

- In `Section 3.2.2 (Model selection)`, we have explicitly clarified that the model selection procedure defaults to the standard linear models (M) for the hourly series.
- In `Section 4.2.1 (Daily)`, we have added a sentence stating the percentage of stations and grid points where M_0, M, and M^* models were selected.

- In `Section 4.2.2 (Hourly)`, we have added a sentence stating the percentage of stations and grid points where M_0 and M models were selected.

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 more than M_0, whereas models M_1/M_1^* and M_2/M_2^* have only 1 degree of freedom more. Choosing the model with the absolute smallest p-value across different degrees of freedom can bias the selection toward simpler and thus more constrained models. 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 favors more flexible models (i.e. both parameters changing) when statistically supported.

Manuscript modifications:

- In `Section 3.2.2 (Model selection)`, we have added the following clarifications: "*We apply for each non-stationary model M_j, $j \geq 1$, a likelihood ratio test to compare the goodness of fit of M_j to that of the stationary model M_0, accounting for their respective number of parameters to avoid overfitting [...] This two-step process prioritizes models with simultaneous temporal effects on μ and σ when statistically justified, ensuring complexity is introduced only when it provides plus-value.*"

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 lacking in the first version of this article.

Manuscript modifications:

- In the revised manuscript, we have added an new `Section 3.3 (Assessing the agreement between AROME and the stations)`:

Assessing the agreement between AROME and the stations

To evaluate the CNRM-AROME model against observations, each Météo-France station is matched to the closest model grid point (2.5 km × 2.5 km) based on its geographic location. This pairing allows assessing the spatial and temporal coherence of simulated fields against raingauge measurements.

We assess the model's agreement using three main metrics:

1. **Pearson correlation coefficient (r):** Used to quantify the spatial consistency between observed and simulated fields across all station locations.
2. **Mean Error (ME, or absolute bias):** Measures the average difference between the simulated and observed statistics, defined as:

$$ME = \frac{1}{n}\sum_{i=1}^{n}(Y_{i,\mathrm{AROME}} - Y_{i,\mathrm{OBS}}) \tag{1}$$

   where Y_i represents the climatological or trend statistic computed at station i (or its closest model grid point) and n is the total number of compared stations.

3. **Relative Bias (RB, %):** Expresses the bias as a percentage of the observed baseline, which is particularly useful for comparing regions with highly contrasted precipitation baselines:

$$RB = 100 \times \frac{\sum_{i=1}^{n}(Y_{i,\mathrm{AROME}} - Y_{i,\mathrm{OBS}})}{\sum_{i=1}^{n} Y_{i,\mathrm{OBS}}} = 100 \times \frac{ME}{\bar{Y}_{\mathrm{OBS}}} \tag{2}$$

   where $\bar{Y}_{\mathrm{OBS}}$ is the mean of the observed values across all compared stations.

These metrics are calculated for five core precipitation indices: the annual frequency of wet days (daily precipitation exceeding 1,mm), the mean annual total precipitation, the daily and hourly 10-year return levels, and the relative trends in daily and hourly 10-year return levels. Climatological indices are evaluated in a stationary framework, whereas relative trends are obtained from the estimated non-stationary GEV distributions (cf. Section 3.2.1). In the latter case, we do not compute the relative bias as relative trends are already %.

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.

Manuscript modifications:

- In `Section 3.3 (Assessing the agreement between AROME and the stations)`, we have defined the Relative Bias (RB, %) metric to quantify systematic errors relative to the observed baselines.
- In `Section 4.1 (Evaluation of the precipitation climatology of AROME)`, the climatology maps (Figure 3) have been updated to add the relative bias (in %) in the fourth row, and the caption has been updated accordingly.
- In the same section, we have added a second panel to Figure 4 showing the seasonal evolution of the relative bias (in %) for the four climatological indices.

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

Manuscript modifications:

- At the end of `Section 4.1 (Evaluation of the precipitation climatology of AROME)`, the summary paragraph has been revised to explicitly mention the systematic underestimation of hourly extremes alongside the lower spatial agreement in spring:

  > *"In summary, AROME shows a good representation of the climatology of precipitation extremes that encourages us to continue evaluation on trends, keeping however in mind that hourly extremes are systematically underestimated by the model throughout the year (especially during spring and summer) and exhibit a lower spatial agreement in spring."*

Comment 7: Climatology Maps Customization (Figure 3)

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 significantly improve the readability of the maps.

Manuscript modifications:

1. In `Section 4.1 (Evaluation of the precipitation climatology of AROME)`, we have added a fourth row to the climatology maps (Figure 3) to present the relative bias (%) alongside the absolute difference in the third row, making the regional comparisons clearer while preserving the absolute magnitude information.
2. In all map figures across the manuscript (Figure 3 in `Section 4.1`, Figures 5 and 6 in `Sections 4.2/4.3`, and the monthly trend maps in `Appendix C`), we have used a uniform dot size for all displayed stations so that larger dots do not obscure adjacent stations.
3. In `Section 4.1 (Evaluation of the precipitation climatology of AROME)`, the color scale of the climatology maps (Figure 3) has been modified so that a zero bias is represented by a clearly visible grey fill.
4. In the trend maps of `Section 4.2 (Spatial distribution of precipitation trends)`, `Section 4.3 (Seasonal variations of hourly trends)`, and the monthly maps in `Appendix C`, we adopted the following display conventions for station markers and grid cell masks:

Type	Display
Non-significant (stations)	Open circle: white fill, light grey outline
Significant 0 (stations)	Filled grey disk, dark outline
Significant coloured (stations)	Trend colour on the map scale; **uniform marker size**
AROME grid	Non-significant grid points **not displayed**

