1

Analysis of maritime conditions via nonparametric directional methods

María Alonso-Pena

Understanding maritime conditions, particularly wave direction and height, is vital for informed decision-making in activities such as shipping, navigation, and seafood gathering. This study applies nonparametric directional methods to analyze these critical variables, employing methods such as kernel density estimation (KDE) and flexible regression models to uncover complex relationships. By addressing the circular nature of the data and the cylindrical relationships between wave direction, wave height, and covariates such as wind direction and speed, the analysis provides a solid framework for understanding these interactions.

1.1 Introduction

A key variable in understanding maritime conditions is wave direction, which is inherently circular in nature (see, e.g., [28, 23, 16, 27]), as is wind direction, another fundamental factor influencing sea behavior. Due to their circular support (defined on the unit circle rather than the real line) it is crucial to apply appropriate statistical methods that respect the geometry of these variables.

Beyond individual variables, it is important to consider interactions between wave direction and wave height, since the origin and strength of waves are often jointly influenced by meteorological forces. Likewise, wind direction and wind speed interact in ways that impact not only sea state but also navigational safety, coastal erosion, and energy extraction. These pairings, direction with magnitude, constitute examples of cylindrical variables, where one component is circular and the other is linear (see, e.g., [29, 1]).

Although models for cylindrical data do exist, many impose assumptions that are overly restrictive, limiting their capacity to capture complex real-world behaviors. In the context of real-valued variables, data-driven methods for nonparametric density estimation have been well developed ([40, 37, 42]), and several works have extended these ideas to circular ([22, 6, 15], toroidal ([10]), and cylindrical domains ([18]). Moreover, [19] proposed a goodness-of-fit test for parametric families of directional-linear densities.

In addition to examining joint distributions, it is also crucial to model how wave direction varies in response to wind behavior. Several parametric regression models have been proposed for circular responses with real-valued covariates ([21, 24, 17, 34]) and with circular covariates ([13, 26]). However, in practice, the regression functions implied by these models are often too simplistic to capture the true underlying structure of the data.

To address this, data-driven regression models based on kernel smoothing have been extensively studied in the real-valued setting ([14, 7]), and adapted to the case of circular responses by, among others, [11]. Moreover, such estimators can be used to develop significance tests and group comparisons ([2]) as well as goodness-of-fit tests ([**?**]). More advanced regression models have also been developed to go beyond the conditional mean, such as quantile regression for circular responses ([12]) and modal regression ([3]), offering a more nuanced understanding of directional data under covariate effects.

In this work, we study the behavior of wave direction at a key location on the Spanish coast, focusing on its interaction with both wave height and wind conditions. We adopt a nonparametric, data-driven approach that allows for flexible modeling of circular and cylindrical variables, capturing complex dependencies that parametric models may miss. The paper begins with a description of the dataset, which includes maritime conditions (wave direction and height) and meteorological variables (wind direction and speed) recorded at a strategic point along the Spanish coast. Next, we introduce the methodology used to analyze wave behavior, including kernel-based density estimation techniques for circular and cylindrical variables, regression models for circular responses, nonparametric significance tests, and modal regression. This is followed by a detailed analysis of wave dynamics on the northwestern Spanish coast, with interpretations of the observed patterns. We conclude with final remarks and suggestions for future research.

1.1.1 Wave data in Galicia, Spain

We analyze maritime and meteorological data collected from a key coastal location in northwestern Spain. The variables of interest include wave origin direction, wave height, wind origin direction, and wind speed. These data were obtained from the Spanish Ministry of Transport and Sustainable Mobility (https://www.puertos.es/es-es/oceanografia/paginas/portus.aspx) and are part of their ongoing oceanographic monitoring efforts.

The measurements were taken at the Villano-Sisargas Buoy, positioned at latitude 43°29.4’ N and longitude 9°12.6’ W, just off the coast of the Costa da Morte, literally *Coast of Death*, in the region of Galicia. This area is well known for its dangerous maritime conditions, where frequent storms meet a rugged, rocky shoreline. The buoy plays a critical role in monitoring this high-risk zone, contributing to maritime safety and environmental awareness. Beyond its environmental significance, this region has strong economic ties to the sea. Local livelihoods depend heavily on marine activities such as fishing and shellfish harvesting, both of which are highly sensitive to changes in oceanographic and meteorological conditions.

Observations were recorded every 12 hours over a one-year period, yielding a total of 712 measurements. Wave and wind directions are expressed in degrees, with 0°indicating North and 90°indicating East.

1.2 Statistical methodology

To analyze the relationships between maritime variables such as wave direction, wave height, and wind characteristics, we use statistical tools that allow for flexible exploration of the data without imposing rigid assumptions about its structure. In particular, because some of these variables, such as wave and wind direction, are directional in nature, we require methods that can properly account for their circular properties. This section introduces the nonparametric techniques used in our analysis, with a focus on kernel-based methods that are well-suited for handling circular and cylindrical data. These tools make it possible to estimate distributions and model relationships in a data-driven way, capturing key features without relying on predefined parametric forms. We begin by discussing kernel density estimation for real-valued, circular and cylindrical variables, followed by regression models for circular responses, techniques for assessing the influence of covariates, and a kernel-based approach to circular modal regression.

