Rebuttal of Review 2

Dear reviewer,

We appreciate the time and effort taken to review our manuscript. Please, find below our answers to your comments.

General Comments

While the manuscript primarily reviews existing concepts of return periods for compound events, it is unclear how these concepts can be generalized to all typologies of compound events. Specifically, the application of return periods to preconditioned and temporally compounding events is not well addressed. To achieve the stated goal of extending the concept to any type of compound event, the manuscript would benefit from illustrative examples that demonstrate this applicability.

Our study follows the classification by Zscheischler et al. (2020) and emphasizes that preconditioned, multivariate, temporally compounding, and spatially compounding events can all be analyzed within a unified conceptual framework under our methodology.

While the manuscript already employs a multivariate approach that captures dependencies between different variables, we recognize the importance of explicitly clarifying how our methodology applies to each typology of compound events. To enhance clarity, we will refine the manuscript by explicitly stating that all typologies of compound events can be framed within the same conceptual approach, ensuring a unified interpretation. Additionally, we will clarify how our methodology applies to each typology described in Zscheischler et al. (2020), reinforcing that they can all be analyzed within the same methodological framework. To further support this, we will provide a detailed description of how the method is applied in each case:

1. Preconditioned events: These events are modeled by incorporating the initial system state as a variable within the dependency analysis. For example, in the case of flooding, where the occurrence of an extreme event depends on prior soil moisture conditions, this factor is explicitly included in the return period estimation through conditional copulas or probability functions. This allows for the assessment of how antecedent conditions influence the likelihood of extreme events. After the selection process, a multivariate set of events is obtained, preserving the influence of prior conditions.

2. Multivariate events: Our methodology is specifically designed to capture dependencies between multiple simultaneous extreme variables, such as the concurrent occurrence of intense precipitation and storm surges within the same system. Multivariate copulas are used to characterize these interactions and compute joint return periods, reflecting the probability of multiple

drivers occurring together. Once the selection process is completed, a multivariate set of events is obtained, maintaining the dependency structure between extreme variables.

3. Temporally compounding events: These events are handled by incorporating the relationship between extreme events occurring within a short time frame. The dependency between consecutive events is accounted for, allowing us to assess how the occurrence of one extreme event increases the probability of another occurring shortly after. This is reflected in the estimation of cumulative return periods, which consider event persistence and its impact on system recovery. At the end of the selection process, a multivariate set of events is obtained, preserving temporal dependencies and the sequential occurrence of extreme events.

4. Spatially compounding events: Our methodology enables the assessment of correlations between extreme events in different regions by incorporating spatial dependency structures. Spatial copula models are applied to estimate the joint probability of extremes occurring across multiple locations, allowing for the calculation of regional return periods that capture the simultaneous occurrence of events in different geographic areas. After the selection process, a multivariate set of events is obtained, maintaining the spatial correlation structure among the analyzed regions.

All the categories described above, end up as a collection of variables whose joint distribution needs to be characterized by means of copulas. That is, mathematically, all the typologies end up being the same problem, where copulas need to be fit and the fitted copula analyzed to derive the joint return periods. Therefore, the distinction between typologies comes from the interpretation of the results, but methodologically all the four types can be treated with the same methodology, without making any methodological different except for the event selection phase.

Additionally, we would like to highlight that our work is structured into three parts, which demonstrates the applicability of our methodological framework to different types of compound events:

- The first part, presented in this manuscript, develops the methodological framework for estimating return periods in compound events.
- The second part, addressed in a complementary study, applies this methodological framework specifically to spatially compounding events.
- The third part, although not submitted to this journal, applies the methodology to analyze multivariate events in an estuarine region, aiming to characterize flood risk considering multiple simultaneous drivers.

Although the manuscript aims to include machine learning techniques, their description lacks detail and validation. The following areas require clarification and elaboration:

Page 17, Points 5C-5D: The manuscript discusses the computation of copula values in higher dimensions via integration. However, the role of Gaussian processes in this context is insufficiently explained. The algorithm is not described in detail, its numerical performance is not evaluated, and there is no validation of the method.

Gaussian Processes are used as a universal approximator or emulator of the func-

tion that we want to evaluate. Since computing values of a multivariate cumulative distribution function (CDF) is a costly endeavor, the Gaussian Processes serve to approximate such a non-linear function with less computational cost.

To improve the clarity of the manuscript, we will include a more detailed methodological description of Gaussian Process Regression (GPR) in the supplementary information. This will provide a better understanding of the role of GPR in our methodology and its applicability in estimating computationally expensive functions.

The use of GPR has been extensively studied and applied in the approximation of high-cost computational functions in various contexts. For example, Ba and Joseph (2012) developed a composite Gaussian process model to emulate computationally expensive functions, capturing both global trends and local details.

