

Modelling metocean extremes using GAMs - a comparison study

P. Bohlinger^a, P. Jonathan^b, B. Youngman^c, T. Economou^c, Ø. Breivik^{a,d}

Contact: patrikb@met.no

Motivation

Marine standards recommend approaches to calculate return levels and exceedance probabilities for extreme values of marine key variables (such as wind, wave, surge). However, some approaches date back many decades. Examples include: the initial distribution method using a Weibull distribution for marginal models on all data; modelling directional wave heights using independent models for each sector; and the Weibull log-normal model for joint probability [1,2]. While they have a practical appeal, these approaches do not take into account recent developments in extreme value theory (EVT). When extrapolating far into the tail in particular, empirical models should be informed by EVT.

However, despite multiple studies it is still not obvious how these modern approaches perform compared to existing standards in various settings. Therefore we contribute with a comparison of our in-house developed workflow for non-stationary, joint **Environmental Extremes EnvEx** applying modern EVT methodology [3,4]. Here, we focus on the fair comparison of the sea state parameter Hs and joint variables because Hmax and Cmax are downstream response variables with their quality being contingent on the correct estimation of the sea state parameters and the distribution functions for the individual waves and crests.

Conclusion

EnvEx provides a framework for applying non-stationary extreme value analysis including UQ for our metocean variables of interest. The impact of variables on return variables and exceedance probabilities can be investigated as all dependencies are traceable through the model.

EnvEx is able to predict a meaningful distribution of 100-year maxima. Models that do not take into account autocorrelation, i.e. serial correlation and covariate effects, are mis-specified and develop biases that are most severe for long tailed data.

Depending on the purpose it may be beneficial to consider the full predictive distribution of 100-year maxima rather than the return levels to better account for the variability in the underlying stochastic process.

Plans

- Incorporate subsampling based uncertainty
- Incorporate more models for Hmax and Cmax
- Create publicly available software package
- Benchmark using different datasets and
- Investigate more sensitivity

Acknowledgements

FORESEE (Nr. 334188) is funded by the *Researcher Project for Young Talents (FRIPRO)* granted by the Norwegian Research Council.

References

- [1] E.M. Bitner-Gregersen and S. Haver (1991). Joint Environmental Model for Reliability Calculations. Proceedings of the 1st International Offshore and Polar Engineering Conference, Edinburgh, United Kingdom, August 11-15, 1991.
- [2] DNV-RP C205 (2014), Environmental conditions and environmental loads.
- [3] Jonathan and Ewans (2013). Statistical modelling of extreme ocean environments for marine design: A review. *Ocean Engineering*, 62:91–109, 2013. ISSN 0029-8018. doi: 10.1016/j.oceaneng.2013.01.004.
- [4] M. Speers, D. Randell, J. Tawn, and P. Jonathan. Estimating metocean environments associated with extreme structural response to demonstrate the dangers of environmental contour methods. *Ocean Engineering*, 311:118754, 2024. ISSN 0029-8018. doi: <https://doi.org/10.1016/j.oceaneng.2024.118754>.
- [5] Heffernan and Tawn (2004). A Conditional Approach for Multivariate Extreme Values. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 66(3):497–546, 07. doi: 10.1111/j.1467-9868.2004.02050.x.
- [6] Youngman, B. D. (2022). *evgam*: An R package for Generalized Additive Extreme Value Models. *Journal of Statistical Software*. doi: 10.18637/jss.v103.i03.
- [7] Youngman and Economou (2017). Generalised additive point process models for natural hazard occurrence. *Environmetrics*, 28. doi: <https://doi.org/10.1002/env.2444>. e2444 env.2444.
- [8] Breivik et al. (2022). The impact of a reduced high-wind Charnock parameter on wave growth with application to the North Sea, the Norwegian Sea, and the Arctic Ocean. *Journal of Geophysical Research: Oceans*, 127, e2021JC018196. doi: 10.1029/2021JC018196

Non-stationary and joint modelling strategy in *EnvEx*

We assume that 1) storm events and their storm peak values are independent given storm peak covariates, 2) storm events can be sufficiently characterized by their respective storm peak variables and the storm peak covariates. Additionally, the storm evolution itself can be described by the probability density of a multivariate time series of sea state characteristics throughout the storm event, other sea state covariates, and the unknown and variable storm length, 3) the dependence between individual waves given their sea state is negligible.