1.2.1 Kernel density estimation for cylindrical variables

Let X be a real-valued random variable and $X_1, \ldots, X_n$ a sample of n independent observations from X, which has density function f. The kernel density estimator of f ([35, 33]) is given by

$$\widehat{f}(x) = \frac{1}{n} \sum_{i=1}^{n} L_h(X_i - x), \tag{1.1}$$

with the kernel function L_h being a symmetric density, with support on the real line (or on an interval of the real line) and centered at zero, with h, called bandwidth, controlling the dispersion of the density. Large values of h lead to an estimator $f(x)$ with large bias and small variance, while smaller values of h reduce the bias at the cost of increasing the variance. Therefore, in practice, a data-driven selection of this parameter is of crucial importance.

Kernel density estimation of a circular density f was introduced in the spherical context by [22] and [6], and considered specifically for circular data by [15]. Let Θ be a circular random variable, and $\Theta_1, \ldots, \Theta_n$ a sample of n independent observations from Θ. We consider the estimator

$$\widehat{f}(\theta) = \frac{1}{n} \sum_{i=1}^{n} K_\rho(\Theta_i - \theta), \tag{1.2}$$

where K_ρ is a circular kernel function, i.e., a symmetric circular density function centered at zero with ρ controlling the concentration of the density. The parameter ρ plays a role opposite to the bandwidth in linear kernel density estimation: it governs the trade-off between bias and variance. A large value of ρ results in a more concentrated kernel, producing a rougher estimate with lower bias but higher variance. Conversely, a lower ρ leads to a smoother density estimate with increased bias and reduced variance.

Bivariate kernel density estimation in the case where one of the variables is spherical was studied by [18], which includes cylindrical variables as a particular case. Let (Θ, X) be a cylindrical random variable with density function f. The estimator of f is given by a product kernel of the form

$$\widehat{f}(\theta, x) = \frac{1}{n} \sum_{i=1}^{n} K_\rho(\Theta_i - \theta) L_h(X_i - x). \tag{1.3}$$

Thus, this estimator depends on two smoothing parameters: ρ and h, which should also be chosen in a data-driven way.

1.2.2 Kernel regression for a circular response

Now we consider the problem of estimating the regression function when the response variable, denoted by Φ, has a circular support. Let Δ be a general covariate, which can have support on the circumference or on the real line. We consider the model

$$\Phi = [m(\Delta) + \varepsilon] (\text{mod } 2\pi), \tag{1.4}$$

where ε is a circular random error such that $\mathbb{E}[\sin \varepsilon | \Delta = \delta] = 0$. The regression function m is the mean direction of Φ conditioned to the value of Δ, that is

$$m(\delta) = \text{atan2}[\mathbb{E}(\sin \Phi | \Delta = \delta), \mathbb{E}(\cos \Phi | \Delta = \delta)].$$

Given a bivariate random sample of (Δ, Φ), i.e., $\{(\Delta_i, \Phi_i)\}_{i=1,\ldots,n}$, [11] proposed estimating m as

$$\widehat{m}_\nu(\delta) = \text{atan2}[\widehat{m}_{1,\nu}(\delta), \widehat{m}_{2,\nu}(\delta)], \tag{1.5}$$

with

$$\widehat{m}_{1,\nu}(\delta) = \sum_{i=1}^{n} R_\nu(\Delta_i - \delta) \sin \Phi_i,$$
$$\widehat{m}_{2,\nu}(\delta) = \sum_{i=1}^{n} R_\nu(\Delta_i - \delta) \cos \Phi_i,$$

where the atan2 operator is the two-argument arctangent which returns the angle between the x-axis and the vector from the origin to (x, y). Additionally, R_ν is either a circular kernel function ($R_\nu(\Delta_i - \delta) = K_\rho(\Theta_i - \theta)$) if the covariate is circular or a regular kernel function for a real-valued covariate ($R_\nu(\Delta_i - \delta) = L_h(X_i - x)$).

As in the density estimation problem, the selection of the smoothing parameter ν (equivalently ρ when $\Delta = \Theta$ and h if $\Delta = X$) is of great importance, as it will control the bias-variance trade-off.

Next, we consider the problem of a circular response variable Φ and two covariates of different nature, Θ and X. Therefore, we assume the model

$$\Phi = [m(X, \Theta) + \varepsilon] (\text{mod } 2\pi), \tag{1.6}$$

where the regression m depends on the two covariates and is analogously defined as the expected direction of Φ conditioned to the values of the covariates. [30] extended the previous kernel estimator to the case of multiple real-valued covariates by using multivariate kernels. In our case, where there is one circular and one real-valued covariate, we employ a product kernel as in (1.3). Thus, given a trivariate iid sample $\{(\Theta_i, X_i, \Phi_i)\}_{i=1,\ldots,n}$ from (Θ, X, Φ), the kernel estimator takes the form

$$\widehat{m}_\nu(x, \theta) = \text{atan2}[\widehat{m}_{1,\nu}(x, \theta), \widehat{m}_{2,\nu}(x, \theta)], \tag{1.7}$$

with

$$\widehat{m}_{1,\nu}(x, \theta) = \sum_{i=1}^{n} L_h(X_i - x) K_\rho(\Theta_i - \theta) \sin \Phi_i,$$
$$\widehat{m}_{2,\nu}(x, \theta) = \sum_{i=1}^{n} L_h(X_i - x) K_\rho(\Theta_i - \theta) \cos \Phi_i.$$

The bivariate regression estimator thus depends on two different smoothing parameters, h and ρ.

