Modelling mean wave direction distribution with the von Mises model

J. L. Vega & G. Rodríguez
Departamento de Física, Universidad de Las Palmas de Gran Canaria, Spain

Abstract

This paper presents the probabilistic modelling of the mean wave direction derived from directional spectral analysis of waves recorded by directional buoys. The analysis is performed on the mean wave direction in terms of the climatic season, the sea state severity and the period of the dominant waves. The usefulness of the von Mises theoretical models to describe the empirical kernel density estimates is examined. It is observed that the single von Mises theoretical model results are useful to fit the observed distribution only for moderate and severe sea states while the mixture of two von Mises distributions enhances significantly the degree of fitness.
Keywords: wave modelling, mean wave direction, kernel density estimation, circular variables, von Mises distribution, von Mises mixtures.

1 Introduction

Probabilistic design and assessment of marine structures interacting with sea waves requires a reliable knowledge of the long-term wave climate. In this context, it is generally assumed that directional wave spectra provide a complete description of a given sea state. However, it is common practice to accept that a sea state in simpler terms is reasonably well characterized by means of three parameters derived from it. These are the significant wave height H_{m0}, the spectral peak period T_p and the mean direction θ_m. Accordingly, it is generally assumed that long time series of these parameters allows one to obtain a convenient description of the long-term wave climate by estimating the joint and

WIT Transactions on Ecology and the Environment, Vol 126, © 2009 WIT Press
www.witpress.com, ISSN 1743-3541 (on-line)
doi:10.2495/CP090011

marginal probability density functions of these three parameters. Nonetheless, the sea state severity is commonly presented in the form of univariate and bivariate histograms of significant wave height and the peak period, or the mean zero-upcrossing wave period.

The above commented procedure does not give any consideration to the directions of waves approaching at a site. It implicitly assumes that all wave directions are equally likely to occur, or in other words, there are no preferred directions for the sea states approaching the point of measurement. However, directional wave information is often required for a variety of applications, including coastal engineering and nearshore dynamics, waste dispersal and pollution studies, sediment transport and beach erosion. Thus, for example, wave-driven currents are the principal mechanism for sand transport on most of the world coasts. The direction of these currents is governed by the direction of the deepwater ocean waves and their subsequent refraction over the shoaling zone. As a consequence, the long-term wave climate is properly described by the joint probability density function of the significant wave height, spectral peak period and average wave direction. Nevertheless, in practical applications using directional information, it is common to use the conditional joint distributions of wave height and period given the mean wave direction, for a certain number of directional sectors.

This study focuses on the long-term scale mean wave direction variability, which is necessary to identify the wave climate in a given area. In particular, the purpose of the current paper is to establish the main properties of the mean wave directional regime in the area of the directional buoy located off Estaca de Bares, a location in the Galician coast, at the Spanish North-Atlantic coast, and to assess the usefulness of the von Mises probability model to fit the observed probability distribution, considering the single symmetric and unimodal von Mises and the two components mixture of von Mises theoretical probability models.

To reach this goal, long time series of H_{m0}, T_p, and θ_m, are analysed. The marginal probabilistic structure of the mean wave direction, as well as its variability as a function of climatic seasons is examined, and contrasted with the single von Mises and a mixture of two von Mises models for circular random variables. Furthermore, the conditional distributions of mean wave direction for various wave height and period thresholds are estimated and the usefulness of these theoretical models to fit the observed empirical distributions is assessed.

The presentation of the study is structured as follows. Section 2 describes the instrumental wave measurement data set examined in the study and indicates the location of the buoy used for recording the data. This section includes a brief summary of the kernel density estimation procedure with emphasis on its use for circular data. Next, in section 3, the main properties of the von Mises probability distribution family, used as theoretical model to fit the observed density functions are summarised. The discussion of results obtained by examining the observed time series is presented in section 4. Concluding remarks are given in section 5.