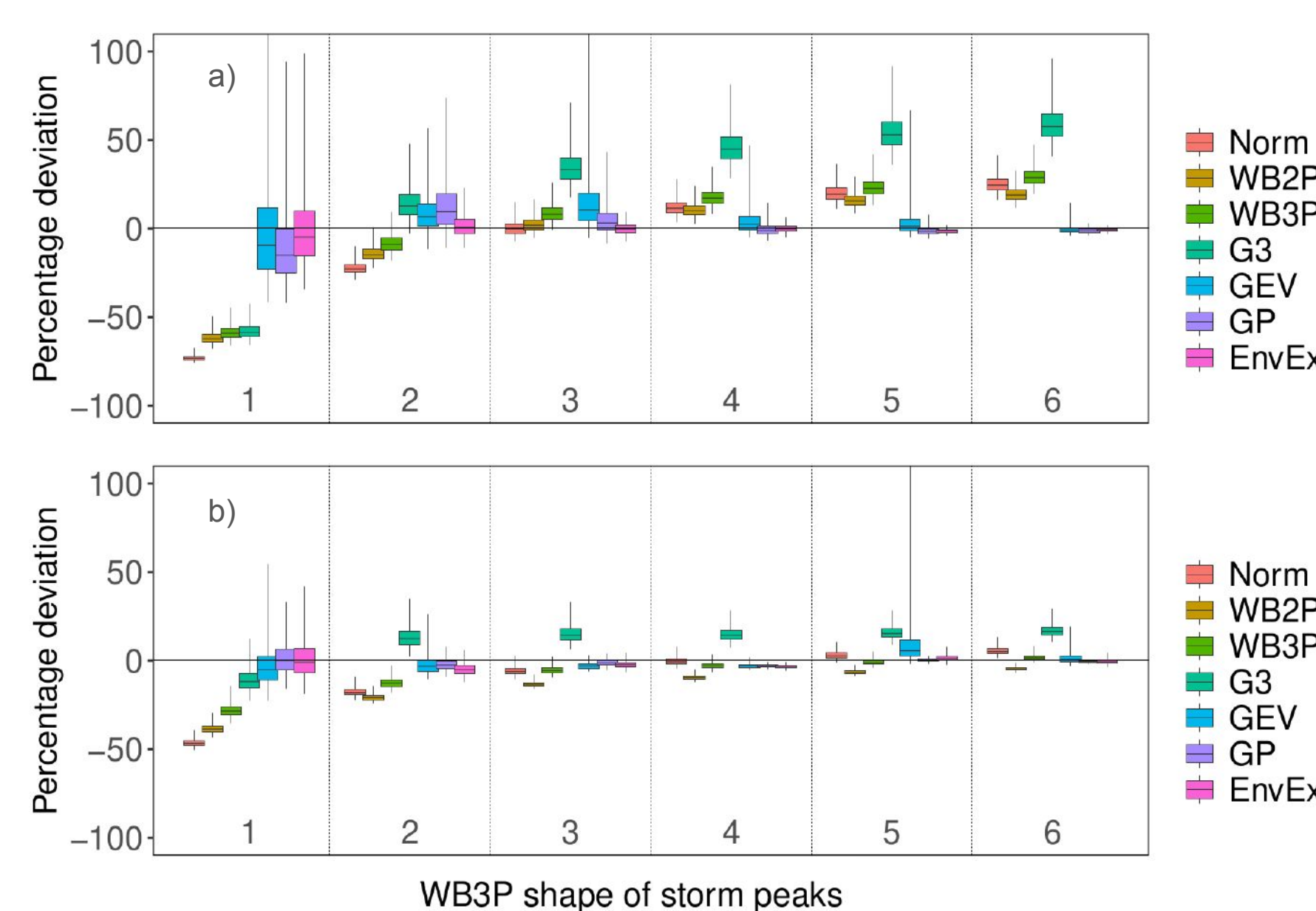
$$\begin{aligned} F_{H_{max},s}(h) &= \int_{\theta^{SP}} \int_{\mathbf{x}^{SP}} \int_{\{(\mathbf{x}_s, \theta_s)\}, \tau \in S_{\tau}} \\ A) \quad & \prod_{s \in S_{\tau}} F_{H_{max},L}(\mathbf{x}_s, \theta_s)(h | \mathbf{x}_s, \theta_s) \\ B) \quad & \times f_{\{(\mathbf{x}_s, \theta_s)\}, T \in S_{\tau} | \mathbf{x}^{SP}, \theta^{SP}}(\{(\mathbf{x}_s, \theta_s)\}_{s \in S_{\tau}}, \tau | \mathbf{x}^{SP}, \theta^{SP}) \\ C) \quad & \times f_{\mathbf{x}^{SP} | \theta^{SP}}(\mathbf{x}^{SP} | \theta^{SP}) \\ D) \quad & \times f_{\theta^{SP}}(\theta^{SP}) \\ & d\{(\mathbf{x}_s, \theta_s)\}_{s \in S_{\tau}, \tau} d\mathbf{x}^{SP} d\theta^{SP} \end{aligned}$$