1.2.3 Testing the effect of a covariate

Before estimating the regression functions, it is useful to test if there is an actual effect of the covariate on the mean direction of the response when considering model (1.4). A methodology to test the significance of the covariate in this context was proposed by [2]. The hypotheses of the test are given by

$$H_0: \quad \Phi = [\gamma + \varepsilon](\text{mod } 2\pi) \tag{1.8}$$

$$H_1: \quad \Phi = [m(\Delta) + \varepsilon](\text{mod } 2\pi), \tag{1.9}$$

with $\mathbb{E}[\sin\varepsilon|\Delta = \delta] = 0$ and where, under the alternative hypothesis, there exists a set of values of Δ of non-zero measure for which the regression function, $m(\Delta)$, is not equal to γ.

Under the null hypothesis, γ is the mean direction of Φ, and can be estimated as

$$\widehat{\gamma} = \text{atan2}\left(\sum_{i=1}^{n} \sin\Phi_i, \sum_{i=1}^{n} \cos\Phi_i\right).$$

Under the alternative, the regression function is estimated as in (1.5). The test statistic is then given by

$$T = \frac{\sum_{i=1}^{n}[1 - \cos(\Phi_i - \widehat{\gamma})] - \sum_{i=1}^{n}[1 - \cos(\Phi_i - \widehat{m}(\Delta_i))]}{\sum_{i=1}^{n}[1 - \cos(\Phi_i - \widehat{m}(\Delta_i))]}.$$

This statistic measures the relative reduction in average angular distance to the conditional mean when moving from the null to the alternative model, normalized by the average distance under the alternative. The distribution of T under H_0 is approximated by a bootstrap strategy described in [2].

When applying this test, one must account for the fact that nonparametric procedures are inherently sensitive to the choice of smoothing parameter, a well known issue discussed by [7]. Since $\widehat{m}$ suffers from some bias, a smoothing level that works well for estimation might not yield reliable results for hypothesis testing, and [2] recommend using values of the smoothing parameter which undersmooth the regression function. In addition, it's good practice to evaluate the test over multiple smoothing parameters within a sensible range, rather than relying on a single value.

1.2.4 Circular modal regression based on kernel smoothing

In some settings, the conditional distribution of the response may exhibit more than one mode for a given covariate value. To account for this, we also consider a multimodal regression approach ([9]), which focuses on estimating the local modes of the conditional distribution rather than its mean. This method complements the classical kernel regression described earlier and is particularly useful when the response displays clustered or heterogeneous behavior across the range of the covariate.The presence of multiple local modes may suggest the influence of an unobserved covariate or an interaction effect with the covariate under study.

Multimodal regression in the context of circular variables was studied by [3]. For a circular response Φ and a general covariate Δ, the regression multifunction is defined as the set of modes of the conditional density, that is

$$M(\delta) = \left\{\phi \in [0, 2\pi) : \frac{\partial}{\partial\delta} f(\phi|\delta) = 0, \frac{\partial^2}{\partial\delta^2} f(\phi|\delta) < 0\right\}. \tag{1.10}$$

Note that in (1.10) we could replace the conditional density function $f(\phi|\delta)$ by the joint density function $f(\delta, \phi)$. Thus, [3] estimates (1.10) in a nonparametric way by first estimating the joint density function and then computing its local maxima with respect of the response variable:

$$\widehat{M}(\delta) = \left\{ \phi \in [0, 2\pi) : \frac{\partial}{\partial \delta} \widehat{f}(\phi, \delta) = 0, \frac{\partial^2}{\partial \delta^2} \widehat{f}(\phi, \delta) < 0 \right\}. \tag{1.11}$$

The estimator $\widehat{f}(\phi, \delta)$ is given by (1.3) in the case where Δ is a real-valued variable. If Δ has circular support, the estimator is given by a product of circular kernels. To compute the local maxima, a conditional version of the directional mean shift algorithm (see [45, 44]) is used.

Again, two smoothing parameters must be selected because the estimation of the joint density involves two smoothing parameters. However, their role is somewhat different than in the density estimation scenario: the parameter ρ associated to the response variable controls the number of local modes for a given value of δ, with higher values leading to multiple local modes and smaller values leading to a unimodal estimator. On the other hand, the parameter associated to the covariate Δ controls the smoothness of the regression multifunction.

1.3 Data analysis and discussion

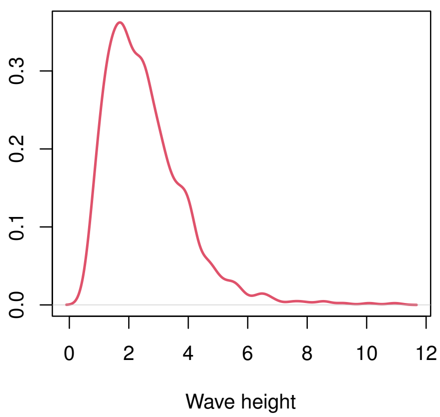
This section presents an exploratory and inferential analysis of the available data, with the goal of understanding the relationship between wave direction and various wind-related variables. We begin by examining the distributional characteristics of each variable individually and in pairs, using both univariate and bivariate kernel density estimates. This is followed by formal testing to assess whether wind speed and direction have a significant effect on the mean wave direction. Finally, we approach the regression problem from both classical and multimodal perspectives, first considering each covariate separately and then jointly, to better capture the structure and possible complexities in the data. For most methods, the *R* package *NPCirc* (see [4]), an updated version of [31], is used.