2 The field site and data analysis

2.1 Field site and recorded data

Wave observations used in this study were performed by a buoy of the offshore network implemented by Puertos del Estado (Ministerio de Fomento). This is a Seawatch directional buoy deployed in the Galician coast (Northwest of Spain), offshore from A Coruña, at a point of latitude 44º 03,94' N and longitude 07º 37,27' W. This location is shown in Figure 1 for helping the comments of results. The water depth at the measuring point was 387 meters.

The Seawatch buoy measured short-term records at three hour intervals during 1997 and hourly till 2003. Measurements used in this study span over a relatively long period, starting on January 1997 and lasting on December 2003.

Figure 1: Location of the directional buoy offshore of Estaca de Bares, (Galician Coast) Northwest Spain.

Due to problems with power supply, change of the internal battery of the buoy, failure in remote data transmission via satellite, retrieval for cleaning biofouling growth on the outer surface, and other logistical mishaps caused loss of data during some periods. Hence 100% data could not be collected. Also the collected time series were subjected to error checks and only the records which were found suitable were included for posterior analysis. A total of 43227 usable records were obtained over this period of seven years.

Short-term analysis of directional wind-wave spectra permits to derive various characteristic parameters, such as the significant wave height H_{m0}, the peak period T_p and the mean wave direction θ_m, among others.

Thus, the data base used in this study consisted of an hourly (except during 1997) valued trivariate time series $\{H_{m0}, T_p, \theta_m\}$.

2.2 Kernel density estimation

Kernel density estimation is a nonparametric method of estimating a pdf from data that is related to the histogram [1]. Given a set of n observations $x_1; \ldots; x_n$, the kernel density estimate (*kde*) is defined as.

$$f(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right) \tag{1}$$

where K is a function named as the kernel and given by some smooth density. In practice however, the choice of kernel appears to have very little effect on the performance of the kernel estimator, and in most cases the Gaussian kernel is used for simplicity, such as in this study. In contrast, the choice of bandwidth, h, is of crucial importance for the performance of the kde. The bandwidth is a positive number. The value of h basically decides how many observations are included in the estimation of f(x) at the point x. So a small choice of bandwidth means that only observations very close to x are used in the estimation, while a large bandwidth includes most of the observations in the sample. Since the observations close to x are more likely to carry information about the density's behaviour at that point, we would expect precision of the density estimator to increase, and thereby the bias to decrease, as we decrease h. On the other hand, as we decrease h, fewer observations are used to estimate f(x), so we would expect the variance of our estimator to increase as we decrease h. So, there is a tradeoff between choosing a small vs. a large bandwidth. Nevertheless, some practical rules which permit to obtain reasonable values have been proposed. In the present paper that suggested in Fisher [2] has been used. That is,

$$h = \sqrt{7}\,\frac{\sigma}{n^{-1/5}} = \sqrt{\frac{7}{\kappa}}\,n^{-1/5} \tag{2}$$

In the particular case of directional data, the natural domain of definition of random variables is limited to an interval bounded on both sides. A useful approach to deal with data on the finite interval [0, 2π) is to impose periodic boundary conditions. That is, to wrap the kernel round the circle. Nevertheless, from a computational point of view, a simpler procedure consists in augmenting the data set by replicating it twice on the intervals [-2π, 0] and [2π, 4π]. The estimated probability density is then drawn in the range [0, 2π), such as suggested by Silverman [1]. This procedure has been used by Vega and Rodriguez [3].

3 The von Mises probability distribution

A circular random variable, θ, is defined as a random variable with support on the unit circle, i.e. the angle is in the range (0, 2π) radians or (0º, 360º). The probability density function of a circular variable, f(θ), is always positive

$$f(\theta) \geq 0 \quad \forall \quad 0 \leq \theta \leq 2\pi \tag{3}$$

satisfies the normalization condition

$$\int_0^{2\pi} f(\theta) d\theta = 1 \tag{4}$$

and the periodicity condition

$$f(\theta) = f(\theta + n2\pi) \quad ; n = 0, \pm 1, \pm 2, \cdots \tag{5}$$

This condition is peculiar to circular random variables. A linear random variable, R, is defined as a continuous random variable with support on the whole real line or an interval on it. Consequently, circular random variables must be analysed by using techniques differing from those appropriate for the usual Euclidean type variables because the circumference is a bounded closed space, for which the concept of origin is arbitrary or undefined.

