1 Response to reviewer comments (04/04/2026)

1.1 Response to reviewer 1

I appreciate the authors' hard work in revising the manuscript. The revisions have significantly improved the clarity and quality of the paper. I believe the manuscript could be ready for publication after addressing the following comments:

We thank the reviewer for this positive and useful feedback. We will address each comment point-by-point below.

L210-220: The added rescaling method is interesting, but its formulation and motivation are not yet sufficiently clear. In the response to reviewers, the authors mentioned rescaling the data to [0, 0.9] (which appeared to be a heuristic choice?), but in the revised manuscript, the upper bound of the scaling is computed using an explicit formula. It would be helpful if the authors could elaborate on the reasoning behind this formulation and clarify how this equation is derived. For example, the substitution for 'min' (i.e., $F_G^{-1}(1/R)$) is introduced without explanation. I also wonder if this rescaling bound is the same for every pixel. Are you applying it to each pixel independently or just computing one rescaling bound for the whole domain?

We realise we did not sufficiently address this in the revised manuscript. In the original response to reviewers we rescaled to [0,0.9] but thought that this was too subjective. We have updated this section to better explain the reasoning for this new rescaling method. The new text is:

> An alternative approach that allows us to specify a sensible maximum range for the generated data, using only the training data, is to set the maximum range using quantiles of the training distribution $\mathcal{G}$, which we can specify to correspond specific return periods. For example, we can set it so that the maximum possible value our method can produce, at any point on the domain, corresponds to an intensity that is on average exceeded once every R events at that point (or similarly, once every $R' = \lambda R$ years, with λ the number of events per year). For the distribution $\mathcal{G}$, the minimum and maximum values associated with R are $F_{\mathcal{G}}^{-1}(1/R)$ and $F_{\mathcal{G}}^{-1}(1-1/R)$, respectively. Following this approach, the data can be rescaled according to
>
> $$y' = \frac{y - F_{\mathcal{G}}^{-1}(1/R)}{F_{\mathcal{G}}^{-1}(1-1/R) - F_{\mathcal{G}}^{-1}(1/R)}, \tag{1}$$
>
> assuming that R is sufficiently large that $F_{\mathcal{G}}^{-1}(1/R)$ is smaller than the smallest value in the training data and $F_{\mathcal{G}}^{-1}(1-1/R)$ is larger than the largest value in the training data. While this will result in the same maximum and minimum values across the $\mathcal{G}$-distributed training data, it will map back to different values in physical space.

The ambiguity in this part further makes the application of the method in L330-335 unclear. Additionally, the motivation for multiplying by log(R)/log(N) for untransformed data is not immediately clear.

We have clarified this text to follow more naturally from Section 2 and added more context for the log(R)/log(N) scaling. We have updated the text to:

> To rescale the training data to the interval $(0,1)$ for training a deep learning model, we use the return period-based rescaling method described in Sect. 2. We used a 1-in-$R = 10{,}000$ event return level to specify the maximum value any generated marginal can reach (choosing event-based rather than year-based scaling for simplicity). This was converted into a maximum value of $1-1/R$ for the uniform distribution and $F_{\mathcal{G}}^{-1}(1-1/R)$ for the Gaussian and Gumbel distributions. The lower bounds were calculated analogously, *i.e.*, using a minimum value of $1/R$ for the uniform distribution and $F_{\mathcal{G}}^{-1}(1/R)$ for the Gaussian and Gumbel distributions. As a heuristic, the no-transform data is rescaled to $(0,1)$ by setting 1 to correspond to the maximum value multiplied by $\log(R)/\log(N)$, where N is the number of independent hazard events—approximating the ratio of the R-event to N-event maximum under Gumbel tails. While we note this is inexact, it suffices for our comparative purposes.

Figure 5: Using white pixels to indicate AD test failure is not appropriate since white is already included in your colorbar.

All white has been removed from the colorbar so NA pixels are now clear.

1.2 Response to reviewer 2

I am a new reviewer in the second round of reviews. The manuscript "Simulating spatial multi-hazards with generative deep learning" presents a novel approach to simulate storm events using generative adversarial networks (GANs). The study focuses on the development of a GAN combined to extreme value distributions in order to simulate realistic extreme events. The application to storm events in the Bay of Bengal further illustrates and evaluates the ability of this approach to replicate many features of the most extreme storm events, including severity measures. Overall, I find the manuscript well documented, the choice of the tools are clearly motivated, and many interesting results are obtained. I have just some comments which mainly aim at providing a more objective presentation of the study and its limitations.