Following [3,4] we can now design the hierarchical model for the distribution of the maximum individual wave height within any storm. The hierarchy consists of A) a model for Hmax within a storm event with random length $\tau \in T$ consisting of sea states of

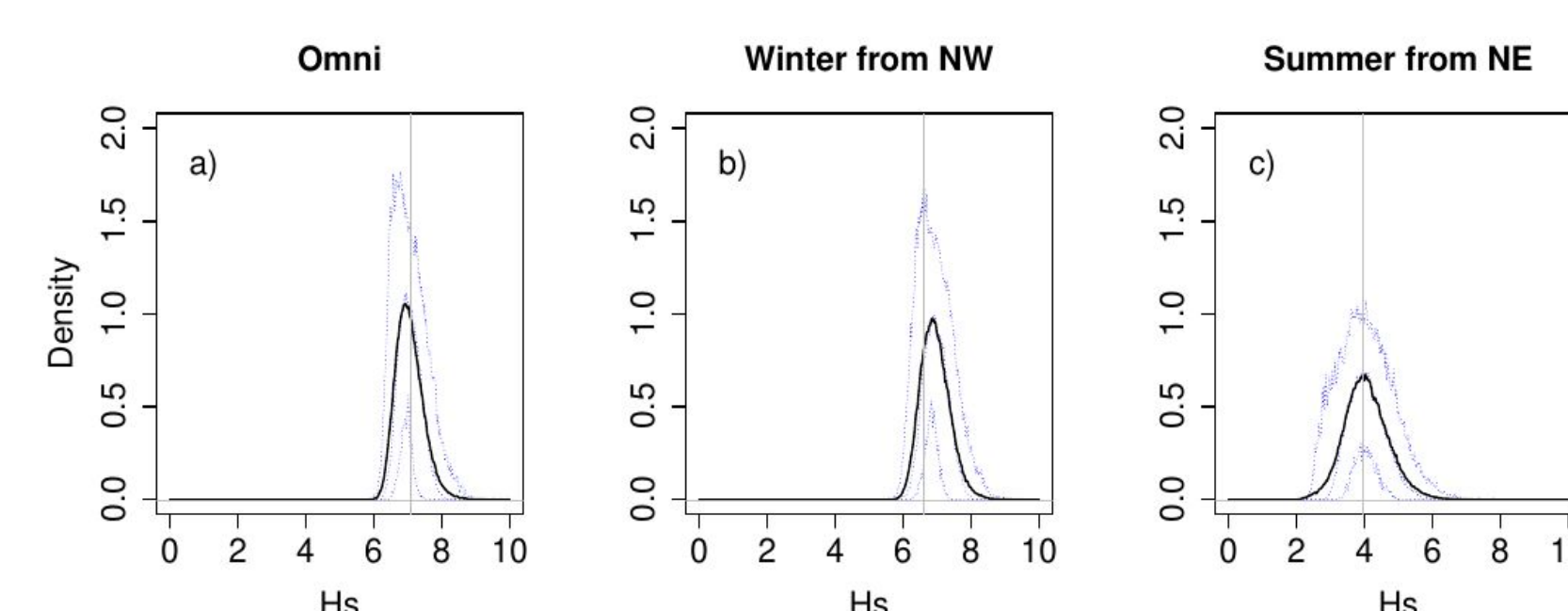
unknown length L, B) the evolution of all multivariate storm events characterized by all multivariate storm peaks given their covariates, C) the joint model (Heffernan and Tawn) [5] or non-stationary marginal modelling of the independent storm peak variables given their covariates [6], and D) the occurrence of storm peak covariates [7].