The probability distribution model most frequently used in applications involving circular random variables is the von Mises family. A circular random variable, θ, is said to follow a von Mises distribution with parameters μ and κ, VM(μ, κ), if its probability density function is given by

$$f(\theta) = \frac{1}{2\pi I_0(\kappa)} \exp[\kappa \cos(\theta - \mu)] \quad ; 0 \leq \mu \leq 2\pi \quad , \kappa > 0 \tag{6}$$

where μ is the mean direction and κ is called the concentration parameter, and $I_0(\kappa)$ is the modified Bessel function of the first kind and order zero. Details on the estimation of these parameters can be found in Mardia and Jupp [4].

Unfortunately, the von Mises density function of a circular random variable is unimodal and symmetric. These facts make the von Mises model unsuitable for analysing circular data with more complicated features such as multimodality and/or skewness. One possible alternative in these situations is to use a mixture of von Mises distributions, given by

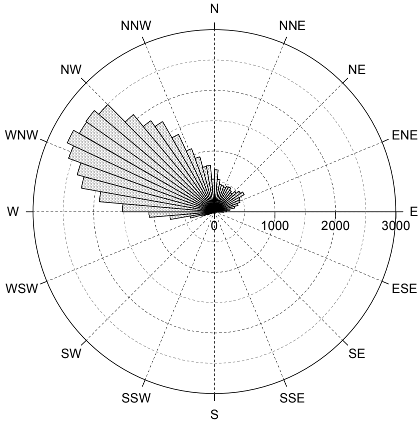
$$f(\theta) = \sum_{i=1}^{N} \lambda_i \frac{1}{2\pi I_0(\kappa_i)} \exp[\kappa_i \cos(\theta - \mu_i)] \tag{7}$$

where, for the *i*th component, p_i is the mixing proportion, μ_I is the mean direction and κ_i is the concentration parameter. The finite mixtures of von Mises distributions in both mean direction and concentration parameters are widely used in many disciplines, including astronomy, ecology, geology and medicine. However, it is worthwhile to mention that the use of mixtures of von Mises distributions present two main drawbacks. One is that testing the order or the number of components necessary in a circular mixture is a challenging problem, and the other one is the complexity of numerical routines necessary to fit these models with the maximum likelihood estimation [5]. For these reasons, the study limits to the use of second order von Mises mixtures and its comparison with a single von Mises distribution. Various methods for estimating the parameters in a von Mises mixture have been suggested in the literature. A comparison of several of these procedures has been presented in Spurr and Koutbeiy [6]. In this study, the estimation of the unknown parameters has been carried out by means of a non-linear least squares method, based on the Levenberg-Marquardt algorithm.

4 Results and discussion

The mean wave direction of the complete data set is represented, as a rose diagram in Figure 2 (left). It is observed that the largest part of sea states arrives at the zone coming from the WNW-NW sector. This is a clear indication of the fetch restrictions due to the geographical location of the buoy. Sea states approaching the buoy with South component are strongly restricted by the orientation of the North Spanish coast. Furthermore, the Gulf of Vizcaya conforms a relatively short basin in the sector N-E, in comparison with the sector N-W, open to Northwest Atlantic, where longer fetches can be delineated.

Detailed examination of the wave direction rose of Figure 2 reveals the presence of a secondary peak around the NE-ENE sector. The bimodality precludes the possibility that the single von Mises distribution fits properly the empirical distribution, such as observed in the right part of Figure 2. In this case the agreement between the mixture of two VM and the empirical distribution seems to be adequate, even though it probably could be improved by adding one more component to the von Mises mixture.