We thank the reviewer for their thoughtful and constructive review and for their positive comments regarding the manuscript.

Compound events: The introduction starts with a discussion on compound events, which seems the main motivation for the study. However, compound events are not necessarily caused by multiple simultaneous extreme hazards. Compound events can arise from various combinations of low- to moderate-intensity hazards linked to specific geographical, geomorphological, cryospheric, and meteorological conditions (https://doi.org/10.5194/egusphere-2026-971, https://onlinelibrary.wiley.com/doi/full/10.1002/

Persistent meteorological conditions (long droughts or very wet conditions) often lead to a greater severity of many natural hazards. For example, the most disastrous flooding events are often the consequence of intense precipitation and saturated soils (https://www.nature.com/articles/s43247-025-02691-6). Temporal dependence or Nonconcurrence of hazards are thus key aspects of many compound events. The persistence is not treated in this study and this aspect should be discussed in Section 5 and in the conclusions.

Thank you for these useful references. We agree that combinations of low- to moderate-intensity conditions or temporally compounding events (arising from persistent conditions or events clustered in times) are of critical importance; however, we consider them largely out of scope for this work. We have altered the first paragraph of the introduction to clarify that we are only focused on multivariate and spatially compounding hazards and not preconditioned or cascading types (flagging that we are following Zscheischler et al. (2020)'s typology, where readers can find more detailed discussions about compound hazards). We have also added some thoughts to the discussion about the degree to which the model can represent some limited temporal conditions, e.g., 30-day antecedant rainfall.

New first paragraph:

> Hazards that entail several extreme weather variables and extend over large spatial domains are responsible for some of the most damaging natural catastrophes (Zscheischler et al., 2018). Compound hazards can be broadly grouped into four typologies: preconditioned, temporally compounding, multivariate, and spatially compounding (Zscheischler et al., 2020). This article focuses on the multivariate and spatially compounding types, specifically on the co-occurrence of hazard extremes across space or across variables. Multivariate or *multi*-hazards (UNDRR, 2017) occur when [...]

Additional discussion:

> The method developed in this paper is suitable for generating the multivariate and spatially compounding varieties of compound hazards (Zscheischler et al., 2020), but does not address preconditioned or temporally compounding hazards. While the temporal problem is generally beyond the scope of this work, it may be possible to represent some types of preconditioning effects by simply using a longer temporal window for the temporal aggregation function $h_{k|t}(\mathbf{x})$. Letting, for example, one field represent 30-day cumulative rainfall before a storm. Otherwise, extending this method to capture either cascading events or the temporal dynamics within storms is more challenging and would require a significant extension of the method.

- Stochastic weather generators: The paragraph at l. 33-48 in the introduction is unclear and should be revised in my opinion.

- First, it should be indicated that “standard” (i.e. parameterized) stochastic weather generators are able to generate stochastic fields that take into account the spatial dependence and the dependence between multiple weather variables (https://doi.org/10.1007/s00477-024-02897-8).

- In these paragraphs, some arguments given for the limitations of the models (dimensions, do not ‘look like’ realistic weather) are not well founded (see comments below).

- However, I agree that these models struggle to adequately represent the asymptotic dependence structures and the extremal dependence between multiple variables.

We have changed the first part of the paragraph in question to:

> Alternatively, a purely stochastic approach to hazard modelling can be taken by fitting statistical models to weather data and simulating large events sets from those models (Wilks and Wilby, 1999). Recent advances in stochastic weather generation have enabled the generation of multivariate and spatiotemporal fields (Obakrim et al., 2025); however, adequately representing the extremal behaviour of multiple weather variables generally requires the use of specialised methods from multivariate extreme value theory (Davison et al., 2019; Lamb et al., 2010, 2019; Speight et al., 2017; Becher et al., 2023).

The ‘look like’ comment is addressed separately.

The GAN proposed in this study aims at improving the simulations of weather events where one variable exceeds a certain threshold. This can be of interest in many applications but it should be indicated that stochastic weather generators simulate not only extreme events, but long continuous time series. For many applications, these long continuous time series are essential to assess some natural hazards (e.g. antecedent conditions for flooding events, https://doi.org/10.5194/egusphere-2024-909).