Comparison using artificial data (the truth is known)

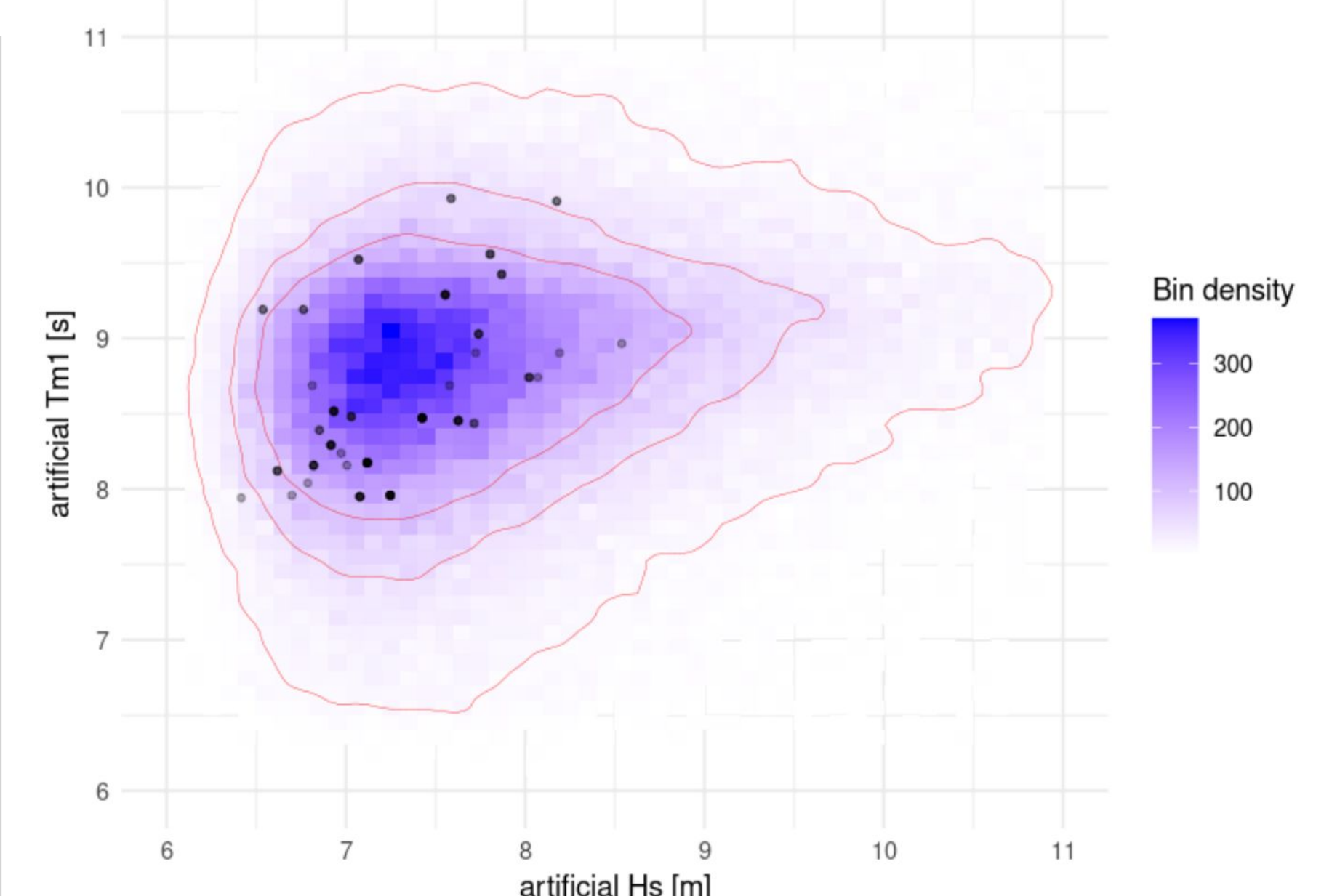
We fit a range of candidate models as commonly applied and let them predict a distribution of 100-year maxima. Those should group around the true 100-year return value for the storm peaks for the probability: $\mathbb{P}(H_{sT}^{SP} > u_T^{SP})$. All models are provided with the same 50 years of data to which they are fitted to. The predictions based on these models are subsequently compared.