Additionally, Zhuang et al. (2025) applied GPR in the context of high-dimensional American option pricing, highlighting its potential to mitigate computational challenges in high-dimensional settings. Their study demonstrates that GPR-based approaches can be adapted to efficiently model complex systems without a significant increase in computational cost, which aligns with our use of GPR to approximate the Vine copula CDF while maintaining computational efficiency.

These studies are examples of the effectiveness of Gaussian processes in approximating computationally expensive functions, supporting their application in our study. We believe this addition will clarify the role of GPR within our methodological framework and provide a stronger justification for its use.

The validation and calibration performance is carried out in Part II of the paper, since this first part is a detailed description of the methodology, presenting potential variations and citing all the relevant sources that we know about.

Page 18: The determination of the highest probability density point in the critical layer is attributed to "computational optimization techniques," but these techniques are neither named nor described.

Once the critical layer has been defined, there are two potential ways forward to define events of interest. The first one is to determine the most likely event from the critical layer. To determine this event, we have used the MLE (Maximum Likelihood Estimation) algorithm of the Spotpy Python library, which -in spite of its name- does not maximize the statistical likelihood, but it does a classical gradient descent optimization -considering the derivatives of the function-. This procedure works well for up to three dimensions, because above that number there is a high probability of it getting stuck in a local maximum.

The alternative strategy is to not look for the most likely event, but simply to sample events from the critical layer proportionally to their likelihood. However, this procedure faces two complications. First, there is a need to compute the value of the CDF at each point, which implies a multidimensional numerical integration, which is highly time consuming. Second, to sample from the critical layer we need the CDF for all the events at the layer, which is difficult to calculate accurately.

To remediate the first problem, we use Gaussian Process Regression (GPR) as a non-linear approximator to the real value of the multivariate CDF. We compute the value of the CDF using Monte Carlo integration methods for a number of points that serve as interpolation basis for GPR. Then, this methods provides approximated values of the CDF, speeding up the process. The number of points

of the interpolation base serve to control the approximation error to any desired level.

The second problem is dealt with using the Metropolis-Hastings algorithm, which allows us to sample from complicated probability distributions where the CDF cannot be computed. Using an auxiliary probability distribution, this method allows us to generate a sample that grows iteratively, which, in the end, converges to the desired unknown distribution.

Sampling for the critical layer in this way, allows us to generate a collection of events from which the most likely one -or an event almost indistinguishable from the most likely one- can be obtained. Even more, since a complete collection of events is obtained, multiple events related to the same return period can be analyzed and see how the effect of such events may differ in the impact variable.

To improve the clarity of the methodology presented in the manuscript, we will incorporate the following description in the methodological section:

The identification of the most probable design event within the critical layer was performed following two complementary approaches, depending on the dimensionality of the problem.

First, for cases involving up to three dimensions, we employed Maximum Likelihood Estimation (MLE) as implemented in Spotpy (Spotpy, 2024). Contrary to its name, this algorithm does not maximize the statistical likelihood but rather applies a classical gradient-descent optimization, leveraging function derivatives to iteratively refine the estimate of the most probable event. This method is computationally efficient; however, for problems beyond three dimensions, it presents a high risk of converging to local maxima, reducing its effectiveness.

As an alternative, when dealing with higher-dimensional spaces, we employed a sampling-based strategy that does not rely on direct maximization but instead proportionally samples events from the critical layer based on their likelihood. This approach, however, faces two main computational challenges:

1. *The need to evaluate the cumulative distribution function (CDF) at multiple points, requiring multidimensional numerical integration, which is computationally expensive.*
2. *The necessity to estimate the CDF for all events within the critical layer, which is challenging to compute accurately.*

To address the first issue, we leveraged Gaussian Process Regression (GPR) as a surrogate model to approximate the true multivariate CDF. We computed reference CDF values using Monte Carlo integration over a selected set of interpolation points, which were then used to train the GPR model. This approach significantly reduces computational costs while maintaining control over approximation errors.

For the second issue, we utilized the Metropolis-Hastings algorithm, which allows us to efficiently sample from complex probability distributions where direct CDF computations are infeasible. By iteratively refining an auxiliary probability distribution, this method generates a representative sample of events that converge to the true distribution of extreme events in the critical region.

This refined sampling process allows us to obtain a collection of extreme events, from which we can extract the most probable one—or an event nearly indistinguishable from it. Furthermore, this method enables the analysis of multiple events with

the same return period, providing deeper insight into how different realizations of extreme events may impact the target variable.

The complete implementation of this methodology, including its application in high-dimensional contexts, is detailed in Part II of this study, where we validate its effectiveness and illustrate its practical relevance in extreme event characterization.

We believe that these modifications will contribute to a better integration between Part I (methodological framework) and Part II (case study) while enhancing clarity in the presentation of the optimization process. We sincerely appreciate the reviewer's observation, which has allowed us to strengthen the exposition of our methodology.