We first estimate the density function of wave height, wind speed, wave direction and wind direction individually. For the real-valued variables, we use the kernel estimator in (1.1), where the kernel L_h is a Gaussian density. Many different methods have been proposed in the literature to select the smoothing parameter h in a data-driven way, from a simple rule of thumb ([40]), to more elaborated plug-in approaches or other methods such as cross validation ([38]). Here, we employ the two-step plug-in approach of ([39]). Figures 1.1(a) and 1.1(b) illustrate the resulting kernel density estimates for wave height and wind speed, respectively. Wave height shows a markedly skewed distribution, with low-height waves occurring frequently and waves exceeding 6 meters being relatively rare. Wind speed also exhibits a slight right skewness, with speeds above 18 m/s being uncommon.

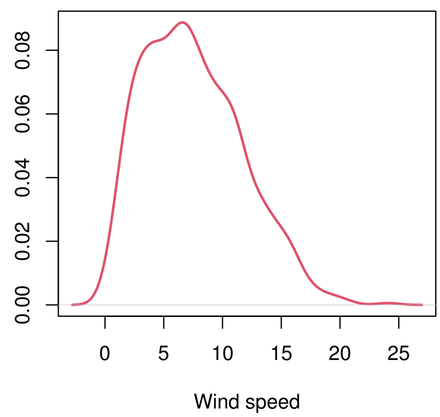
Next we estimate the density functions of the two circular functions with the estimator in (1.2). We employ the von Mises kernel, given by

$$K_\rho(\theta) = \frac{1}{2\pi I_0(\rho)} \exp\{\rho \cos\theta\},$$

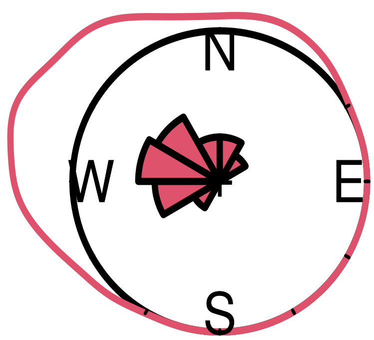
with I_0 being the Bessel function of the first kind and order zero. Although the problem of the data-driven selection of ρ has not been as studied as in the real-valued case, there are several proposals ([41, 32, 20, 5, 47]). We follow the approach of [5], selecting ρ as a two-stage plug-in estimate of the optimal concentration minimizing the asymptotic mean integrated squared error of the density estimator. The resulting estimates of the density



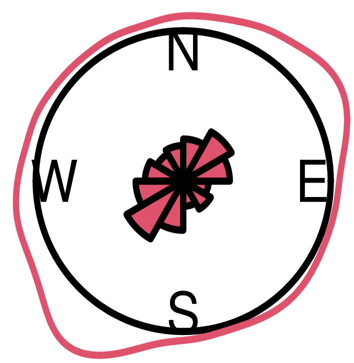
(a) Wave height



(b) Wind speed



(c) Wave origin direction



(d) Wind origin direction

FIGURE 1.1
Kernel density estimators of wave height (a), wind speed (b), wave direction (c) and wind direction (d).

functions of wave and wind origin directions are displayed in Figures 1.1(c) and 1.1(d), respectively. We see that most waves originate from the North-West, although there seems to be a small mode in the North-East direction. Wind directions exhibit a bimodal pattern, predominantly from the North-East and South-West.

We now examine the joint density of the two cylindrical variables associated with wave and wind behavior. We employ the product kernel density estimate in (1.3), with L_h being a Gaussian kernel and K_ρ being the von Mises kernel. Although the problem of selecting the two smoothing parameters in practice has not received much attention, [18] derived explicit expressions for the theoretical optimal smoothing parameters in terms of minimizing the asymptotic mean integrated squared errors under a specific parametrization and [46] derived a explicit formula for the least squares cross-validation loss. In this work, we use a plug-in approach were we minimize the mean integreated squared error where the unknown quantities are replaced by parametric estimators based on a mixture of von Mises and Gaussian densities, extending the ideas of [32] and [20] to the bivariate case.