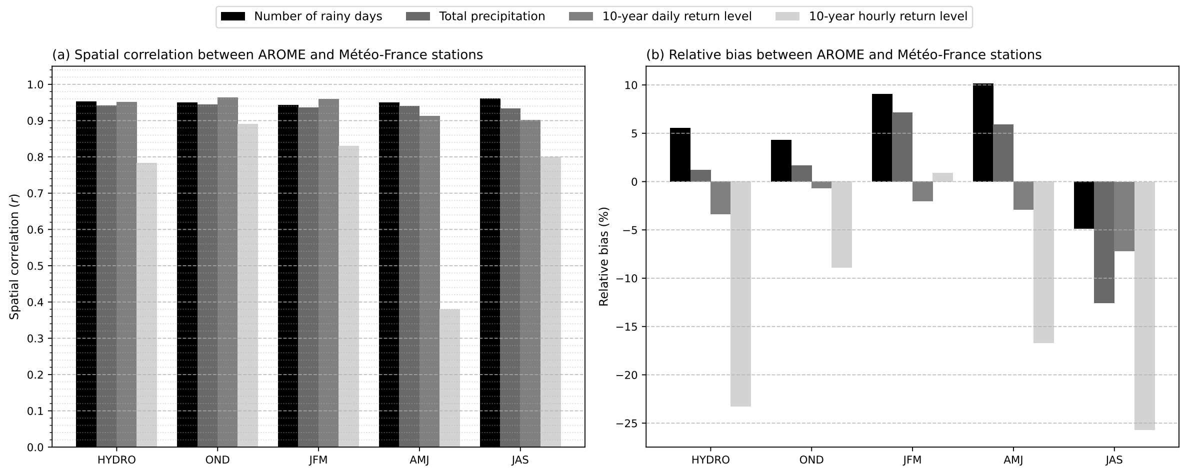
Comment 8: Seasonal Bias Synthesis (Figure 4)

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

- **Response:** We agree. We have added a second panel to the seasonal synthesis figure showing the seasonal evolution of the relative bias (%) for the four precipitation indices. This provides a concise and complete seasonal comparison of both spatial correlation and bias.

Manuscript modifications:

- In `Section 4.1 (Evaluation of the precipitation climatology of AROME)`, Figure 4 (seasonal synthesis) has been updated to include a second panel (b) presenting the seasonal relative bias (in %) for the four precipitation indices, and the caption has been updated to describe both panels.



Article Figure 4. *Seasonal synthesis of the climatological agreement between the AROME model and Météo-France stations for the four precipitation indices: (a) spatial Pearson correlation (r) and (b) relative bias (RB, in %) depending on the season.*

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 (example or organization below, or any other combination allowing to show all those metrics). It could support what you then mention at line 377 in the discussion.

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

1. We have moved the monthly trend maps to the Appendix.
2. We have redesigned the summary figure (Figure 7) as a three-panel bar-chart synthesis for both daily (1959–2022) and hourly (1990–2022) extremes: **(a)** monthly median GEV relative trends, with side-by-side subpanels for Météo-France stations and AROME; **(b)** monthly mean error (ME, %) in relative trends between AROME and stations; and **(c)** monthly spatial correlation (r) of GEV relative trends. This structured summary makes the monthly analysis much more quantitative and readable.

Manuscript modifications:

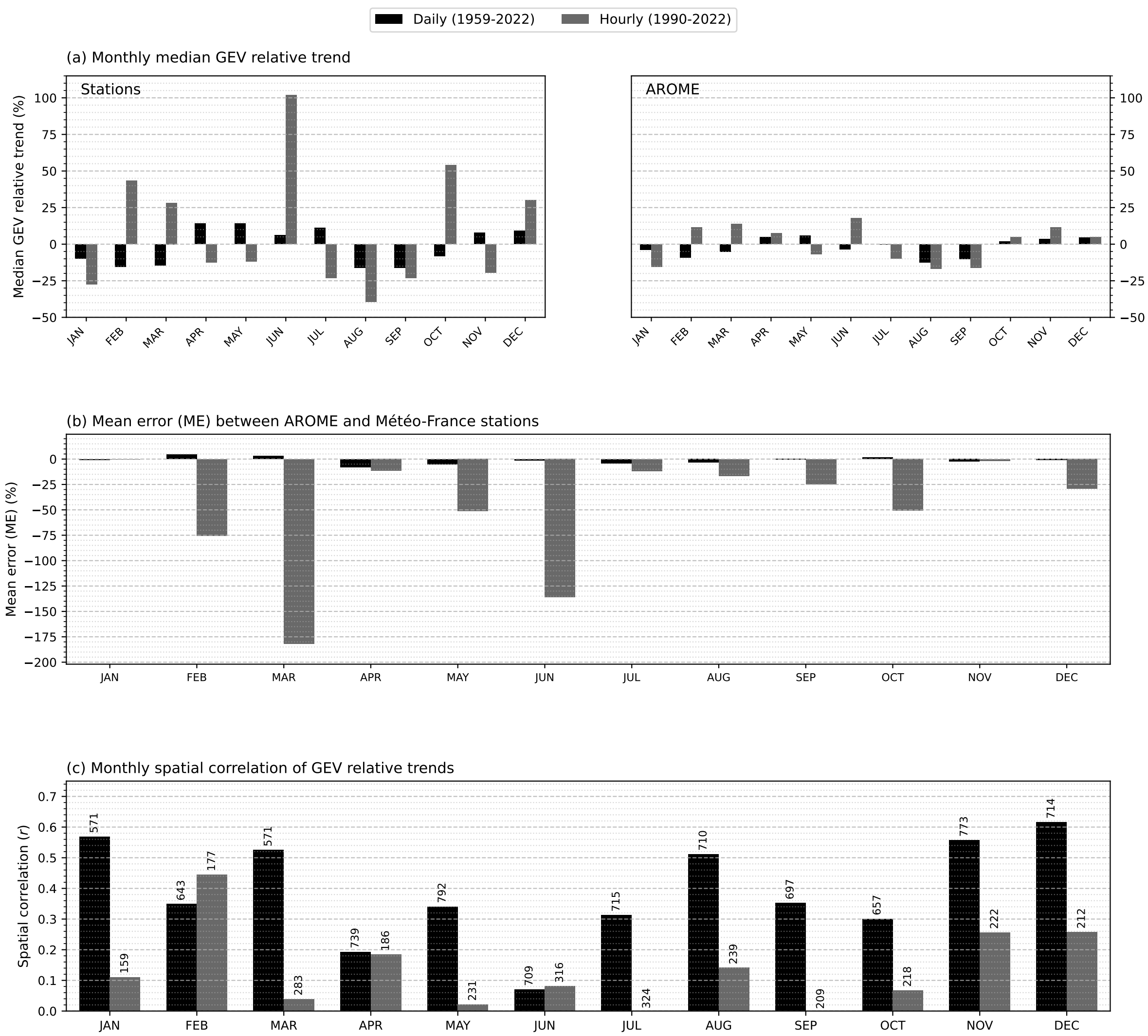
- In `Section 4.2.2 (Hourly)`, the maps of monthly trends (formerly Figure 8) have been moved to Appendix C as Figure C1.
- In `Section 4.2.2 (Hourly)`, the section describing monthly trends has been rewritten to be more quantitative, referencing the new Figure 8 to detail the median trends, relative biases, and spatial correlations:

> **Section 4.2.2 (Monthly refinement):** "The monthly analysis (Figure 7) further highlights the specificity of convective regimes and the strong temporal concentration of some signals. Figure 7a compares the monthly median GEV relative trends at Météo-France stations and the corresponding AROME grid points (side-by-side subpanels). At the daily scale, station median trends remain moderate throughout the year, ranging from −16.27% in September to +14.05% in April/May. In contrast, station-based hourly GEV trends show far larger values concentrated in specific convective windows, with monthly medians reaching +43.28% in February, +28.26% in March, +101.86% in June, +53.88% in October, and +30.09% in December. AROME reproduces the positive sign of these monthly trend peaks but systematically underestimates their magnitude. For instance, in June, AROME simulates a median hourly GEV trend of +17.70% (against +101.86% for stations), and in February, it simulates +11.46% (against +43.28% for stations).
>
> This systematic underestimation is reflected in the monthly mean errors (ME) between the corresponding AROME grid points and stations shown in Figure 7b, with large negative values at the hourly scale during convective months (e.g., $ME = -75.6\%$ in February, $ME = -182.0\%$ in March, and $ME = -136.1\%$ in June), whereas daily mean errors remain much smaller (e.g., $ME = -0.9\%$ in January and $ME = 3.1\%$ in March).
>
> Finally, Figure 7c shows that spatial correlation (r) is consistently higher at the daily scale (peaking at $r = 0.62$ in December) than at the hourly scale. At the hourly scale, spatial correlations are extremely low or close to zero during the convective months (e.g., $r = 0.08$ in June, $r = 0.07$ in October, $r = 0.04$ in March, and $r = 0.02$ in May), and virtually zero in July and September ($r \approx 0.00$), reflecting the high local spatial variability of hourly trends. February stands out as the only month showing moderate spatial agreement ($r = 0.45$)."

- The discussion section in `Section 5.3.2` (discussing the comparison of hourly signals) has been updated to refer to the new Figure 7 and include the quantitative statistics:

> **Section 5.3.2:** "AROME reproduces part of the seasonal structure of these hourly signals—concurring on the months associated with enhanced extreme

activity (Figure 7a)—but fails to capture their spatial organization and magnitude. Figure 7a and 7b show that the model systematically underestimates the magnitude of positive hourly GEV trends, leading to a much larger underestimation of relative trends at hourly than at daily scale (e.g., $ME = -136.1\%$ vs. -1.4% in June, see Fig. 7a,b). Furthermore, hourly spatial correlations shown in Figure 7c remain close to zero for most convective months (e.g., $r = 0.08$ in June, $r = 0.04$ in March), reflecting shifts in the simulated convective precipitation zones, especially in mountain regions (such as the Massif Central and the Alps) where summer convective activity dominates. Part of this trend magnitude underestimation is consistent with the underestimation of temperature trends in AROME (Figure 8)."



Article Figure 7. *Monthly GEV relative trend synthesis for both daily (1959–2022) and hourly (1990–2022) scales, calculated only over locations with statistically significant trends at the stations (at the 10% level):* ***(a)*** *median GEV relative trend (%) at Météo-France stations (left) and corresponding AROME grid points (right);* ***(b)*** *mean error (ME, in %) between the corresponding AROME grid points and Météo-France stations; and* ***(c)*** *spatial correlation (r) of GEV relative trends. Values above each bar in panel (c) indicate the number of significant stations used in the calculations. All statistics are evaluated at AROME grid points corresponding to these stations.*

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 agree. We have added a discussion linking the rain gauge undercatch to the model underestimation, showing that correcting for undercatch would likely mean that the model's underestimation of hourly extremes is even larger than reported. We have also replaced the term "heterogeneity" with "inhomogeneities" and clarified its meaning (non-climatic shifts) and potential impact. In addition, we note that part of the model underestimation may be related to the 2.5 km horizontal resolution of AROME: Brousseau et al. (2016, https://doi.org/10.1002/qj.2822) showed that increasing the resolution to 1.3 km leads to a better representation of maximum precipitation intensities, with less pronounced underestimation.