Section 3.6.2: The manuscript references the use of the Metropolis-Hastings algorithm. However, no algorithm details are provided, and no numerical validation or examples are included.

We propose modifying Section 3.6.2 of the manuscript to include a more detailed description of the Metropolis-Hastings algorithm, which was previously mentioned in line 475 of Part I of the article. However, we acknowledge that the original manuscript did not include the proper reference to the foundational work on the method. This will be corrected by citing Hastings (1970), who formalized the algorithm, along with the original formulation by Metropolis et al. (1953). In addition to this correction, we will clarify its implementation and provide numerical evidence to support its functionality. The suggested wording for the revised manuscript is as follows:

"To obtain a representative sample of the probability distribution within the critical layer, the Metropolis-Hastings algorithm is implemented (Metropolis et al., 1953; Hastings, 1970). The process begins with the selection of an initial event within the critical layer, from which a sequence of events is generated using an auxiliary proposal distribution. In each iteration, a new candidate event is proposed, and its probability density is compared to that of the current event. If the candidate has a higher density, it is automatically accepted; otherwise, it is accepted with a probability proportional to the ratio of both densities. This iterative procedure allows the construction of a sample that, after a sufficient number of iterations, converges to the target distribution, ensuring an adequate representation of the variability in extreme events.

In Part II of this study, numerical validation is presented to demonstrate the effectiveness of the method in generating representative samples of the probability distribution. This enables both the identification of the event with the highest density and the analysis of multiple events associated with the same return period."

With this modification, we aim to improve the methodological clarity of the process, particularly by emphasizing the selection of the proposal distribution, as it plays a key role in the efficiency of the sampling procedure and the convergence to the target distribution. This revision provides a clearer understanding of the role of Metropolis-Hastings in our methodology.

In summary, while the manuscript provides a comprehensive review of return period concepts, further clarification and illustrative examples are needed to demonstrate the applicability of these concepts to diverse types of compound events. Additionally, the integration of machine learning techniques should be substantiated with detailed methodologies and numerical validations.

We appreciate the reviewer's comment and agree that including additional illustrative examples can improve the understanding of the applicability of the concepts presented in the manuscript. We will revise the text to incorporate more examples where possible and relevant.

However, we would like to emphasize that this manuscript is the first part of a three-part study. In particular, the second part of this study, which is currently under review, focuses on applying the methodology to spatially and temporally compounding events, providing detailed numerical validations and case studies. Additionally, there is a third part of the study, which, although not submitted to this journal, applies the same methodological framework to multivariate compound events in estuarine environments, considering river discharge, wind, and marine/coastal variables to analyze interactions between different extreme phenomena affecting estuarine dynamics.

Since this manuscript primarily focuses on methodological formulation, most of the validation and practical applications are presented in the second part of the study.

Specific comments

Section 3.1.2: It is unclear to me the difference between parametric and non-parametric measures of dependence. I guess that the distinction is between rank-invariant measures and other measures like Pearson's correlation. Also the intuitive distinction between Spearman's rho and Kendall's tau at line 150 should be better explained.

We appreciate the reviewer's insightful comment and agree that a more detailed explanation of these concepts will enhance the manuscript's clarity. Differences between Parametric and Non-Parametric Dependence Measures:

- Parametric Measures: These measures, such as Pearson's correlation coefficient, assume a linear relationship between variables and require that the data follow a normal distribution. They are sensitive to outliers and may not effectively capture non-linear relationships.

  Non-Parametric Measures: These include coefficients like Spearman's rho and Kendall's tau, which do not assume a specific distribution and are based on the ranks of the data rather than their actual values. This makes them more robust to non-normal distributions and outliers, allowing them to capture monotonic relationships that are not necessarily linear.

Differences between Spearman's Rho and Kendall's Tau:

- Spearman's Rho: Calculated as the Pearson correlation coefficient between the ranked variables, Spearman's rho assesses how well the relationship between two variables can be described using a monotonic function. It is sensitive to differences in ranks and can be influenced by the presence of tied ranks.
- Kendall's Tau: This coefficient measures the difference between the probability of observing concordant versus discordant pairs in the data (Okoye and Hosseini 2024). It is considered more robust in the presence of ties and provides a more direct probabilistic interpretation of the strength of association between two variables.

To address this observation in the manuscript, we will implement the following revisions:

1. Enhance the explanation in Section 3.1.2 by explicitly differentiating between parametric and non-parametric dependence measures, supported by the relevant references.
2. Provide a clearer intuitive explanation around line 150 regarding the conceptual differences between Spearman's rho and Kendall's tau, highlighting their respective applications and limitations, and substantiating with appropriate citations.

Section 3.2: please, notice that AIC is not a test, but a selection criterion.