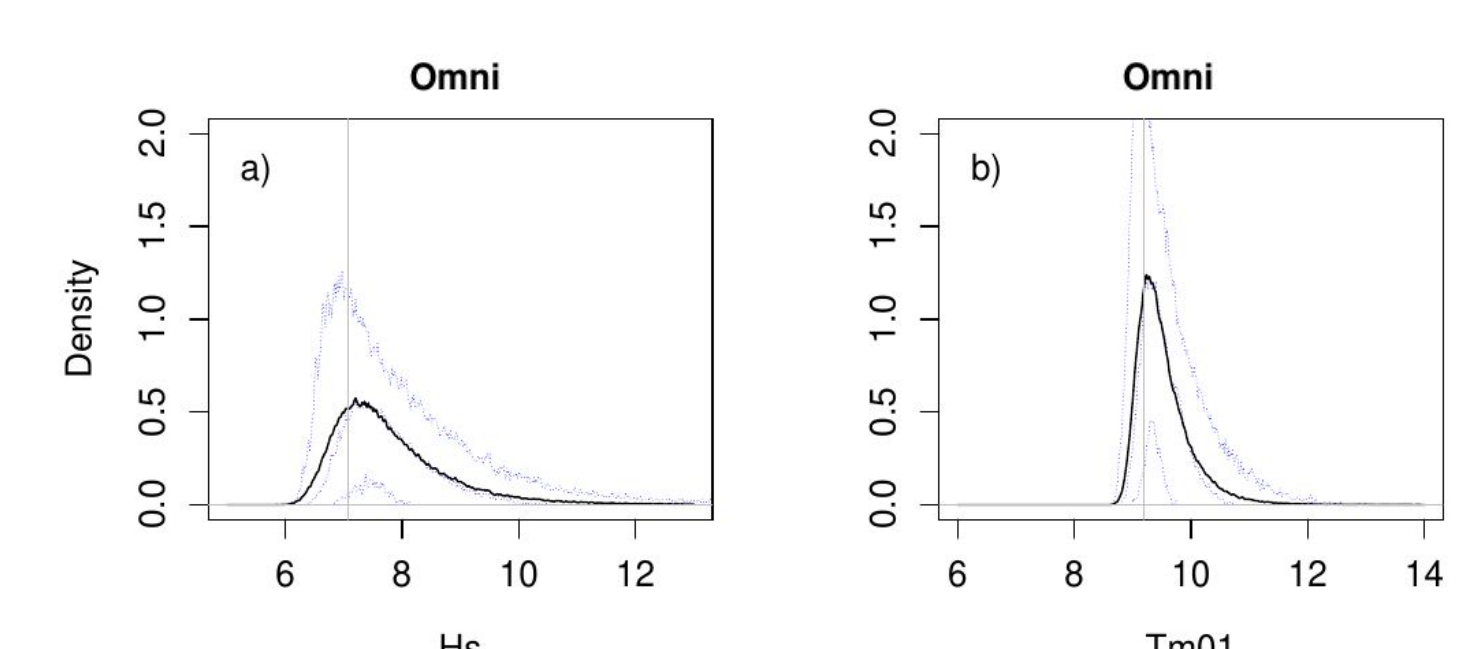
Boxplots of the percentage deviations from the true 100-year return value for all candidate methods evaluated for different WB3P storm peak shapes, a) seasonal dependence, b) serial correlation. The boxes represent the interquartile range (0.25pct - 0.75pct) and the whiskers include 99% (0.005pct - 0.995pct). *EnvEx* is seasonal in all threshold, point-process, and GP-parameters.



EnvEx predictions for 100-year Hs maxima from a seasonal-directional marginal model for Hs against the true value represented by the gray line. Seasonal-directional subregions are b) winter and direction from 100-150 degrees and c) summer and directions from 200-250 degrees. The blue lines indicate the 5pct, 50pct, and the 95pct. The black line represents the combined density from all bootstrap members.