Manuscript modifications:

- In `Section 5.1.2 (Data)`, the paragraph discussing rain gauge limitations has been updated as follows:

> "First, station precipitation measurements are frequently affected by undercatch errors, which are particularly pronounced in cold and mountainous regions with strong winds. Because undercatch typically leads to an underestimation of actual precipitation (especially for high-intensity, short-duration convective events), 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 here. Part of this underestimation may also stem from the 2.5 km horizontal resolution of AROME: Brousseau et al. (2016) showed that increasing the resolution to 1.3 km results in a better representation of maximum precipitation intensities, with less pronounced underestimation. Furthermore, station time series can also exhibit some inhomogeneities (non-climatic shifts in the time series caused by changes in station location, environment, or sensor types over decades), which could introduce local artificial trend components, although the data used here have been quality-controlled by Météo-France."

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. We agree that the Simplified Metastasis Extreme Value (SMEV) distribution is a very valuable alternative to the GEV for short records and sub-daily extremes. We have added a mention of the SMEV distribution in the limitations section, citing the three recommended papers.

Manuscript modifications:

- In `Section 5.1.2 (Data)`, the paragraph discussing alternative extreme value approaches has been updated as follows:

  > "To make the estimation more robust, other solutions include taking into account threshold exceedances within a generalized Pareto distribution (GPD) framework, or using a distribution modeling all precipitation values, such as the recent extended generalized Pareto distribution (Haruna et al., 2025) or the Simplified Metastasis Extreme Value (SMEV) distribution (Marra et al., 2020; Correa-Sanchez et al., 2025; Lompi et al., 2025), or using a neighborhood or region-of-influence approach (Blanchet et al., 2021) to reduce sampling bias by using more data."

Minor Comments

- **Line 11 (and 395). "added value" ... with respect to what?**
  - **Response:** We agree. Since this paper does not perform a direct comparison with coarser regional climate models, we have rephrased "added value" to "value" (or "clear value") in both the Abstract and the Conclusion (Section 6) to focus on the model's usefulness for studying extremes rather than framing it as a relative added value.
  - **Manuscript modifications:** In the Abstract and Conclusion (Section 6), the text has been corrected to refer to the "value" of the explicit-convection model.
- **Line 19. What does it mean "under calm conditions"?**
  - **Response:** We agree. We have rephrased this term to be physically precise.
  - **Manuscript modifications:** In `Section 2.2`, we have changed "under calm conditions" to "for non-extreme or average precipitation events".
- **Line 54. "sensitivities": maybe more clearly "temperature-scaling rate"**
  - **Response:** We agree. We have replaced "sensitivities" with "temperature-scaling rates".
  - **Manuscript modifications:** In `Section 1`, we have changed "sensitivities" to "temperature-scaling rates".

- **Line 57. “hourly extremes”: what kind? Annual Maxima, percentiles, return levels?**
  - **Response:** We agree. We have specified the exact metrics in the text.
  - **Manuscript modifications:** In `Section 1`, we have replaced “hourly extremes” with “hourly return levels and percentiles”.
- **Line 70. “over long period”: this is not generally true for CPM. This is way your study is valuable.**
  - **Response:** We agree. We have rephrased the text to emphasize the high computational cost of long CPM simulations, which makes our 63-year hindcast particularly valuable.
  - **Manuscript modifications:** In `Section 1`, we have updated the text to: “conducting multi-decadal hindcasts is computationally challenging for CPMs, which makes our 63-year simulation particularly valuable”.
- **Line 121. 50km? you mentioned 25-31 km at line 79**
  - **Response:** We thank the reviewer for noting this apparent inconsistency. Our hindcast uses ERA5 boundary forcing at 50 km, whereas the native horizontal resolution of ERA5 is ~31 km. We have clarified this in Sections 1 and 2.2.
  - **Manuscript modifications:** In `Sections 1` and `2.2`, we now state that the simulation is forced by ERA5 at 50 km (native ERA5 resolution ~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 have moved the temperature trend comparison section and Figure 2 to the Discussion section.
  - **Manuscript modifications:** The temperature trend comparison text and Figure 2 have been moved from `Section 3.1` to `Section 5.1.1` under model limitations.
- **Line 133: Evaluation is made per year and season ... and month.**
  - **Response:** We have added “and month” to this description.
  - **Manuscript modifications:** In `Section 2.3`, the text now reads: “per year, season, and month”.
- **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 figure in the main text as an illustration, but we have clarified in the caption that it shows a specific example.
  - **Manuscript modifications:** The caption of Figure 2 has been updated to explicitly state: “This figure illustrates an example of the M_3^* model showing a positive trend”.
- **Line 253. “moderately” capture... in some seasons (highest r=0.4).**

- **Response:** We have toned down the text to use more cautious language.
- **Manuscript modifications:** In `Section 4.1.2`, we have changed "capture" to "partially capture" or "moderately capture" for seasons with low to moderate correlations.

- **Lines 261-264. You mention twice "no trend" for seasons where about 1/3 of stations show significant trend. I guess you intended something like "no clear spatial pattern" or similar. I suggest to clarify/correct.**
  - **Response:** We agree that "no trend" was inaccurate. We have corrected this to describe the lack of a clear regional spatial pattern.
  - **Manuscript modifications:** In `Section 4.2.2`, we have changed "no trend" to "no clear spatial pattern" or "spatially incoherent trends".

- **Line 273. -84.7% in figure 7**
  - **Response:** We have checked the sign and corrected it in the text.
  - **Manuscript modifications:** In `Section 4.2.2`, the sign has been corrected to match the figure.

- **Line 291. Also July and October null/very small**
  - **Response:** We have added July and October to the list of months with very low spatial correlations.
  - **Manuscript modifications:** In `Section 4.2.2`, we have updated the text to: "spatial correlations are almost null in March, May, July, September, and October".