We have corrected the wording to make this point more clear.

Eq. (1): It is unclear what is x and what is R_i(x).

We appreciate the reviewer's observation and acknowledge that an adjustment in the mathematical notation is necessary to enhance clarity and ensure a precise interpretation of the expressions used.

In this context, $R(x)$ represents the rank of the observation x within the dataset. As described in Section 3B of the manuscript, to compute the empirical copula, it is common practice to transform the original variables into the standard uniform space ($[0, 1]$). This transformation is performed using the expression:

$$\widehat{F}(x) := \frac{R(x)}{n+1}$$

where $R(x)$ is the rank of x among all observations for a given variable, and n is the total number of observations. By applying this transformation, the original variables $(X, Y, ..., W)$ are converted into pseudo-observations $(U_X, U_Y, ..., U_W)$, which follow a uniform distribution in $([0, 1])$.

To address this observation in the manuscript, we will implement the following revisions:

1. Clarify the definition of $R(x)$ in the text, explicitly explaining its role in the variable transformation for empirical copula construction.
2. Include a brief note in the corresponding equation, ensuring that readers understand that $R(x)$ represents the rank of the observation, rather than an additional function.

   These modifications will improve the clarity of the manuscript and facilitate the interpretation of the notation used.

Eq. (3) and (4): It is difficult to understand in which sense X and Y are conditional events rather than random variables. Analogously, F_X(X) should be a random variable and not a conditional event.

We acknowledge the reviewer's comment and recognize that the notation in Equations (3) and (4) was not presented clearly. Specifically, the way X and Y were defined may have led to the incorrect interpretation that they represent conditional events rather than random variables.

In our approach, X and Y are continuous random variables, while $U = F_X(X)$ and $V = F_Y(Y)$ correspond to their transformation via the probability integral transform. This process allows any continuous random variable to be mapped into the uniform space $[0, 1]$ and is widely used in copula modeling to represent dependence structures without imposing assumptions on the marginal distributions.

To improve clarity in the manuscript and avoid any possible ambiguities, we will implement the following modifications:

1. Clarify in the text that X and Y are random variables, while U and V result from applying the probability integral transform, ensuring they follow a uniform distribution in $[0, 1]$.
2. Reword the explanation of Equations (3) and (4) to prevent any misinterpretation regarding the term "conditional events."

We believe these adjustments will enhance the clarity of the manuscript and ensure that the interpretation of Equations (3) and (4) is well understood. We sincerely appreciate the reviewer's insightful suggestion, which will undoubtedly contribute to improving the presentation of our work.

Eq. (5): W is not the copula function but the random variable defined by C(U,V).

In this context, $W = C(U, V)$ is a univariate random variable and not the copula function itself. Equation (5) defines the Kendall distribution function, which is expressed as:

$$K_c(t) = \Pr[W \leq t] = \Pr[C(U, V) \leq t]$$

where W represents a scalar value derived from the copula function, and t is a probability threshold that separates the supercritical and non-critical regions.

To improve clarity in the manuscript and avoid any misinterpretation, we will implement the following adjustment:

1. Explicitly state in the text that W is a univariate random variable derived from the copula and not the copula function itself.

We believe this modification will enhance the understanding of the role of W in Equation (5).

Sections 4 and 5 are quite similar. They should be merged and their scope should be better defined.

We do agree with the referee that the Discussion and Conclusions sections do not belong in the paper, since, without reading Part II, it is difficult to provide solid conclusions to our study. We have removed both sections and instead included two new ones: *Problems and Limitations* and *Concluding Remarks*.

The former will address all the aspects discussed throughout this review that may challenge the practical implementation of these techniques or even render them unsuitable in certain scenarios. The latter will summarize the key points and guide the reader towards Part II of the paper, where we will present a practical application of the methodology introduced in this study. This second part will demonstrate the applicability of our approach using real-world data and case studies, further supporting its robustness and relevance.

References

Ba, Shan, and V. Roshan Joseph. 2012. "Composite Gaussian Process Models for Emulating Expensive Functions." *The Annals of Applied Statistics* 6 (4): 1838–60. https://doi.org/10.1214/12-AOAS570.

Okoye, Kingsley, and Samira Hosseini. 2024. "Correlation Tests in R: Pearson Cor, Kendall's Tau, and Spearman's Rho." In *R Programming: Statistical Data Analysis in Research*, edited by Kingsley Okoye and Samira Hosseini, 247–77. Singapore: Springer Nature. https://doi.org/10.1007/978-981-97-3385-9_12.

Zhuang, Jirong, Deng Ding, Weiguo Lu, Xuan Wu, and Gangnan Yuan. 2025. "A Gaussian Process Based Method with Deep Kernel Learning for Pricing High-Dimensional American Options." *Computational Economics*, January. https://doi.org/10.1007/s10614-024-10833-9.