We agree that this is a critical problem although we consider it mostly out of scope for this work. We have altered the introduction to make it clear from the outset that this paper only attempts to model spatial and multivariate extremes.

Estimates of the GPD parameters: I disagree with some interpretations for the parameter estimates of the GPD. At l. 313-315, it is indicated that “Additionally, the shape parameter estimates for all variables are within the range [-0.5, 0.5], which suggests that the marginal distributions are light-tailed, meaning they decay relatively quickly and are less prone to producing extremely large values.”

GPD with shape parameters above 0.2 are generally considered heavy-tailed, and for shape parameters above 0.5, the GPD is so heavy-tailed that it does not have a variance anymore. Furthermore, the spatial pattern of the shape parameter is very patchy in Figure 5 and indicates that these estimates are very uncertain, especially for the total precipitation.

The color scale does not help to interpret these results, and I would suggest using a color scale that distinguishes cases where the GPD has an upper bound (negative shape par.) and cases where it has a heavier tail (positive shape par.), with a darker color when xi exceeds 0.3.

We have updated the colour scale as suggested. The text has also been updated to reflect new insights from the colour map. Namely that negative shape parameters are predominantly onshore, and the majority of other variables are light or heavy-tailed. A patch of very heavy-tailed rainfall stands out in the southwest of the domain. In future, a nonstationary model or pooled fit may improve some these tail estimates.

Limitations: Section 5 seems like an extended summary of the study. I would have expected a more critical discussion of the pros and cons of this approach (numerical aspects, application to other sets of variables, etc.). In my opinion, it would be important to stress that only simultaneous extreme variables can be simulated.

We have significantly revised the discussion to address this comment, removing some of the summarising text and included more discussion of the limitations of using small datasets and reiterating that this method does not address temporal dynamics.

l. 38, 42: In Quinn et al., 2019, 2,400 is the number of gauge series records. In Huser and Wadsworth, 2022), 3,600 is the dimension of satellite rainfall measurements. I do not understand why these numbers are given as limiting factors for the dimensions in applications of stochastic approaches since they were just the dimensions of the observations.

We have updated the text to make sure it is not implying that these are hard limits, but rather the highest-dimensional studies we are aware of, some which required significant computational resources. The updated text is:

> However, the conditional exceedance model remains limited in its ability to scale to high dimensions (see, for example, Quinn et al. 2019, who required >800 CPU cores to model the dependence between 2,400 river gauges over the U.S.) and cannot be used to generate hazard scenarios at new locations. Spatial models such as max-stable models and r-Pareto processes address this using geostatistical methods to parametrise dependence structures across the spatial domain. The gradient-based r-Pareto processes of de Fondeville and Davison (2018) are particularly powerful, and have successfully modelled up to 3,600 dimensions, although this is not necessarily an upper bound on their potential (Huser and Wadsworth, 2022). Most spatial process models, however, suffer a trade-off between flexibility and scalability: many [...]

l. 48: "yet the events do not necessarily 'look like' realistic weather, as might be observed in a rainfall radar." This sentence seems to be a very subjective statement. In my experience, there are advanced stochastic weather generators that use Gaussian fields and simulate stochastic fields that look similar to radar images.

We have refined this sentence to be more grounded in the literature:

> A further challenge is that the majority of spatial statistical methods model pairwise dependence structures (Davison et al., 2012), entailing a loss of information (Serinaldi et al., 2015) and making it challenging to capture complex high-order meteorological features such as rainbands, spiralling vortices, and fronts. Papalexiou et al. (2021) for example, couple random fields with velocity fields and anisotropy in order to model such features in a spatiotemporal setting.

l. 51: The acronym GAN should be introduced here.

That has been added.

Introduction: In a recent paper, GANs have been used to downscale CMIP6 simulations of hourly precipitation (https://agupubs.onlinelibrar

Thanks, we have added this to the introduction.

l. 231: "augmentationsto"

This has been fixed, thanks.

l. 257: Missing period.

This has been fixed, thanks.

Figure 11: Pearson's correlations help to assess the similarity of the spatial patterns but it does not indicate that inter-variable extremal correlation estimates are generally lower in ERA5. It would be fair to indicate the mean differences between the inter-variable extremal correlation estimates obtained with ERA5 and with hazGAN.

Added to text:

> In all cases, the generated data had a tendency to overestimate the strength of the inter-variable dependence, with the mean overestimation (across all pairs of variables and locations) of 0.292 (uniform), 0.256 (Gaussian), 0.205 (Gumbel), and 0.241 (rescaling-only).