- **Line 345. Where? Same regions where you find underestimation?**
  - **Response:** We have specified that this mismatch occurs primarily in regions dominated by summer convective activity.
  - **Manuscript modifications:** In `Section 5.3.2`, we have added: "especially in mountain regions (such as the Massif Central and the Alps) where summer convective activity dominates".

- **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:** We agree. Grid-box smoothing lowers the absolute intensity values but should not affect relative temporal trends. We have corrected the text to remove grid-box smoothing as an explanation for relative trend underestimation.
  - **Manuscript modifications:** In `Section 5.1.1`, we have removed the reference to spatial smoothing for relative trends and focused on the temperature trend underestimation.

Response to Anonymous Referee #2

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 #2 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.

General Comments

Comment 1: Return Period Choice

There is a 10-year return period chosen for both daily and hourly data, even though 10-year return period for hourly is much more extreme compared to daily given the amount of hourly data in a 10-year period compared to daily. Has a lower return value been tested? Given the noisy result, and the shorter period of data a better justification for the high return period for hourly data should be given, and it is of interest to show a lower return period.

- **Response:** We thank the reviewer for this comment. We agree that estimating a 10-year return level (RL_{10}) at the hourly scale from a 33-year series (1990–2022) is more uncertain than at lower return periods, because each hourly annual maximum carries less information about the tail of the distribution.

As a sensitivity check, we computed relative trends in RL_2 and RL_5 using the same non-stationary GEV models and the same 1992–2022 trend window as for RL_{10} (Response Figure 1 and Response Figure 2). Quantitative comparisons (hydrological year, all stations with a selected GEV model) show that the spatial patterns are robust to the choice of return period, while trend magnitudes increase with T, as expected for higher quantiles:

Daily (1,583 stations; 704 with significant RL_{10} trends, 44.5%). Among significant stations, pairwise Pearson correlations of relative trends are $r = 0.99$ between RL_5 and RL_{10},

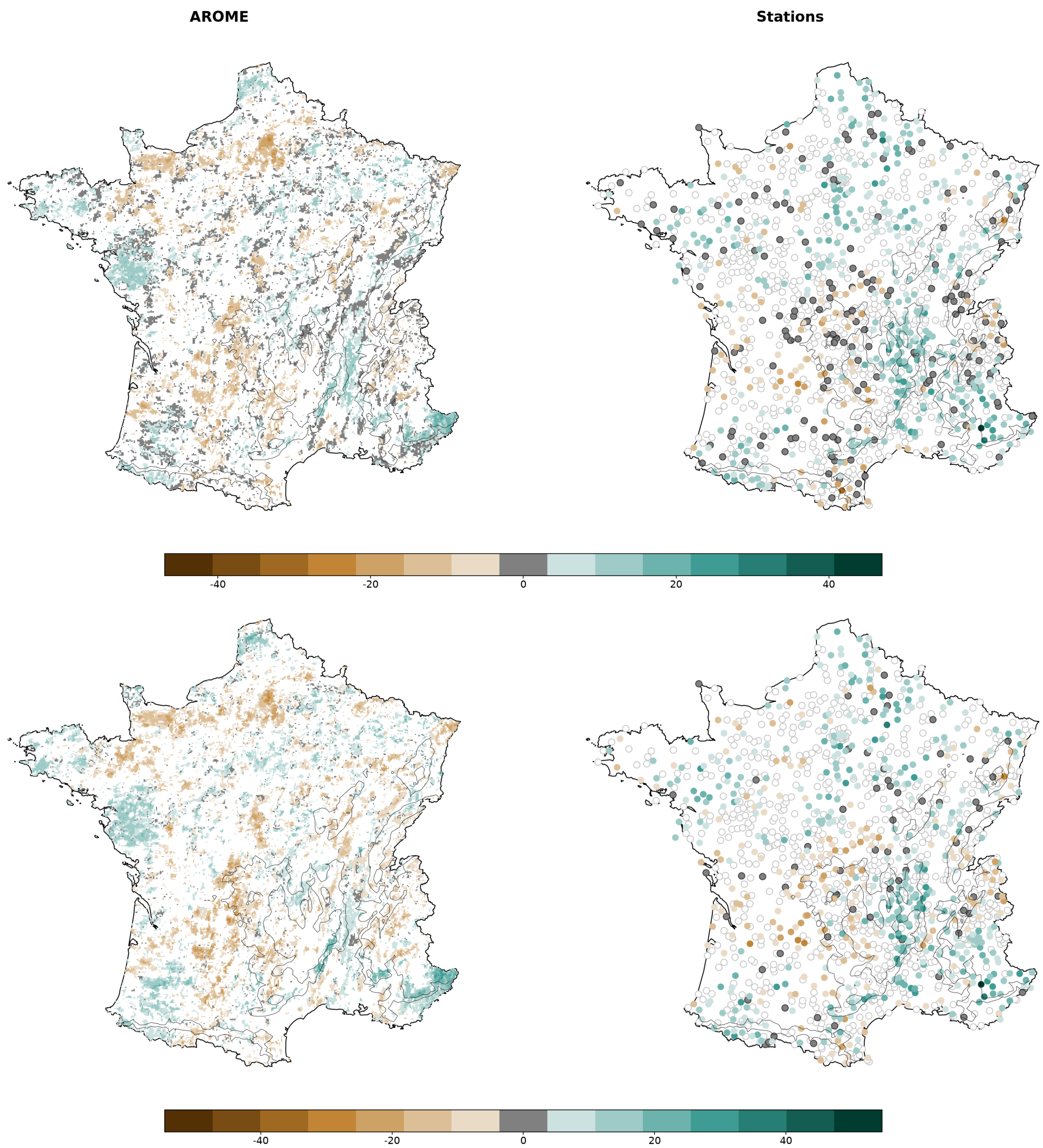
$r = 0.86$ between RL_2 and RL_{10}, and $r = 0.92$ between RL_2 and RL_5. Median absolute trends are similar across return periods (9.3%, 10.9%, and 11.3% for RL_2, RL_5, and RL_{10}, respectively). The same regional signals appear on all three maps—notably positive trends in the southern Alps and Mediterranean arc, and weaker or negative trends in parts of western France—with RL_5 virtually indistinguishable from RL_{10} at the national scale.

