The Journal of Open Source Software

MultiHazard: Copula-based Joint Probability Analysis in R

Robert Jane [1¶], **Thomas Wahl** [1], **Francisco Pena** [2], **Jayantha Obeysekera** [3], **Callum Murphy-Barltrop** [4,5], **Javed Ali** [1], **Pravin Maduwantha** [1], **Huazhi Li** [6], **and Victor Malagon Santos** [7]

1 University of Central Florida, USA **2** South Florida Water Management District, USA **3** Florida International University, USA **4** TU Dresden, DE **5** ScaDS.AI, DE **6** Vrije Universiteit Amsterdam, NL **7** Royal Netherlands Institute for Sea Research, NL ¶ Corresponding author

DOI: 10.21105/joss.08350

Software

- Review
- Repository
- Archive

Editor: Oskar Laverny

Reviewers:

- @mingzehuang
- @EFAcar

Submitted: 21 February 2025
Published: 14 January 2026

Summary

Compound events occur when combinations of drivers and/or hazards contribute to a societal/environmental impact (Zscheischler et al., 2020). Even if none of the individual drivers or hazards are extreme, their combination can produce extreme impacts. Assessing the potential for compound extreme events is therefore critical for effective risk management and mitigation planning. To determine the probability of compound events, statistical models are applied to time series data of the drivers or hazards, typically as the first step in the risk-analysis modeling chain.

The `MultiHazard` R package is designed to enable practitioners to estimate the likelihood of compound events. Although the methods in the package are well-established in the scientific literature, they are not widely adopted by the engineering community despite guidelines increasingly mandating the estimation of compound event likelihoods. Functions within the package are designed to allow practitioners to apply their best judgement in making subjective choices. Inputs are time series representing the drivers/hazards; these may be historical observations or numerically generated with models (for the past or future). Outputs are:

- Estimates of the joint return periods for specific combinations of drivers/hazards.
- Isolines with uncertainty bounds containing drivers/hazards with a specified joint return period along with the "most-likely" or an ensemble of events sampled along an isoline.
- Synthetic sets of events where the peak magnitude of at least one driver/hazard is extreme.

Statement of need

The `MultiHazard` package was developed in collaboration with the South Florida Water Management District (SFWMD) to improve the level of service assessments for coastal infrastructure affected by both inland and coastal drivers (Jane et al., 2020). Initially, the package was created to implement the conditional sampling - copula theory approach. In two-sided conditional sampling, a driver is conditioned to be extreme and paired with the maximum value of the other driver within a specified lag-time. The process is repeated with the drivers reversed yielding two conditional samples. The best fitting of 40 copulas are tested to model the dependence between drivers in each conditional sample. The isoline corresponding to a user specified return period is given by the outer envelope when overlaying the (conditional) contours from the copula model fit to each sample (Bender et al., 2016), see Figure 1a,b. To obtain a single design event, the probability of events on the isolines is calculated using an empirical density estimate, selecting the event with the highest likelihood. Some experts recommend sampling an ensemble of events from the isoline to account for uncertainty in design event selection which is also possible in the package. The conditional sampling – copula theory approach has continued to gain traction (Kim et al., 2023; Maduwantha et al., 2024) and in a review of the best-available, actionable science was highlighted as an approach that Federal agencies in the United States may wish to develop detailed technical guidance on how to use it meet their needs (FFRMS, 2023).

The package possesses the functionaity to compute the desired isoline by interpolating joint return periods calculated over a user-defined grid which can result in smoother curves than overlaying the partial isolines. Isoline generation for time series comprising events from multiple (independent) populations e.g. different types of storms is faciliated by multiplying independent annual-non exceedance probabilities associated with each population as outlined in Maduwantha et al. (2024). Also coded in the package are the methods proposed in Murphy-Barltrop et al. (2023) and Murphy-Barltrop et al. (2024) to derive isolines using the HT04 (Heffernan & Tawn, 2004) and WT13 (Wadsworth & Tawn, 2013) models, along with a novel bootstrap procedure for evaluating isoline uncertainty (see Figure 1c,d), and a diagnotic tool proposed for evaluating the goodness of fit for a curve estimate. The HT04 and WT13 models allow for the analysis of compound events without assuming any copula form. Furthermore, the results from Murphy-Barltrop et al. (2023) guarantee that the estimated isolines from the HT04 and WT13 approaches satisfy certain theoretical properties.

