A mixture of expert model for sea state characterization

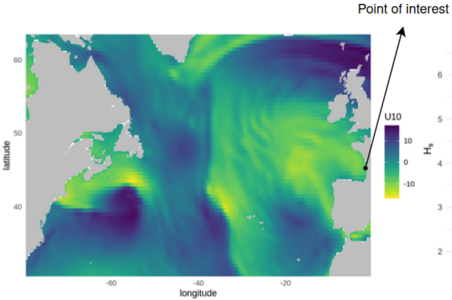
Said Obakrim, Pierre Ailliot, Valérie Monbet, and Nicolas Raillard

Data Science pour les risques côtiers

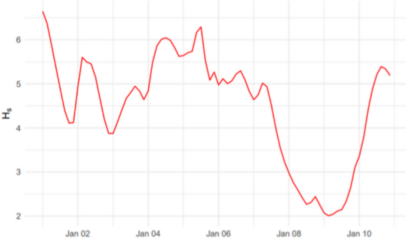
Monday 20th November, 2023

Objective

- Predict/characterize H_s at a nearshore location in the Bay of Biscay using North Atlantic wind conditions



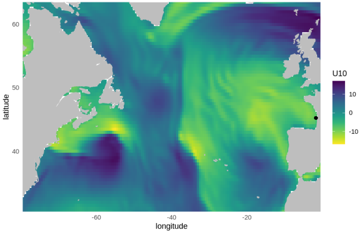
West-east wind components on January 14, 1994



H_s at January 1994 at the point of interest

Wind Data

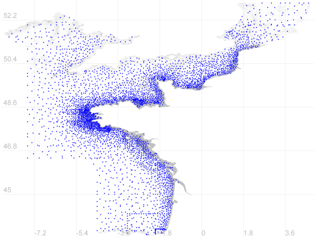
- CFSR[a]: global reanalysis, developed at NCEP
- 0.5°×0.5° spatial resolution
- Temporal resolution of 3 hours from 1994 to 2016



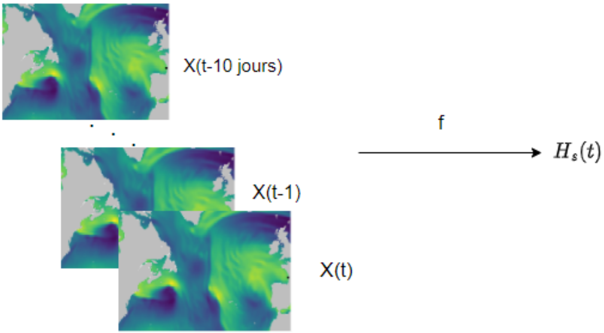
[a] Saha et al. 2013[1]

Wave Data

- Homere[a]: sea state data reanalysis based on the WAVEWATCH III model
- Temporal resolution of 3 hours from 1994 to 2016



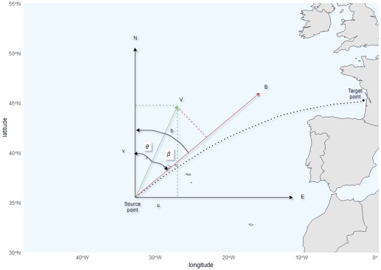
[a] Boudière et al. 2013 [2]



- To predict H_s in our target location, it is necessary to consider at least 10 days of past wind conditions.
- In the data considered here, there are 8102 covariates of wind to be taken into account.
- It will be computationally costly to directly use the data as it is in a statistical model.