Figure 12: The caption is incomplete.

This has been added.

References

Becher, O., Pant, R., Verschuur, J., Mandal, A., Paltan, H., Lawless, M., Raven, E., and Hall, J.: A multi-hazard risk framework to stress-test water supply systems to climate-related disruptions, Earth's Future, 11, https://doi.org/10.1029/2022EF002946, 2023.

Davison, A. C., Padoan, S. A., and Ribatet, M.: Statistical modeling of spatial extremes, Statistical Science, 27, 161 – 186, https://doi.org/10.1214/11-STS376, 2012.

Davison, A. C., Huser, R., and Thibaud, E.: Spatial extremes, in: Handbook of Environmental and Ecological Statistics, edited by Gelfand, A. E., Fuentes, M., Hoeting, J. A., and Smith, R. L., CRC Press, 2019.

de Fondeville, R. and Davison, A. C.: High-dimensional peaks-over-threshold inference, Biometrika, 105, 575–592, 2018.

Huser, R. and Wadsworth, J. L.: Advances in statistical modeling of spatial extremes, Wiley Interdisciplinary Reviews: Computational Statistics, 14, e1537, https://doi.org/10.1002/wics.1537, 2022.

Lamb, R., Keef, C., Tawn, J., Laeger, S., Meadowcroft, I., Surendran, S., Dunning, P., and Batstone, C.: A new method to assess the risk of local and widespread flooding on rivers and coasts, Journal of Flood Risk Management, 3, 323–336, https://doi.org/10.1111/j.1753-318X.2010.01081.x, 2010.

Lamb, R., Garside, P., Pant, R., and Hall, J. W.: A probabilistic model of the economic risk to Britain's railway network from bridge scour during floods, Risk analysis, 39, 2457–2478, https://doi.org/10.1111/risa.13370, 2019.

Obakrim, S., Benoit, L., and Allard, D.: A multivariate and space-time stochastic weather generator using a latent Gaussian framework, Stochastic Environmental Research and Risk Assessment, 39, 3677–3701, https://doi.org/10.1007/s00477-024-02897-8, 2025.

Papalexiou, S. M., Serinaldi, F., and Porcu, E.: Advancing Space-Time Simulation of Random Fields: From Storms to Cyclones and Beyond, Water Resources Research, 57, e2020WR029 466, https://doi.org/10.1029/2020WR029466, 2021.

Quinn, N., Bates, P. D., Neal, J., Smith, A., Wing, O., Sampson, C., Smith, J., and Heffernan, J.: The spatial dependence of flood hazard and risk in the United States, Water Resources Research, 55, 1890–1911, https://doi.org/10.1029/2018WR024205, 2019.

Serinaldi, F., Bárdossy, A., and Kilsby, C. G.: Upper tail dependence in rainfall extremes: would we know it if we saw it?, Stochastic Environmental Research and Risk Assessment, 29, 1211–1233, https://doi.org/doi.org/10.1007/s00477-014-0946-8, 2015.

Speight, L., Hall, J., and Kilsby, C.: A multi-scale framework for flood risk analysis at spatially distributed locations, Journal of Flood Risk Management, 10, 124–137, https://doi.org/10.1111/jfr3.12175, 2017.

UNDRR: The Sendai Framework Terminology on Disaster Risk Reduction: Hazard, https://www.undrr.org/terminology/hazard, [Accessed: 9 February 2026], 2017.

Wilks, D. S. and Wilby, R. L.: The weather generation game: a review of stochastic weather models, Progress in Physical Geography, 23, 329–357, https://doi.org/10.1177/030913339902300302, 1999.

Zscheischler, J., Westra, S., Van Den Hurk, B. J., Seneviratne, S. I., Ward, P. J., Pitman, A., AghaKouchak, A., Bresch, D. N., Leonard, M., Wahl, T., et al.: Future climate risk from compound events, Nature Climate Change, 8, 469–477, https://doi.org/10.1038/s41558-018-0156-3, 2018.

Zscheischler, J., Martius, O., Westra, S., Bevacqua, E., Raymond, C., Horton, R. M., van den Hurk, B., AghaKouchak, A., Jézéquel, A., Mahecha, M. D., et al.: A typology of compound weather and climate events, Nature Reviews Earth & Environment, 1, 333–347, https://doi.org/10.1038/s43017-020-0060-z, 2020.