

Revision for “Climatology and trends of extreme precipitation in France:evaluation of an explicit-convection regional climate model” by Decoopman et al.

This study presents the evaluation of precipitation extremes in France for a 63-year run of the AROME ERA-driven convection-permitting climate model (CPM). In particular, the focus is on daily and hourly 10-yr return levels, both on their climatology and temporal trend. As a benchmark, daily and hourly rain gauge observations are used (minimum 50 year and 25 years record length, respectively).

These objectives are of interest in the hydrological community, given the increasing use of these type of climate models and the need of robust evaluation for extremes. This is still limited, particularly for the observed trend because the lack of longer than 10-20 year simulations. The paper is thus relevant for HESS journal, but I suggest moderate revision. In particular, it could benefit from improved/highlighted information about relative biases (now it’s more focused on correlation metrics), and some updates to figures.

Main comments:

Line 88-89 (but also linked with lines 323-324 in the discussion). 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.

Line 162: I have a few concerns about the non-stationary models. 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?

Line 178: not clear to me why you use the 2-step procedure, why not simply choosing the model with the smallest p-value?

Line 199. 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)

Lines 216-225. 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!

Lines 230. This seems mostly referring to spatial correlation; but also bias should be mentioned (e.g., underestimation at 1h)

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 bias (not much visible as it is now).

Figure 5. I suggest a second panel summarizing %biases for the 4 variables.

Lines 274-285 and figure 8, and figure 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.

DAILY	HOURLY
Trend stations and arome	Trend stations and arome
Relative bias	Relative bias
Correlation	Correlation

Line 298. 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”?

Line 305-308. 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.advwatres.2025.105071>)

Minor comments:

Line 11 (and 395). “added value” ... with respect to what?

Line 19. What does it mean “under calm conditions”?

Line 54. “sensitivities”: maybe more clearly “temperature-scaling rate”

Line 57. “hourly extremes”: what kind? Annual Maxima, percentiles, return levels?

Line 70. “over long period”: this is not generally true for CPM. This is way your study is valuable.

Line 121. 50km? you mentioned 25-31 km at line 79

Lines 126-128. I wonder if this should be placed in the discussion (and a smaller figure, considering the simplicity of its information)

Line 133: Evaluation is made per year and 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).

Line 253. “moderately” capture... in some seasons (highest $r=0.4$).

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

Line 273. -84.7% in figure 7

Line 291. Also July and October null/very small

Line 345. Where? Same regions where you find underestimation?

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)