Wind Projection

Notations:
Hereafter, we denote x as the position in space and t in time



$$W(x,t) = U(x,t)\, cos^2\left(\frac{1}{2}(b(x) - \theta(x,t))\right)$$

- $W(x,t)$: projected wind, $U(x,t)$: wind speed, $\theta(x,t)$: wind direction, and $b(x,t)$: great circle direction

Temporal Smoothing

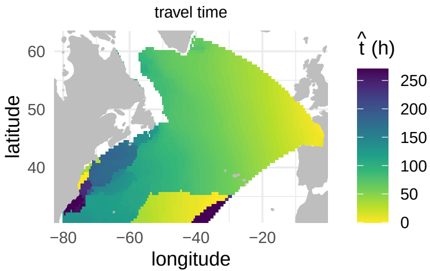
- Covariates X are defined as follows:

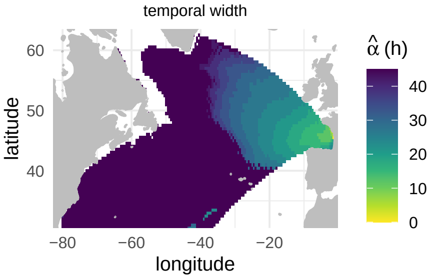
$$X(x,t) = \frac{1}{2\alpha(x)+1} \sum_{i=t-t(x)-\alpha(x)}^{t-t(x)+\alpha(x)} W(x,i)^2$$

$t(x)$: average wave travel time at point x
$\alpha(x)$: temporal window length at point x

- Estimation:

$$(\hat{t}(x), \hat{\alpha}(x)) = \arg \max_{t(x),\alpha(x)} \big(corr(H_s, X(x,t)) \big)$$



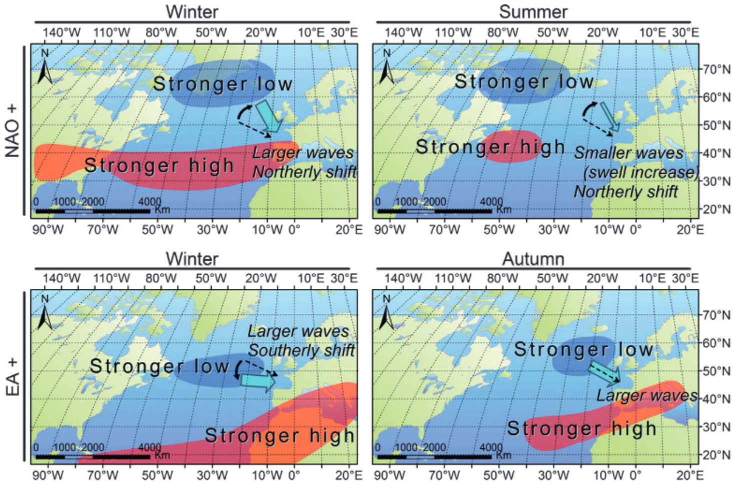


- Both parameters are spatially smooth and increase with the distance between the source and the target point

Remark:

The covariates defined here will be used as predictors for our model and will be noted as X

What we know about waves in the Bay of Biscay ?



Source: Charles et al. (2012). Present wave climate in the Bay of Biscay: spatiotemporal variability and trends from 1958 to 2001. Journal of Climate, 25(6), 2020-2039.

What we know about waves in the Bay of Biscay ?

- Waves at this area are related to large-scale circulation patterns such as North Atlantic Oscillation (NAO)
- These large-scale circulation patterns are known under the names "weather regimes" or "weather types"
- Traditionally, these patterns are found by using empirical orthogonal functions (EOFs) or K-means

▲ Improve the predictions of H_s

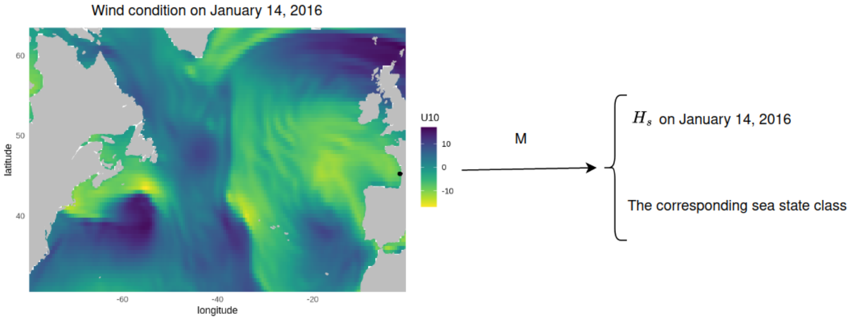
▼ Does not take into account local-scale conditions