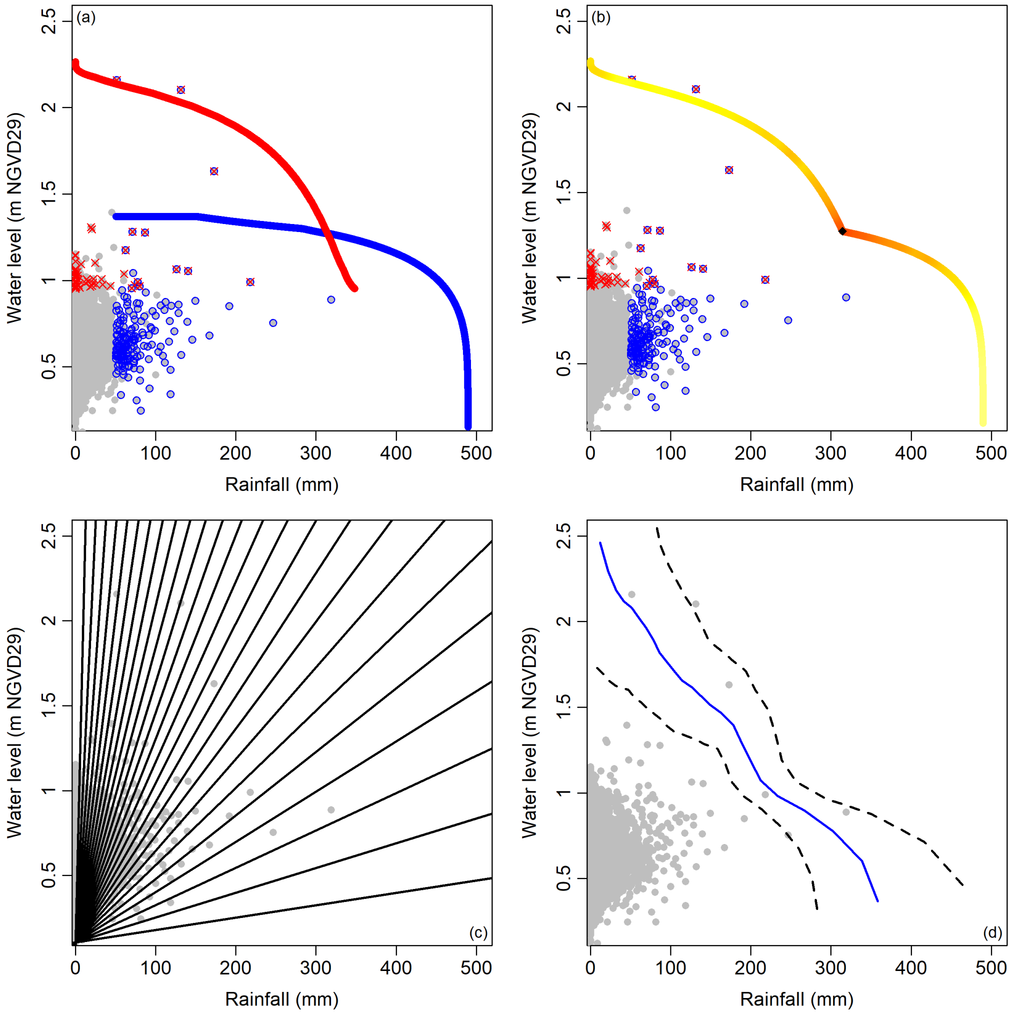

Figure 1: The 100-year isoline at case study site derived using two methods. Two-sided conditional sample - copula theory method: (a) Partial isolines from the samples conditioned on rainfall (red line) and water level (blue line) whose outer envelope when overlaid gives (b) the full isoline. Heffernan and Tawn model method: (c) Sample uncertainties are calculated along 100 rays emanating from the point comprising the minimum observed values of both drivers. (d) The median isoline (red line) and associated 90% confidence interval (dashed black lines) obtained via a block bootstrap procedure .

Compound events are of increasing concern for entities responsible for managing flood risk. `MultiHazard` is designed as a comprehensive user-friendly tool for copula-based joint probability analysis in R. As such, the package provides functions for pre-processing data including imputing missing values, detrending and declustering time series, as well as exploratory data analysis, e.g., analyzing pairwise correlations over a range of time-lags between the two drivers/hazards. The package also contains two automated threshold selection method for the Generalized Pareto distribution (Murphy et al., 2024; Solari et al., 2017) and approaches for robustly capturing the dependence structure when there are more than two relevant drivers/hazards,namely, standard (elliptic/Archimedean) copulas, Pair Copula Constructions (PCCs) and the HT04 model. For the analysis undertaken for the SFWMD, the higher dimensional approaches enabled groundwater level to be included in the analysis. More recently, an approach to generate time varying synthetic events, i.e., hyetographs and hydrographs, was added. Time varying conditions are a prerequisite for non-steady state hydrodynamic modeling.