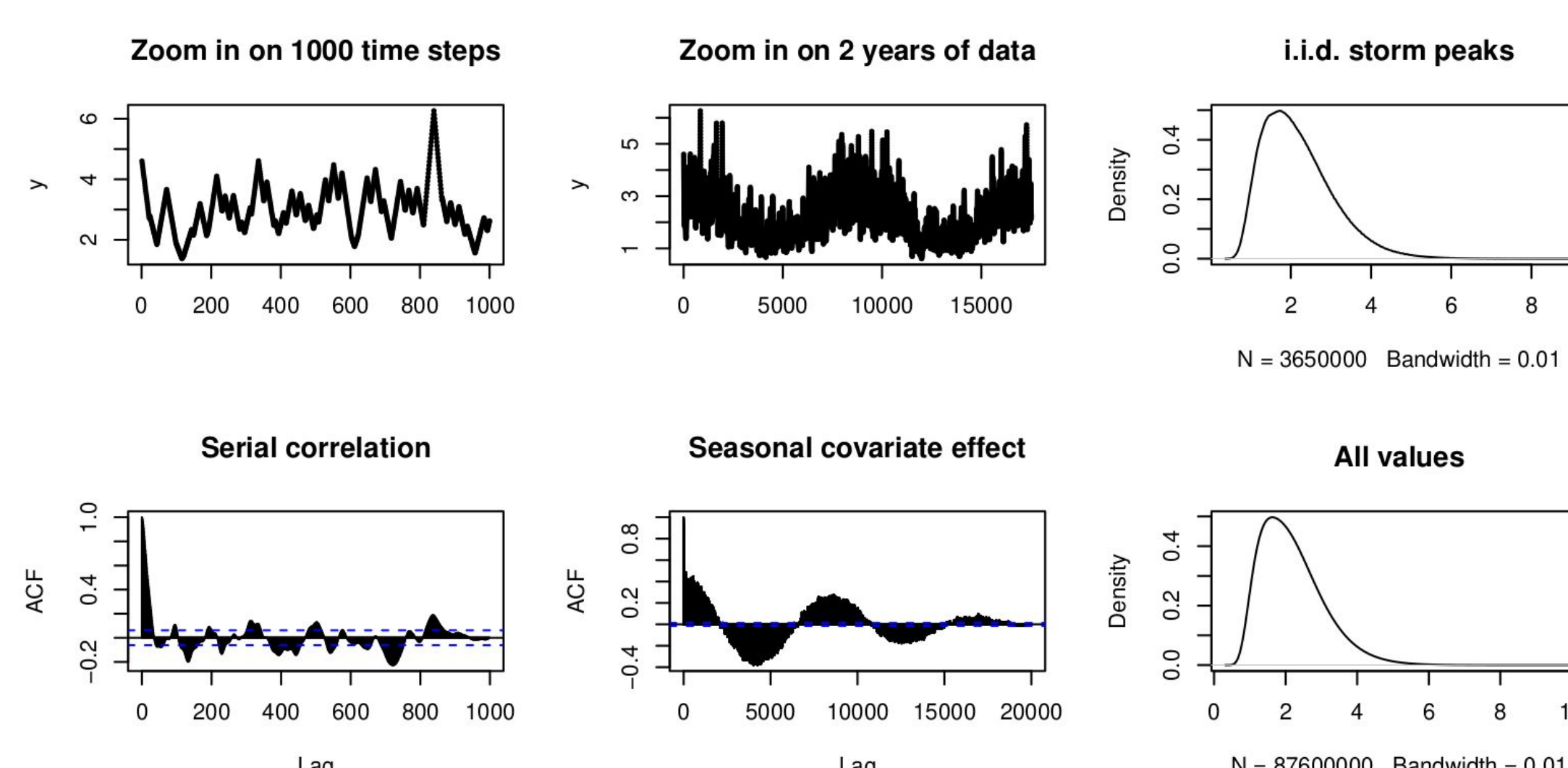
EnvEx predictions for the joint 100-year maxima of Hs and Tm01 from the non-stationary marginal models combined with the Heffernan and Tawn model [5]. Isodensity lines for quantiles $q=0.01, 0.05, 0.1$ are displayed with red lines. True 100-year joint maxima from the artificial dataset are represented by black dots. *EnvEx* is seasonal in all parameters for the threshold, point-process, and GP models for both response variables Hs and Tm01.



Same joint model as above. Predictions of 100-year maxima for both response variables displayed as marginal distributions and shown against the true value represented by the gray line.