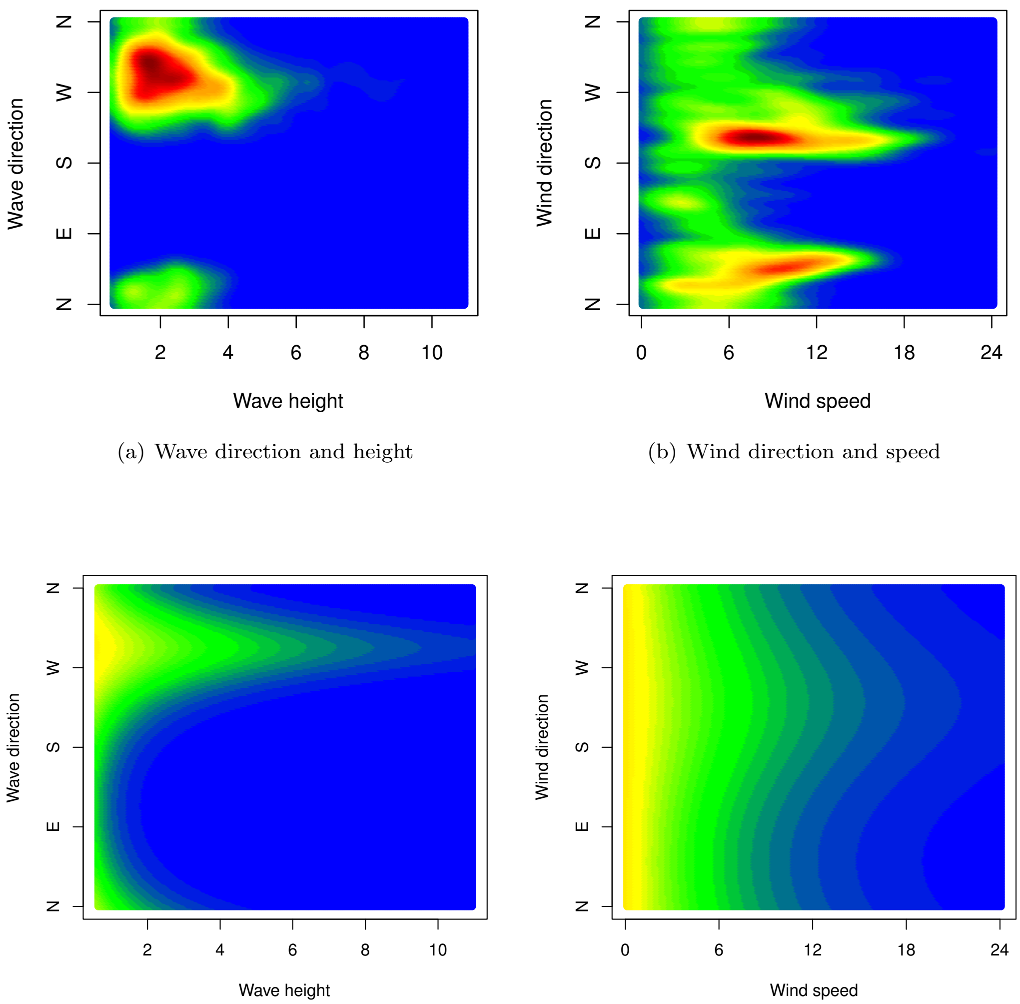
The too panels of Figure 1.2(a) show the cylindrical kernel density estimate for wave direction and height. A distinct mode is observed for low wave heights originating from the North-West. For these lower waves, directions predominantly span the West and North-West, with additional contributions from the North-East and South-West. In contrast, higher waves tend to originate mainly from the West and North-West. Figure 1.2(b) shows the corresponding estimate for wind direction and speed. The primary mode occurs at moderate wind speeds (around 8 m/s) from the South-West, with a secondary local mode for stronger winds from the North-East. Additionally, wind direction is relatively uniform at lower speeds, while higher wind speeds are more concentrated around the South-West and North-East directions. For comparison purposes, the bottom panels of Figure 1.2(a) present the density estimates based on the parametric model of [24], where the parameters are estimated by maximum likelihood. Compared to the nonparametric cylindrical kernel density estimates, the parametric model of Johnson and Wehrly proves too restrictive. While it provides a smooth approximation of the joint distribution, it fails to adequately capture the multimodal structure observed in the data. In particular, the kernel estimates reveal clear bimodality in both the wave and wind direction distributions—most notably for low wave heights and moderate wind speeds—whereas the parametric model tends to smooth over these features, resulting in unimodal or overly simplistic patterns. This limitation highlights the challenges of relying solely on parametric assumptions when modeling complex environmental data, where directional and magnitude interactions often exhibit rich, multimodal behavior.

To assess whether wind speed and wind direction significantly influence the conditional mean direction of the waves, we applied the significance test based on hypotheses (1.9). For each covariate, the test was conducted across 15 different values of the smoothing parameters, including those selected via cross-validation. When considering wind speed as the covariate, the p-value was consistently below 0.001 for all parameter values, indicating a statistically significant effect of wind speed on the mean wave direction. Similar results were obtained for wind direction, supporting the conclusion that it also has a significant influence on the mean wave direction.

We next employ the nonparametric regression estimator in (1.5) to investigate how the mean wave direction varies with wind speed. We employ a Gaussian kernel and select ν by cross-validation, as the minimizer of the function

$$\sum_{i=1}^{n}\{1-\cos[Y_i-\widehat{m}_\nu^{-i}(\Delta_i)]\}. \tag{1.12}$$

The estimated regression function is depicted as a dotted line in Figure 1.3(a). Results



(a) Wave direction and height

(b) Wind direction and speed

(c) Wave direction and height

(d) Wind direction and speed

FIGURE 1.2
Cylindrical density estimators of wave direction and wave height (a) and wind direction and speed (b). Top row shows kernel density estimators and bottom row shows parametric estimators of the Johnson-Wehrly model.