▼ Weather regimes are not evaluated on the prediction of H_s

Combine local + global patterns for sea state classification

What we aim to achieve

- A model capable of predicting H_s at time t and the corresponding sea state class



Penalized Mixture of Experts

$$Z_i \sim M(1, p_i), \quad p_i = (p_{i1}, ..., p_{iK})^T, \quad i = 1, ..., n$$

$$p_{ik} = \frac{\exp(X_i \gamma_k)}{\sum_{l=1}^{K} \exp(X_i \gamma_l)}$$

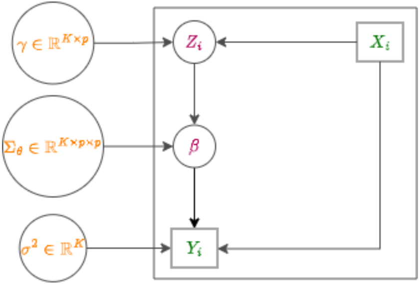
$$\beta | Z_i = k \sim \mathcal{N}(0, \Sigma_{\theta k}), \quad k = 1, .., K$$

$$Y_i | \beta, Z_i = k \sim \mathcal{N}(X_i \beta_k, {\sigma_k}^2)$$

Purple: hidden variables

Green: observed variables

Orange: parameters



Penalized Mixture of Experts

$$Z_i \sim M(1, p_i), \;\; p_i = (p_{i1}, ..., p_{iK})^T, \;\; i = 1, ..., n$$

$$p_{ik} = \frac{\exp(X_i \gamma_k)}{\sum_{l=1}^{K} \exp(X_i \gamma_l)}$$

$$\beta | Z_i = k \sim \mathcal{N}(0, \Sigma_{\theta\, k}), \;\; k = 1, .., K$$

$$Y_i | \beta, Z_i = k \sim \mathcal{N}(X_i \beta_k, {\sigma_k}^2)$$

Motivation and advantages of this model:

- The weather types, through Z, permits to treat heterogeneity of data
- We can put any covariance structure on the regression coefficients through $\Sigma_{\theta\, k}$
- This model permits to penalize the coefficients without the need to use further cross-validation techniques

Estimation

- Models with hidden variables are often estimated using the expectation maximization (EM) algorithm
- Here we use a variational EM algorithm, given the E step is intractable
- For the estimation details: Obakrim, Said. (2022). Statistical downscaling and climate change in the coastal zone. Université Rennes 1. `https://theses.hal.science/tel-03952800/`
- We use the data from 1994 to 2013 as a training set and 2014 to 2016 as a test set
- We found that the optimal number of classes is 3

Results

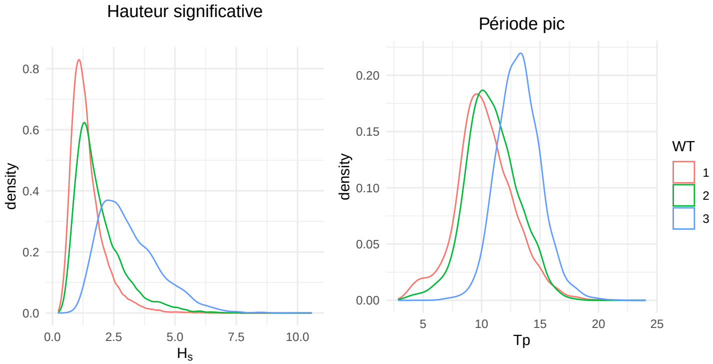


Figure: Empirical density of H_s and Tp based on the obtained classes.