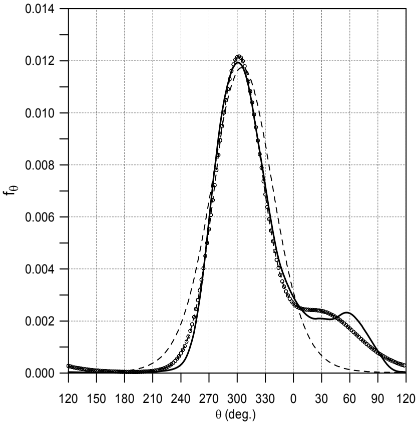


Figure 2: Mean wave direction rose (left) and empirical probability distribution of mean wave direction (solid line), and fitted VM (dashed line) and two VM mixture (dotted line) models (right), for the whole analysed period.

In spite of the bimodal character of the empirical distribution of mean wave direction, it is considerably clustered around a value close to 317 degrees, mainly in the sector between 240º and 360º. Thus, the contribution to the kde of the secondary peak could be due to some atmospheric conditions prevailing during some period during the year, or could be associated to some weather conditions occurring during a non specific time period. With this in mind, sea states have been classified by climatic seasons, sea state severity and the length or period of the dominant waves present in wave trains.

Results in terms of climatic seasons are shown in Figure 3. It can be observed that the distinction between the principal and the secondary peaks is smaller during summer. During this period the frequency of sea states approaching from the NE-ENE sector is lower and the bimodality reduces with an improvement in the fit with the two VM mixtures. The distinction between both peaks enhances progressively toward winter and accordingly the agreement between the mixture model and the *kde* get worse. Hence, separation of sea states by climatic seasons reveals that the bimodal character persists during the year and that the single VM model is not able to fit the observed probability distribution. In this context, the mixture of two VM improves the degree of fitness in all the seasons especially during summer, when the distribution is clearly asymmetric but bimodality reduces. This fact is reflected in the maximum likelihood estimated value of κ, the concentration parameter of the VM model, which is a measurement of the concentration around the mean direction, such as revealed by values given in Table 1, which gives the number of sea states, the average mean wave direction and κ for the full data set and for each season. It is observed that κ increases in summer enhancing the distribution peakedness.

Taking into account that the bimodal character of the observed mean wave direction distribution is not dependent of the climatic season, the variability of the observed distribution has been examined in terms of the sea state severity and by considering whether sea states approaching the measurement site have been locally or remotely generated, that is, filtering the data set for different significant wave height and spectral peak period thresholds.

The directional relationships during calmer periods are less relevant in the design of offshore and coastal systems; the greater interest is in the behaviour at higher sea states. The conditional mean wave direction distribution for six significant wave height thresholds, from 1 to 6 meters, is shown in Figure 4. It is evidenced that empirical distributions narrows and enhance around the modal value of θ_m, which shifts westward, as the H_{m0} threshold increases. This evolution is reflected by the statistical values given in Table 2. It is remarkable that for low thresholds the observed distribution present a similar structure to that associated to the full data set, but as the significant wave height threshold increases the von Mises, model, and naturally the two components mixture, fits more and more properly to the empirical density estimate because of the distribution becomes more concentrated about the mean direction, the symmetry increases and the bimodal character tends to disappear. These results reveal that the secondary peak is associated with sea states of low or moderate severity.