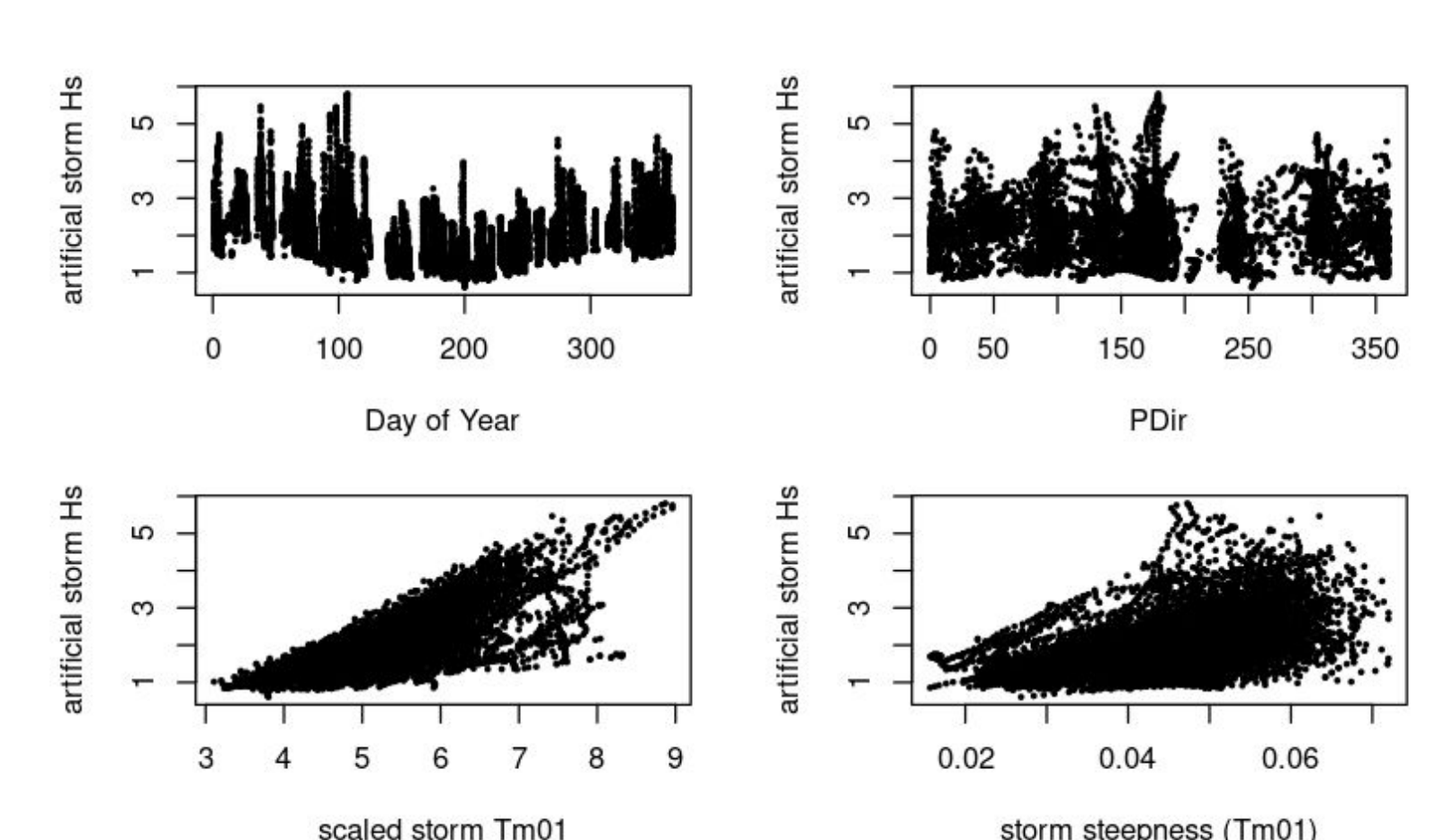
Data

Characteristics of exemplary seasonal artificial data used in lieu of Hs.



WB3P peaks with shape=2 and introduced serial correlation following the power-storm model with $\lambda=1$ and seasonality in all parameters. Consisting of 10000 years worth of hourly data.

Characteristics of hypothetical Ekofisk



WB3P peaks with shape=2 and introduced Ekofisk-like serial correlation and seasonality in all parameters as well as Ekofisk-like dependency on Tm01. Consisting of 10000 years worth of hourly storm data. Dependencies and serial correlation are based on NORA3 at the Ekofisk location [8].