Hourly (574 stations; 193 with significant trends, 33.6%). Correlations remain high but are lower than for daily data, reflecting the shorter records and higher convective variability: $r = 0.98$ (RL_5 vs. RL_{10}), $r = 0.79$ (RL_2 vs. RL_{10}), and $r = 0.87$ (RL_2 vs. RL_5) among significant stations. Median absolute trends increase from 31% (RL_2) to 42% (RL_5) and 45% (RL_{10}), confirming that RL_{10} amplifies tail trends but does not create spurious regional structures: the same broad patterns are present at RL_2 and RL_5, with reduced amplitude and less spatial coherence, as expected for a lower quantile on a 33-year record.

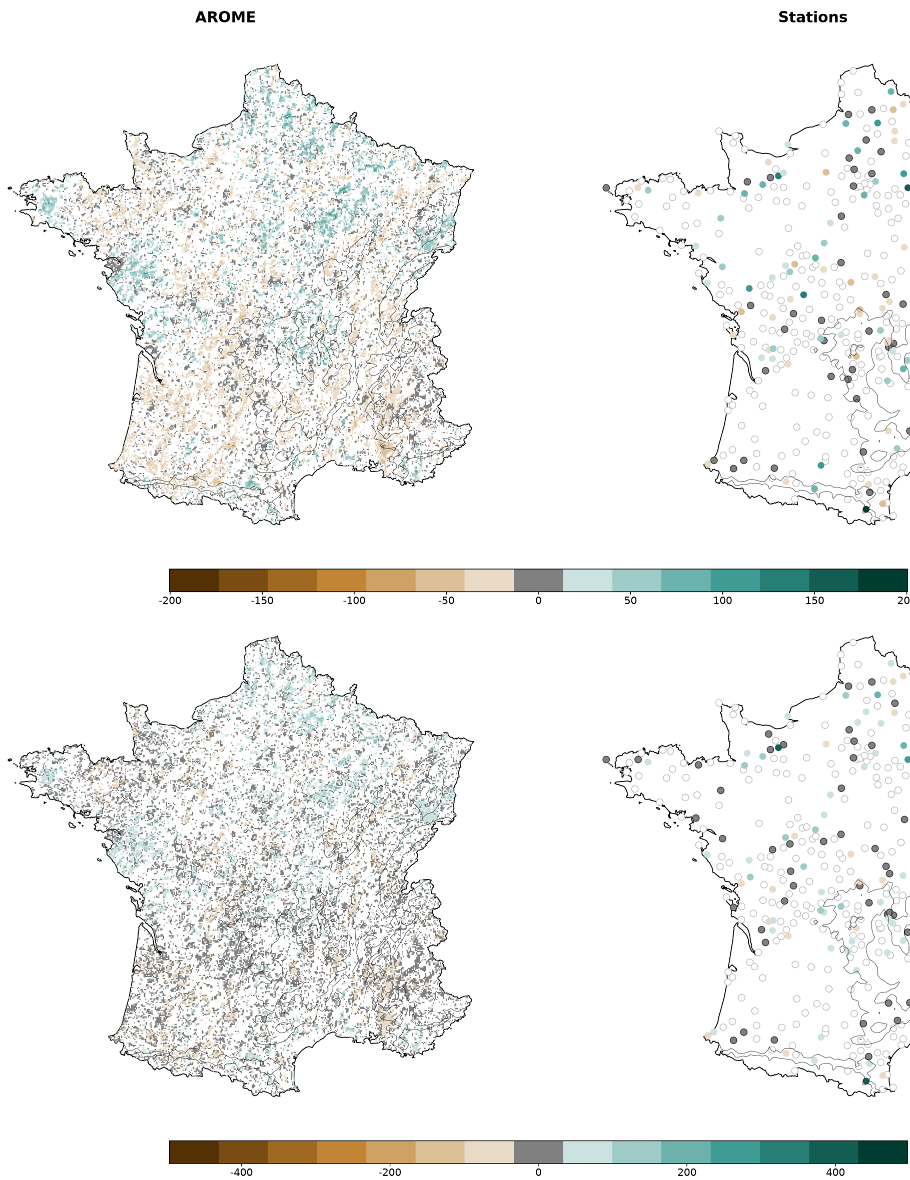
AROME shows the same behaviour: $r = 0.99$ between RL_5 and RL_{10} trends (daily and hourly, significant grid points) and $r = 0.83$–0.86 between RL_2 and RL_{10}, consistent with the station-based analysis.

We note that significance is assessed from the RL_{10} trend uncertainty (10% level); the same 704 daily and 193 hourly stations are therefore flagged across return periods. This is conservative for RL_2 and RL_5, which are less noisy than RL_{10} at the hourly scale. Together with the mean annual-maxima check in our response to Reviewer #1 ($r = 0.88$ for relative trends across 1,558 daily stations), these results indicate that the trend maps (Figures 6 and 7) are not driven by GEV fitting artefacts or by the choice of a high return period alone.