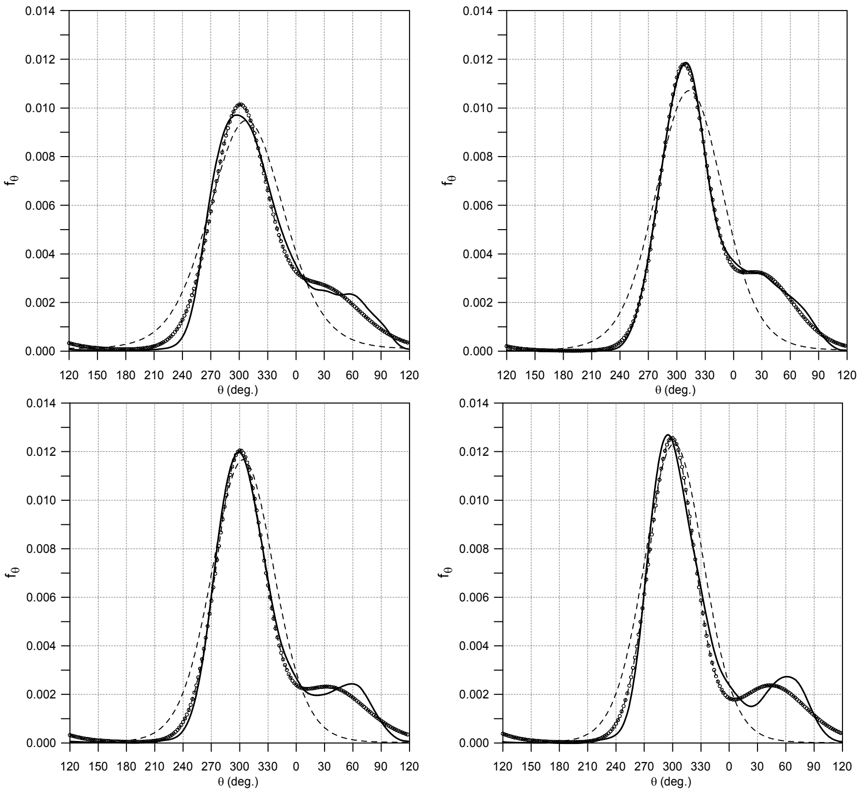


Figure 3: Empirical probability distribution of mean wave direction (solid line), the VM (dashed line) and two VM mixture models fitted (dotted line) for spring (up-left), summer (up-right), autumn (down-left), and winter (down-right) seasons.

The evolution of the empirical density function of the mean wave direction and the fit to a single VM and a two VM mixture for four threshold spectral peak period is shown in Figure 5. The corresponding values of the thresholds, as well as the basic circular statistics are given in Table 3. Results show that empirical distributions narrows and enhance around θ_m, which shifts westward, as the T_P threshold increases. Furthermore, it can be observed that for large values of T_P, that is, by removing low period sea states, the empirical distribution fits better to the von Mises models. This fact is particularly true for the mixture of two VM because the bimodal character disappear by filtering the sea states with low wave spectral period but even for large thresholds the distribution remains skewed, with a smoothly decaying plateau in the NE quadrant, which makes the single VM fail to fit the empirical *kde*.

