

Review. Developing and evaluating a Bayesian weather generator for UK precipitation conditioned on discrete storm types by Paul Bell et al.,

Overview.

This paper presents a Bayesian Generalized Linear Model (GLM) weather generator for UK precipitation, conditioned on a newly developed dataset of physically based storm types (cyclones, fronts, and thunderstorms). The authors evaluate the model using proper scoring rules (CRPS, BS, twCRPS) and introduce the Diebold–Mariano (DM) test to assess the statistical significance of performance. Results show that conditioning on storm types significantly improves predictive skill, though the magnitude of improvement is modest. The Bayesian framework allows for transparent uncertainty quantification, and the approach demonstrates good potential for extending WG applications to process-based climate downscaling. However, the paper also highlights calibration issues, with consistent overprediction of moderate-to-heavy rainfall, suggesting room for refinement in the model's distributional assumptions.

➔ Overall, the paper is a well-structured and innovative study that advances both the methodological and interpretive aspects of weather generator development, especially in linking atmospheric processes with statistical modeling. However, some minor clarification are needed.

Questions.

1. The PIT histograms and posterior predictive checks reveal a consistent overestimation of moderate-to-high precipitation amounts. The discussion could identify whether this bias stems from the Gamma assumption, the linear predictor structure, or the aggregation timescale (6-hourly).
2. The inclusion of storm types improves CRPS by $\sim 0.012 \text{ mm } 6\text{h}^{-1}$, a statistically significant so far small gain. The authors should clarify whether this difference has practical consequences for hydrological or risk modeling applications, or whether it primarily demonstrates methodological robustness.
3. The threshold-weighted CRPS (twCRPS) is an interesting choice for assessing extremes, but its sensitivity to limited high-end samples is a concern. The authors could consider (or at least discuss) complementary extreme metrics, such as quantile skill scores or percentile biases, to confirm the robustness of conclusions for heavy precipitation.
4. One of the motivations of this paper is to apply the WG to GCM projections, it would be useful to discuss how storm-type classification might perform on biased model fields, and whether threshold recalibration or bias correction would be required.

Minor comments.

L.13: Continuous Ranked Probable Score (CRPS) ➔ Continuous Ranked Probability Score (CRPS).

L.51: meterological ➔ meteorological.

L.172: where where we expect ➔ where we expect.

Figure 4 caption could clarify what “Observed o*” represents (is it mean observed occurrence?). In general, some figure captions are long and could be more concise.