A simple and useful regression model for bimodal extreme data

This manuscript develops an interesting regression models for bimodal extreme-valued data. The main developments are based on the BGEV distribution as introduced in Otiniano et al. (2023). Although I am generally positive about the value of this work, I have a few major concerns:

- Originality and main contribution: To my knownledge this manuscript pioneers the development of regression models for bimodal extreme-valued data. Is that the case? If yes, this should be further stressed. Otherwise, this needs to be clarified further.

- Presentation and style: While I believe there is a variety of interesting ideas in this manuscript, as it stands it not engaging nor written in line with the expectations that one would have for a leading journal. For example, paragraphs such as the one below add very little to the readership of *TEST*:

  > "On the other hand, climate variables such as temperature, relative humidity, and precipitation, among others, depend on other variables such as radiation, pressure, and wind speed (Schlatter 1987). In statistics, a regression model can show whether changes observed in the dependent variable (response) are associated with changes in one or more independent variables (covariates)."

- *Background and motivation*, I: At the moment the paper is quite poorly motivated. I believe the introduction should be revamped placing more focus on two central ideas: bimodality in the extremes, and regression methods for extremes. In terms of regression methods for extremes, I believe it would be natural to refer to a few key references on the topic. Chapter 6 of the Handbook on Statistics of Extremes (de Carvalho et al., 2026, to appear) presents additional material on regression methods for univariate extremes, which could be noted in §1. See also the references therein.

- *Background and motivation*, II: The recent dissertation of Carcaiso (Carcaiso, 2025, §5) contains related developments on the subject of heterogenous extremes, and should be referred to in §1.

- *Background and motivation*, III: The focus of the paper is on extremes, as made clear by the title and abstract; accordingly, the Introduction should place greater emphasis on relevant literature in extreme value theory, such as the references mentioned above. In addition to this, it would be appropriate to refer to further books on the topic (e.g., Coles, 2001; Beirlant et al., 2004). For example, in addition to Chapter 6 of the Handbook, Coles (2001) covers a variety of regression models models for extremes. See also de Carvalho et al. (2022) and Majumder and Richards (2025) for two examples of recent developments on this direction.

  Finally, to keep the focus on extremes, I would only refer *en passant* to the literature on regression models for bimodal data (say, 1–2 sentences), and would refer to less references on that topic as well.

- Background and conclusion: In my opinion at the moment, the paper puts too much emphasis on the reparametrization. I would place focus first on the regression, and then say that motivated by it, you develop a more interpretable reparametrization of BGEV.

- §§2–4 need more clarity in terms of whether or not this is a model for a block maxima response.

- §2: Why does the reparametrization in terms of the median is actually intuitive—since the interest in an extreme value analysis is typically on higher quantiles?

- §2: Why to restrict the focus to polynomial basis? Perhaps it could be mentioned that splines could be used as well.

- §2: Does this regression model suffer from the well-known drawback of the GEV regression, in the sense that it may be impossible to evaluate on some new values of the covariate? See Castro-Camilo et al. (2022, pp. 2–3).

- §4: All the text that starts in §4 should have a subsection (e.g., §4.1 Data description and rationale for the analysis; then, §4.2 Adjusting the dew...).
- §4: To produce more modern visualization of the quantile residuals, please replace the probability plots with qq-boxplots (Rodu and Kafadar, 2022).
- §5: Future work could be noted here. For example, it would be interesting to develop a LASSO-version of the proposed model, along the same lines as de Carvalho et al. (2022).

Minor Comments

- I do not think the current title works well. Something like "A Regression-Type Model for Bimodal Extreme-Valued Data" would be more appropriate.
- As noted in (**??**), some covariates may be signifcant for the bulk but not for the extremes. Is it possible to assess this with the proposed model?
- Avoid acronyms for words that are barely used (e.g., replace CDF with distribution function, PDF with...).
- Remove spaces in between em-dashes (i.e., replace ' — ' with '—').
- Comma missing in (3).

Despite the substantial work this revision may require, there is much to appreciate in the manuscript, and I look forward to receiving the revised version.

References

Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J. (2004), *Statistics of Extremes: Theory and Applications*, Wiley, Hoboken, NJ: Wiley.

Carcaiso, V. (2025), "Bayesian Mixture Models for Extremes," Ph.D. thesis, University of Padova, PhD dissertation.

Castro-Camilo, D., Huser, R., and Rue, H. (2022), "Practical strategies for generalized extreme value-based regression models for extremes," *Environmetrics*, 33, e2742.

Coles, S. (2001), *An Introduction to Statistical Modeling of Extreme Values*, London: Springer.

de Carvalho, M., Huser, R., Naveau, P., and Reich, B. J. (2026, to appear), *Handbook on Statistics of Extremes*, Boca Raton, FL: Chapman & Hall/CRC.

de Carvalho, M., Pereira, S., Pereira, P., and de Zea Bermudez, P. (2022), "An Extreme Value Bayesian Lasso for the Conditional Left and Right Tails," *Journal of Agricultural, Biological and Environmental Statistics*, 27, 222–239.

de Carvalho, M., Ramirez, V. P., Henriques-Rodrigues, L., and Lee, J. (2026, to appear), "Regression Models for Extreme Events," in *Handbook on Statistics of Extremes*, ed. M. de Carvalho, R. Huser, P. N. . B. R., Chapman & Hall/CRC.

Majumder, R. and Richards, J. (2025), "Semi-parametric bulk and tail regression using spline-based neural networks," *arXiv:2504.19994*.

Otiniano, C. E., Paiva, B. S., Vila, R., and Bourguignon, M. (2023), "A bimodal model for extremes data," *Environmental and Ecological Statistics*, 30, 261–288.

Rodu, J. and Kafadar, K. (2022), "The q–q Boxplot," *Journal of Computational and Graphical Statistics*, 31, 26–39.