Related packages

Several packages employ copula-based approaches to derive isolines. The MATLAB toolbox MvCAT (Multivariate Copula Analysis Toolbox) (Sadegh et al., 2017) utilizes up to 26 copula families to model the dependence structure between a pair of random variables. MvCAT includes fewer copula families than MultiHazard and does not consider the two-sided conditional sampling - copula theory methodology but rather uses a one-sided sampling approach. On the other hand, MvCAT adopts multiple criteria to select among the candidate copula families and a Bayesian framework to account for the uncertainty range for the copula parameters. Parts of the recently released ReturnCurves (André & Murphy-Barltrop, 2024) R package that implement the methods outlined in Murphy-Barltrop et al. (2023) and Murphy-Barltrop et al. (2024) have been subsummed into the MultiHazard package with permission from the authors. The ReturnCurves package only allows practitioners to use the WT13 model to evaluate isolines, while our package incorporates a range of different modelling frameworks.

Acknowledgements

The Authors thank the South Florida Water Management District for funding the development of this package through several projects over the past few years. Robert Jane and Thomas Wahl acknowledge financial support from the USACE Climate Preparedness and Resilience Community of Practice.

References

André, L., & Murphy-Barltrop, C. (2024). *ReturnCurves: Estimation of return curves*. https://doi.org/10.32614/cran.package.returncurves

Bender, J., Wahl, T., Müller, A., & Jensen, J. (2016). A multivariate design framework for river confluences. *Hydrological Sciences Journal*, *61*(3), 471–482. https://doi.org/10.1080/02626667.2015.1052816

FFRMS. (2023). *Federal flood risk management standard (FFRMS) climate-informed science approach (CISA) state of the science report*.

Heffernan, J. E., & Tawn, J. A. (2004). A conditional approach for multivariate extreme values (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *66*(3), 497–546. https://doi.org/10.1111/j.1467-9868.2004.02050.x

Jane, R., Cadavid, L., Obeysekera, J., & Wahl, T. (2020). Multivariate statistical modelling of the drivers of compound flood events in south florida. *Natural Hazards and Earth System Sciences*, *20*(10), 2681–2699. https://doi.org/10.5194/nhess-20-2681-2020

Kim, H., Villarini, G., Jane, R., Wahl, T., Misra, S., & Michalek, A. (2023). On the generation of high-resolution probabilistic design events capturing the joint occurrence of rainfall and storm surge in coastal basins. *Int. J. Climatol.*, *43*(2), 761–771. https://doi.org/10.1002/joc.7825

Maduwantha, P., Wahl, T., Santamaria-Aguilar, S., Jane, R. A., Booth, J. F., Kim, H., & Villarini, G. (2024). A multivariate statistical framework for mixed storm types in compound flood analysis. *Nat. Hazards Earth Syst. Sci.*, *24*, 4091–4107. https://doi.org/10.5194/nhess-24-4091-2024

Murphy, C., Tawn, J. A., & Z., V. (2024). Automated threshold selection and associated inference uncertainty for univariate extremes. *Technometrics*, *0*, 1–10. https://doi.org/10.1080/00401706.2024.2421744

Murphy-Barltrop, C. J. R., Wadsworth, J. L., & Eastoe, E. F. (2023). New estimation methods for extremal bivariate return curves. *Environmetrics*, *34*(5), e2797. https:

//doi.org/10.1002/env.2797

Murphy-Barltrop, C. J. R., Wadsworth, J. L., & Eastoe, E. F. (2024). Improving estimation for asymptotically independent bivariate extremes via global estimators for the angular dependence function. *Extremes*, *27*(4), 643–671. https://doi.org/10.1007/s10687-024-00490-4

Sadegh, M., Ragno, E., & AghaKouchak, A. (2017). Multivariate copula analysis toolbox (MvCAT): Describing dependence and underlying uncertainty using a bayesian framework. *Water Resources Research*, *53*(6), 5166–5183. https://doi.org/10.1002/2016WR020242

Solari, S., Egüen, M., Polo, M. J., & Losada, M. A. (2017). Peaks over threshold (POT): A methodology for automatic threshold estimation using goodness of fit p-value. *Water Resources Research*, *53*(4), 2833–2849. https://doi.org/10.1002/2016WR019426

Wadsworth, J. L., & Tawn, J. A. (2013). A new representation for multivariate tail probabilities. *Bernoulli*, *19*(5B), 2689–2714. https://doi.org/10.3150/12-BEJ471

Zscheischler, J., Martius, O., Westra, S., Bevacqua, E., Raymond, C., Horton, R. M., Hurk, B. van den, AghaKouchak, A., Jézéquel, A., Mahecha, M. D., Maraun, D., Ramos, A. M., Ridder, N. N., Thiery, W., & Vignotto, E. (2020). A typology of compound weather and climate events. *Nat Rev Earth Environ*, *1*, 333–347. https://doi.org/10.1038/s43017-020-0060-z