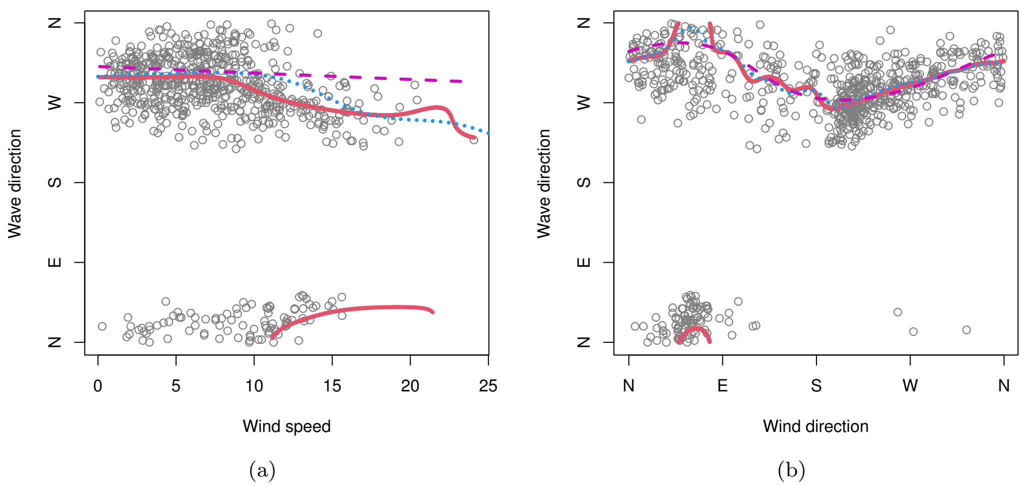


FIGURE 1.3
Scatter plots of wave direction vs. wind speed (a) and wave direction vs. wind direction (b) with parametric estimators of the conditional mean (dashed trace), kernel regression estimators of the conditional mean (dotted trace) and of the conditional modes (continuous trace).

indicate that the mean wave origin is primarily from the North-West, gradually shifting toward the West as wind speed increases. For comparison purposes, we also compare with the parametric regression model of [17], which assumes that the conditional distribution is a von Mises density with mean direction given by $\beta_1 + \text{atan}(\beta_2 x)$. Such estimator is depicted as a dashed line, exhibiting an almost constant mean direction centered around the North-West, regardless of wind speed. This contrast highlights a key limitation of the parametric model: its reduced flexibility in capturing nonlinear trends in the directional response.

Notably, a subset of data points shows wave directions from the North-East at moderately high wind speeds. To account for this, we apply the multimodal regression estimator in (1.11), illustrated by the continuous line in Figure 1.3(a). For this, two smoothing parameters must be selected. One might be tempted to use the same strategy as in multiple kernel density estimation; however, it is widely known that smoothing parameters that are optimal for the estimation of the density function are not optimal for the estimation of its local modes (see [8]). Here, we use the modal cross-validation approach proposed by [3]. This estimator reveals a single local mode at low wind speeds and a second mode emerging when wind speed exceeds 10 m/s. This pattern may reflect interactions with wind direction or the influence of additional variables.

We also examine the marginal effect of wind direction on wave direction. We also apply estimator (1.5), where the kernel is given by a von Mises density and the smoothing parameter is again selected by cross-validation. We also compare to the parametric estimator of [36]. In Figure 1.3(b), both the parametric (dashed line) and nonparametric (dotted line) od the conditional mean are shown, in addition to the conditional local mode estimator (solid line). The three estimates largely coincide across most wind directions, indicating good agreement between methods. However, for wind directions between North and East, a noticeable discrepancy arises: the modal estimate shifts toward the North-East, the parametric mean remains closer to the North-West, and the nonparametric mean lies in between. This behavior suggests a skewed conditional distribution in this range of wind directions. While the parametric model appears appropriate in general, this localized deviation highlights its limited ability to accommodate asymmetry, which can be better captured by nonparametric and modal-based approaches.

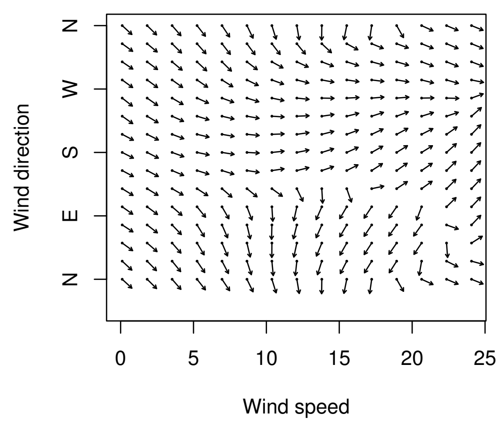


FIGURE 1.4
Representation of the estimated regression function of wave direction with wind direction and speed as covariates. Arrows represent the estimated direction of of waves.

Finally, we analyze the full regression model (1.6), incorporating both wind speed and wind direction as covariates. We use the nonparametric estimator defined in (1.7), where the two smoothing parameters are selected by cross-validation in an analogue way to (1.12). We visualize the estimated wave direction over a grid of covariate values in Figure 1.4. The results indicate that for winds originating from the South and West, the estimated wave direction is primarily from the North-West under weak wind conditions. As wind speed increases, the wave direction gradually shifts toward the West, and for strong winds from the South, it further shifts toward the South-West. Conversely, for winds coming from the North-East, the estimated wave direction transitions from the North-West to the North, and eventually to the North-East.