We retain RL_{10} as the primary metric because it is true extreme and because our daily (63-year) and hourly (33-year) records are sufficient for its estimation. In addition, using the same return period at both daily and hourly scales enables a direct and consistent comparison of the extreme trend characteristics between the two time resolutions. The results for RL_2 and RL_5 confirm the robustness of the detected spatial patterns. We present the lower-return-period maps here for the reviewer's assessment rather than as an additional manuscript appendix, to keep the main text focused.



Response Figure 1: Relative GEV trends (%) of RL_2 (first row) and RL_5 (second row) for daily precipitation (hydrological year): columns are AROME (left) and Météo-France stations (right), with a separate colour scale for each row. Display conventions: open circles for non-significant stations ; filled grey for significant near-zero trends; coloured markers for significant trends (uniform size); non-significant AROME grid cells not displayed.



Response Figure 2: Relative GEV trends (%) of RL_2 (first row) and RL_5 (second row) for hourly precipitation (hydrological year): columns are AROME (left) and Météo-France stations (right), with a separate colour scale for each row. Display conventions: open circles for non-significant stations ; filled grey for significant near-zero trends; coloured markers for significant trends (uniform size); non-significant AROME grid cells not displayed.

Manuscript modifications:

- In `Section 3.2.3 (Trend in return level)`, we have added the justification for using the 10-year return level as the main indicator, while noting the consistency with lower return periods:

  > *"In this study, we focus on the 10-year return level (*RL_{10}*) as the primary metric for extreme precipitation. Our daily series (63 years) and hourly series (33 years) provide sufficient data for its robust estimation. As a sensitivity check, trends in* RL_2 *and* RL_5 *show the same regional patterns as* RL_{10} *(daily:* $r = 0.86$–0.99 *among significant stations; hourly:* $r = 0.79$–0.98*), confirming that the detected signals are robust and not driven by GEV fitting uncertainty."*

Comment 2: Introduction & Physical Processes

Introduction: First paragraph introduces a lot of different processes that are not necessarily connected. For instance, global warming leads to an increase in surface temperature, but it's the temperature increase in the atmosphere that increases the water holding capabilities, and the energy balance is also between the top of atmosphere and surface (and everything in between). A lot of the statements in Introduction need a reference for instance line 47-50 and should include more research articles in addition on the IPCC 2021 report.

- **Response:** We agree that the physical description in the first paragraph was somewhat compressed and mixed different thermodynamic and dynamic concepts. In the revised manuscript, we have restructured this paragraph to:

1. Clarify the physical link: surface warming leads to atmospheric warming, which increases the water-holding capacity of the troposphere according to the Clausius-Clapeyron relationship, while the actual condensation and precipitation are driven by vertical motion and adiabatic cooling of ascending air masses.
2. Add more academic references (e.g., Trenberth et al. 2003, O'Gorman and Muller 2010, Westra et al. 2014) alongside the IPCC 2021 report to ground these statements in the literature.

Manuscript modifications:

- In `Section 1 (Introduction and context)`, the first paragraph has been restructured. The main changes are:
  - Replace *"Due to buoyancy (Archimedes' principle), warm air surrounded by cooler air tends to rise. As warm air ascends in the atmosphere, it undergoes adiabatic cooling, leading to the condensation of water vapor and the formation of precipitation (Lin, 2022)."* with a clearer formulation that distinguishes thermodynamic and dynamic contributions.

– Replace *"However, under calm conditions, the central portion of the precipitation distribution does not fully capitalize on this excess moisture."* with *"However, for non-extreme or average precipitation events, [...]"* to clarify the distinction between mean and extreme precipitation.
– Add references: Trenberth et al. (2003), O'Gorman and Muller (2010), and Westra et al. (2014) alongside the existing IPCC (2021) and Clapeyron (1834) citations.
– Replace *"sensitivities"* (line 141) with *"temperature-scaling rates"* (as also requested by Reviewer #1).

Comment 3: Treatment of Aerosols and Solar Brightening

Figure 2 and Figure 3: How much of the temperature trend could be caused by aerosol changes due to anthropogenic pollution? How is aerosol treated in AROME? The results in these plots indicate that there is an increase in temperature after the peak pollution in the 1980s and that the trend in 10-year return level is increasing after this peak. Even though aerosols are not included or kept constant in AROME, this should be stated and included in the discussion and limitations.

- **Response:** We would like to clarify that aerosols are **not** kept constant in this AROME simulation: they are monthly evolving and vary from year to year, taken from the ALDERA reanalysis (which uses an interactive aerosol scheme to capture historical variations, including the peak pollution in the 1980s and subsequent solar dimming/brightening, as described by Nabat et al. 2020). Green House Gases also evolve annually. In the revised manuscript, we have made this clearer in Section 2.2 and Section 5.1.1 (Discussion).

Regarding the origin of the weaker simulated temperature trend, we do not attribute it to the ERA5 boundary forcing, which reproduces observed temperature trends well over France. More plausible internal model limitations include the model's sensitivity to radiative forcings, the absence of interactive vegetation (vegetation properties are prescribed and do not evolve with climate), and the use of a time-invariant land use/land cover map over the 63-year simulation period.

Manuscript modifications:

- In `Section 2.2 (CNRM-AROME model)` (line 183), the existing sentence *"Aerosols are monthly evolving and come from the ALDERA reanalysis, a CNRM-ALADIN simulation at 12 km driven by ERA5 with an interactive aerosol scheme (Nabat et al., 2020). Green House Gases are evolving annually."* already states this. We have expanded it to explicitly mention the peak pollution and dimming/brightening:

  > *"Aerosols are monthly evolving and vary from year to year, taken from the ALDERA reanalysis, a CNRM-ALADIN simulation at 12 km driven by ERA5*

with an interactive aerosol scheme that captures historical variations including the peak anthropogenic pollution in the 1980s and subsequent solar dimming/brightening (Nabat et al., 2020). Greenhouse gases also evolve annually."