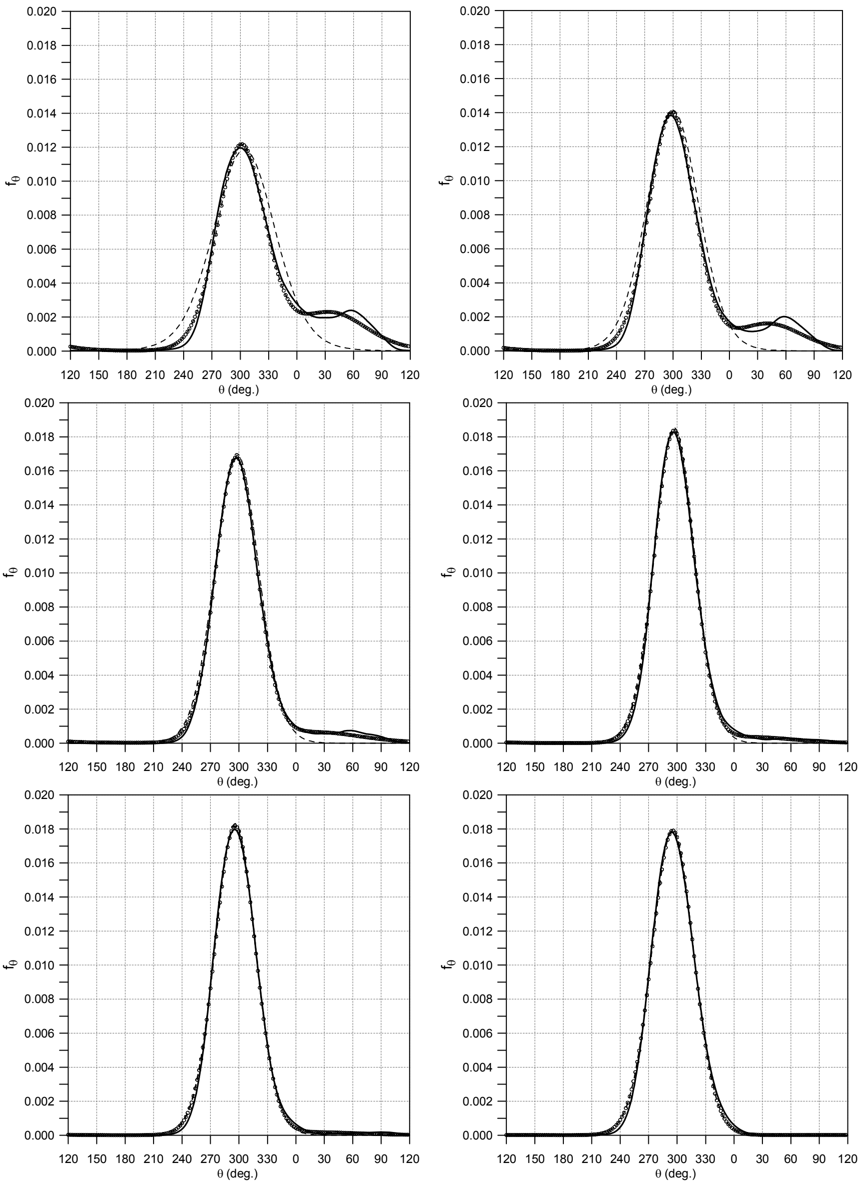


Figure 4: Empirical probability distribution of mean wave direction (solid line), the VM (dashed line) and two VM mixture models fitted (dotted line) for $H_{m0}\geq1$m (up-left), $H_{m0}\geq2$m (up-right), $H_{m0}\geq3$m (middle-left), $H_{m0}\geq4$m (middle-right), $H_{m0}\geq5$m (down-left), and $H_{m0}\geq6$m (down-right) thresholds.