These findings suggest that both wind speed and direction jointly influence wave propagation patterns, with stronger winds exerting a more pronounced directional effect. This highlights the importance of considering multivariate interactions when modeling wave behavior.

1.4 Conclusions

In this study, we have examined the behavior of wave direction under the influence of key maritime and meteorological factors, using flexible nonparametric methods designed for circular and cylindrical data. The analysis, based on observations from a coastal monitoring station in northwestern Spain, reveals rich structure in the relationships between wave direction and covariates such as wave height, wind direction, and wind speed.

Although wave direction prediction is traditionally approached through deterministic models based on differential equations, these models often require statistical fine-tuning to adapt to local or short-term variability. In this context, nonparametric, data-driven methods offer a valuable complement. By avoiding rigid model assumptions, these approaches are particularly well suited to capturing the complex, nonlinear, and possibly multimodal dependencies that characterize real-world oceanographic processes.

Kernel-based estimators enabled us to uncover patterns not easily captured by parametric models, including shifts in dominant wave direction with varying wind speed or direction. Modal regression further highlighted situations where the conditional distribution of wave direction departs significantly from unimodality or symmetry, underscoring the limitations of mean-based analysis in such contexts.

Our significance testing procedures confirmed the relevance of wind speed and direction as individual covariates. However, a key limitation remains: tests were performed separately for each covariate, leaving open the need for multivariate significance testing frameworks that can jointly assess the influence of several covariates. Likewise, modal regression, while useful, was restricted to models with a single covariate. Extending this to handle multiple covariates simultaneously, especially in the cylindrical context, is a promising direction for future work.

Another important avenue is the incorporation of temporal dependence. The sea state is inherently dynamic, and successive observations are not independent. In this sense [43] and [25] proposed parametric methods to account for spatio-temporal processes. Extending these nonparametric methods to account for temporal correlation could be tackled by employing another version of cross-validation for selecting the smoothing parameter, such as the leave-$(2l+1)$-out instead of classical leave-one-out selectors. It would also be convenient to account for temporal dependence in inferential tasks, for example by adapting the bootstrap strategy in Section 1.2.3. In addition, time-varying density estimation, autoregressive structures, or functional data approaches could be explored to enhance both the descriptive and predictive capabilities of the models.

Altogether, our findings illustrate the strengths of flexible, geometry-aware statistical tools in environmental data analysis. These methods serve not only to refine predictions from physical models but also to offer new insights into the structure of directional phenomena. Future research should continue to bridge the gap between physical and statistical modeling, develop more comprehensive inference tools, and explore dynamic extensions suited to evolving environmental conditions.

Bibliography

[1] T. Abe and C. Ley. A tractable, parsimonious and flexible model for cylindrical data, with applications. *Econometrics and Statistics*, 4:91–104, 2017.

[2] M. Alonso-Pena, J. Ameijeiras-Alonso, and R. M. Crujeiras. Nonparametric tests for circular regression. *Journal of Statistical Computation and Simulation*, 19:477–500, 2021.

[3] M. Alonso-Pena and R. M. Crujeiras. Analizing animal escape data with circular nonparametric multimodal regression. *Annals of Applied Statistics*, 17:130–152, 2023.

[4] M. Alonso-Pena, M. Oliveira, J. Ameijeiras-Alonso, R. M. Crujeiras, M. Gijbels, and A. Rodríguez-Casal. *R package `NPCirc`: Nonparametric Circular Methods (version 3.1.1)*, 2022.

[5] J. Ameijeiras-Alonso. A reliable data-based smoothing parameter selection method for circular kernel estimation. *Statistics and Computing*, 34, 2024.

[6] Z. D. Bai, C. R. Rao, and L. C. Zhao. Kernel estimators of density function of directional data. *Journal of Multivariate Analysis*, 27:24–39, 1988.

[7] A. W. Bowman and A. Azzalini. *Applied Smoothing Techniques for Data Analysis: the Kernel Approach with S-Plus Illustrations.* Oxford University Press, 1997.

[8] A. Casa, J. E. Chacón, and G. Menardi. Modal clustering asymptotics with applications to bandwidth selection. *Electronic Journal of Statistics*, 14:835–856, 2020.

[9] Y.-C. Chen, C. R. Genovese, R. J. Tibshirani, and L. Wasserman. Nonparametric modal regression. *The Annals of Statistics*, 44:489–514, 2016.

[10] M. Di Marzio, A. Panzera, and C. C. Taylor. Kernel density estimation on the torus. *Journal of Statistical Planning and Inference*, 141:2156–2173, 2011.

[11] M. Di Marzio, A. Panzera, and C. C. Taylor. Non-parametric regression for circular responses. *Scandinavian Journal of Statistics*, 40:238–255, 2012.

[12] M. Di Marzio, A. Panzera, and C. C. Taylor. Nonparametric circular quantile regression. *Journal of Statistical Planning and Inference*, 170:1–14, 2016.

[13] T. D. Downs and K. V. Mardia. Circular regression. *Biometrika*, 89(3):683–698, 2002.

[14] J. Fan and I. Gijbels. *Local Polynomial Modelling and its Applications.* Chapman & Hall, 1996.

[15] N. I. Fisher. Smoothing a sample of circular data. *Journal of Structural Geology*, 11:775–778, 1989.