- In `Section 5.1.1 (Temperature trends)`, after the existing paragraph on temperature trends (line 207: *"It should be noted that temperature trends [...] the magnitude of the trends we can detect."*), we have added:

  "The weaker simulated temperature trend suggests an internal dampening of the warming signal in France compared to observations. Since aerosols and greenhouse gases evolve realistically in this simulation (Section~2.2) and ERA5 reproduces observed temperature trends well, the underestimation likely stems from the model's internal sensitivity to radiative forcings, the absence of interactive vegetation, and/or the lack of temporal evolution of the land use/land cover map."

Comment 4: Visualization of Non-Significant Trends

For the maps with stations, setting non-significant trends to zeros makes it difficult to differentiate with stations that have close to zero or zero significant trends and stations with strong trends that are non-significant. The authors should consider not showing the stations at all or choosing another marker. The size of the marker also seems to vary with the trend, which can make it look like the stations with strong trends are more numerous than they are.

- **Response:** We agree with the reviewer. Representing non-significant trends as zero makes it difficult to distinguish between a significant trend of 0% and a non-significant trend of 50%. For the climatology maps (Figure 4), we only adopted uniform marker size and a clearly visible grey for zero bias (no significance coding). For the trend maps (Figures 6 and 7, and Appendix C), we adopted the following display conventions (also detailed in our response to Reviewer #1, Comment 7):

Type	Display
Non-significant (stations)	Open circle: white fill, light grey outline
Significant 0 (stations)	Filled grey disk, dark outline
Significant coloured (stations)	Trend colour on the map scale; **uniform marker size**
AROME grid	Non-significant grid points **not displayed**

This keeps the station network visible on hourly maps (where only about one third of stations are significant) while clearly separating non-significant locations from significant near-zero and coloured trends.

Manuscript modifications:

- In `Figures 6 and 7` (trend maps), we have changed the visualization according to the table above.
- The caption of Figures 6 has been updated. Original caption: *"Relative GEV trends over 1959–2022 in the 10-year return level of daily precipitation [...]."* Updated caption adds the display conventions (open circles for non-significant stations, filled grey for significant near-zero trends, coloured markers for significant trends, uniform size; non-significant AROME grid cells not displayed).
- The same changes have been applied to the captions of Figures 6 and 7 (hourly trends) and the monthly trend maps (Appendix C, Figure C1).

The Figure 3 caption has been updated for uniform marker size and zero-bias display; captions for Figures 5 and 6 and Appendix C, Figure C1 describe the significance conventions above. The revised maps are shown in our response to Reviewer #1 (Comment 7): Figure 3 (climatology) and Figures 5 and 6 (trends).

Minor Comments

- **Why choose the hydrological year? The study itself does not take into account snowpack, and in addition the convective extreme precipitation can come in summer that in this study is over the start of the hydrological year (JAS, where summer usually is considered as JJA). This should be justified more.**

  Response: We thank the reviewer for this comment, which highlights an important distinction between two levels of our analysis.

  For **annual maxima** (hydrological year, September to August), the September start is motivated by the need to keep the autumn Mediterranean convective season intact within a single block, as required for the independence assumption of annual maxima GEV modeling. This choice is standard in French climatology and hydrology when studying annual extremes.

  For **seasonal** analyses, the grouping (OND, JFM, AMJ, JAS) is a standard partition in French hydrology for extreme precipitation studies (e.g., Haruna et al., 2026). As shown by Berghald et al. (2025), this grouping is particularly suited for capturing extreme trend patterns compared to traditional meteorological seasons (e.g., SON, DJF).

  To avoid conflating these distinct convective regimes, we also provide **monthly** analyses (Figures and Appendix), which allow summer (June–August) and autumn (September–November) signals to be examined separately. We have clarified this distinction in the methodology section.

  Manuscript modifications: In `Section 3 (Methodology)`, after the definition of seasons and the hydrological year, we have updated the clarification as follows:

"The hydrological year (September to August) is used for annual maxima instead of the calendar year, which is standard in French climatology and hydrology. This choice is motivated by the fact that the most intense extreme precipitation events in France, particularly Mediterranean episodes (e.g., Cévenol events), occur during autumn (September–November). Using a calendar year would risk splitting these autumn events across two years, violating the block-independence assumption required for annual maxima GEV modeling. The seasonal partition (OND, JFM, AMJ, JAS) follows the conventional grouping of months used in France for precipitation extreme studies (Haruna et al., 2026). A previous study by Berghald et al. (2025) demonstrated that this grouping better captures trend patterns compared to the traditional seasonal definition (e.g., SON, DJF)."

- **It can be confusing in the text with GEV models and Arome model, could consider naming one different.**

  Response: We agree. In the revised text, we refer to CNRM-AROME as the "climate model", "simulation(s)", or "CP-RCM" to avoid any confusion, and have specifically replaced "AROME model" with "AROME simulations" in the results sections. We refer to the GEV models as "statistical models" or "statistical GEV distributions".

  Manuscript modifications: Throughout the manuscript, we have revised the terminology to prevent confusion between the climate model and the extreme value models. Specifically:

  - We refer to CNRM-AROME as the "CP-RCM", "climate model", or "simulation(s)", and have replaced "AROME model" with "AROME simulations" in the results sections.
  - We refer to GEV models (M_0 to M_3^*) as "statistical models" or "statistical GEV distributions".

- **Too little font in Figure 8**

  Response: Figure 8 has been moved to the Appendix, where it can be displayed at a larger scale and read more easily.

  Manuscript modifications: Figure 8 (monthly GEV trend synthesis) has been relocated to the Appendix to improve readability.