- The obtained classes depend on H_s and Tp

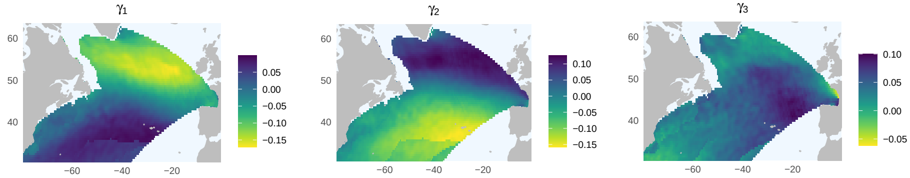


Figure: Estimated coefficients of the Multinomial model

- The probability of being in a specific weather type is related to the origin of the waves
- e.g. when the southern wind is strong, it is very likely that we are in the first weather type

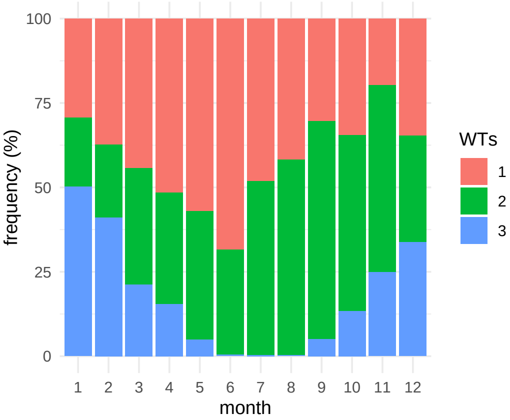


Figure: Monthly frequency of weather types.

- The 3rd weather type mainly occurs in winter
- The 1st and 2nd often occur in summer

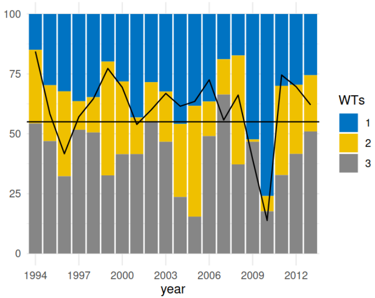
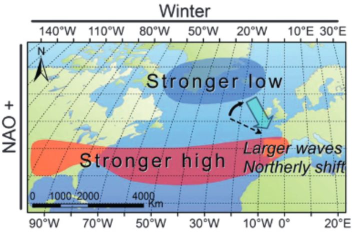
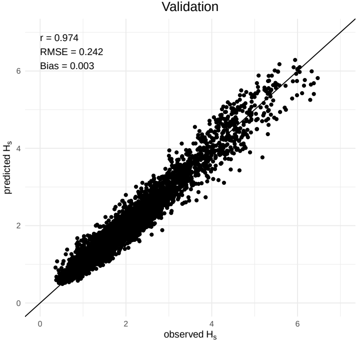


Figure: Winter long-term variability of weather types as a function of NAO

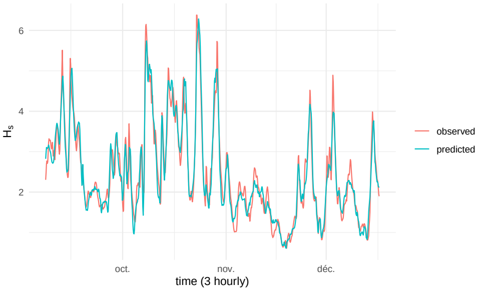
- Strong waves are mostly observed during NAO+



Observed vs predicted H_s in the test set



Time series of observed and predicted values at the end of 2014



Summary

- We proposed a model that does regression and classification at the same time
- The model demonstrates satisfactory prediction accuracy
- The resulting sea state classification (weather types) is interpretable

Considering both local and global-scale conditions when developing weather types is beneficial for prediction accuracy and interpretability

References I

S. Saha, S. Moorthi, H.-L. Pan, X. Wu, J. Wang, S. Nadiga, P. Tripp, R. Kistler, J. Woollen, D. Behringer, *et al.*, "The ncep climate forecast system reanalysis," *Bulletin of the American Meteorological Society*, vol. 91, no. 8, pp. 1015–1058, 2010.

E. Boudière, C. Maisondieu, F. Ardhuin, M. Accensi, L. Pineau-Guillou, and J. Lepesqueur, "A suitable metocean hindcast database for the design of marine energy converters," *International Journal of Marine Energy*, vol. 3, pp. e40–e52, 2013.