[16] N. I. Fisher. *Statistical Analysis of Circular Data.* Cambridge University Press, 1993.

[17] N. I. Fisher and A. J. Lee. Regression models for angular responses. *Biometrics*, 48:645–677, 1992.

[18] E. García-Portugués, R. M. Crujeiras, and W. González-Manteiga. Kernel density estimation for directional-linear data. *Journal of Multivariate Analysis*, 121:152–275, 2013.

[19] E. García-Portugués, R. M. Crujeiras, and W. González-Manteiga. Central limit theorems for directional and linear random variables with applications. *Statistica Sinica*, 25:1207–1229, 2015.

[20] E. García-Portugués. Exact risk improvement of bandwidth selectors for kernel density estimation with directional data. *Electronic Journal of Statistics*, 7:1655–1685, 2013.

[21] A. L. Gould. A regression technique for angular variates. *Biometrics*, 25:683–700, 1969.

[22] P. Hall, G. S. Watson, and J. Cabrera. Kernel density estimation with spherical data. *Biometrika*, 74:751–762, 1987.

[23] S. R. Jammalamadaka and A. SenGupta. *Topics in Circular Statistics.* World Scientific, 2001.

[24] R. A. Johnson and T. E. Wehrly. Some angular-linear distributions and related regression models. *Journal of the American Statistical Association*, 73:602–606, 1978.

[25] G. Jona-Lasinio, A. E. Gelfand, and G. Mastrantonio. Spatial and spatio-temporal circular processes with application to wave directions. In *Applied Directional Statistics*, pages 145–178. Chapman and Hall/CRC, 2018.

[26] S. Kim and A. SenGupta. Multivariate-multiple circular regression. *Journal of Statistical Computation and Simulation*, 87:1277–1291, 2017.

[27] C. Ley and T. Verdebout. *Modern Directional Statistics.* CRC Press, 2017.

[28] K. V. Mardia and P. E. Jupp. *Directional Statistics.* Wiley, 2000.

[29] K. V. Mardia and T. W. Sutton. A model for cylindrical variables with applications. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 40:229–233, 1978.

[30] A. Meilán-Vila, M. Francisco-Fernández, R. M. Crujeiras, and A. Panzera. Nonparametric multiple regression estimation for circular response. *TEST*, 30:650–672, 2021.

[31] M. Oliveira, R. M. Crujeiras, and A. Rodríguez-Casal. NPCirc : An R package for nonparametric circular methods. *Journal of Statistical Software*, 61:1–26, 2014.

[32] M. Oliveira, R.M. Crujeiras, and A. Rodríguez-Casal. A plug-in rule for bandwidth selection in circular density estimation. *Computational Statistics and Data Analysis*, 56(12):3898–3908, 2012.

[33] E. Parzen. On Estimation of a Probability Density Function and Mode. *The Annals of Mathematical Statistics*, 33(3):1065 – 1076, 1962.

[34] B. Presnell, S. P. Morrison, and R. C. Littel. Projected multivariate linear models for directional data. *Journal of the American Statististical Association*, 93:1068–1077, 1998.

[35] M. Rosenblatt. Remarks on Some Nonparametric Estimates of a Density Function. *The Annals of Mathematical Statistics*, 27(3):832 – 837, 1956.

[36] Y. Sarma and S. R. Jammalamadaka. Circular regression. In *Statistical Science and Data Analysis: Proceedings of the Third Pacific Area Statistical Conference*, pages 109–128, Utrecht, Netherlands, 1993. VSP.

[37] D. W. Scott. *Multivariate Density Estimation.* Wiley, 1992.

[38] D. W. Scott and G. R. Terrell. Biased and unbiased cross-validation in density estimation. *Journal of the American Statistical Association*, 82(400):1131–1146, 1987.

[39] S. J. Sheather and M. C. Jones. A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society: Series B (Methodological)*, 53(3):683–690, 2018.

[40] B. W. Silverman. Using kernel density estimates to investigate multimodality. *Journal of the Royal Statistical Society, Series B (Statistical Methodology).*, 43:97–99, 1981.

[41] C. C. Taylor. Automatic bandwidth selection for circular density estimation. *Computational Statistics and Data Analysis*, 52(7):3493–3500, 2008.

[42] M. P. Wand and M. C. Jones. *Kernel Smoothing.* Chapman & Hall, 1995.

[43] F. Wang, A. E. Gelfand, and G. Jona-Lasinio. Joint spatio-temporal analysis of a linear and a directional variable: Space-time modeling of wave heights and wave directions in the adriatic sea. *Statistica Sinica*, 25(1):25–39, 2015.

[44] Y. Zhang and Y.-C. Chen. The EM perspective of directional mean shift algorithm. *arXiv*, arXiv:2101.10058, 2021.

[45] Y. Zhang and Y.-C. Chen. Kernel smoothing, mean shift, and their learning theory with directional data. *Journal of Machine Learning Research*, 22:1–92, 2021.

[46] Y. Zhang and Y.-C. Chen. Mode and ridge estimation in euclidean and directional product spaces: A mean shift approach. *Journal of Computational and Graphical Statistics*, 2025.

[47] S. Zámečník, I. Horová, S. Katina, et al. An adaptive method for bandwidth selection in circular kernel density estimation. *Computational Statistics*, 39:1709–1728, 2024.