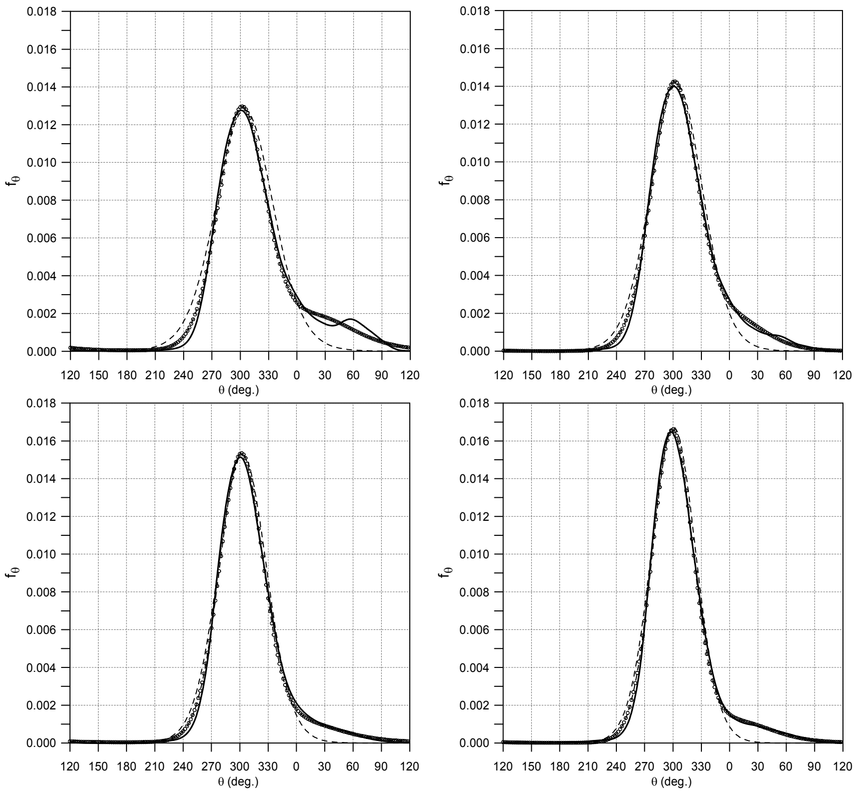


Figure 5: Empirical probability distribution of mean wave direction (solid line), the VM (dashed line) and two VM mixture models fitted (dotted line) for $T_z \geq 6$ s (up-left), $T_z \geq 8$ s (up-right), $T_z \geq 10$ s (down-left), and $T_z \geq 12$ s (down-right) thresholds.

Table 1: Number of sea states, average mean wave direction and maximum likelihood estimate of the concentration parameter for the full data set and for the data set filtered by climatic season.

	N	θ_m (^{o})	κ
Complete set	43 227	317.5	2.18
Spring	10 663	319.8	1.91
Summer	10 727	326.0	2.20
Autumn	11 726	316.3	2.08
Winter	10 111	315.8	2.00

Table 2: Number of sea states, average mean wave direction and maximum likelihood estimate of the concentration parameter for the data set filtered with different significant wave height thresholds.

	N	θ_m (°)	κ
H > 1 m.	33 782	325.5	3.85
H > 2 m.	15 840	319.6	5.73
H > 3 m.	6 219	316.1	10.93
H > 4 m.	2 264	316.7	12.83
H > 5 m.	837	315.5	14.37
H > 6 m.	218	314.6	16.46

Table 3: Number of sea states, average mean wave direction, and maximum likelihood estimate of the concentration parameter for the data set filtered with different spectral peak period thresholds.

	N	θ_m (°)	κ
Ts > 6 s.	39 854	313.7	2.58
Ts > 8 s.	33 004	309.5	3.50
Ts > 10 s.	20 351	307.7	3.98
Ts > 12 s.	9 646	305.4	4.29

5 Concluding remarks

The study reveals that a single VM distribution is not adequate in general to characterize the probabilistic structure of the mean wave direction in the study site, even when considering the various climatic seasons independently. However, this model becomes useful when examining mean wave directions associated to sea states with moderate and large periods, removing low period sea states, or for the more severe wave conditions, in terms of significant wave heights.

The use of a mixture of two VM models significantly improves the degree of fitness in all the cases, especially when the empirical distribution presents a bimodal character, or when is unimodal but significantly skewed, such as in the case of the full data set, the seasonal distributions, or when severe sea states are considered.

The present analysis is site specific and no attempt has been made to draw general conclusions for wider sea areas. However, the used methodology is of wide application. Furthermore, results derived from this study should help the development of joint distributions of the three parameters considered to characterise long-term wave climate.

Acknowledgement

The authors are grateful to Puertos del Estado, Ministerio de Fomento, Spain, for providing the data used in this study.

References

[1] Silverman, B.W. 1986. *Density Estimation for Statistics and Data Analysis*. London: Chapman and Hall.

[2] Fisher, N.I. 1993. *Statistical Analysis of Circular Data*. New York: Cambridge University Press.

[3] Vega J.L and G. Rodriguez, 2007. Modelling long term distribution of mean wave direction, *Proc. of the 12th Int. Conf. of the International Maritime Association of the Mediterranean*, Eds. Guedes Soares and Kolev, Balkema, 839-846.

[4] Mardia, K.V. and Jupp, P.E., 1999. *Directional Statistics*. Wiley, Chichester.

[5] Mooney, J. A., P.J. Helms, and I.T. Jolliffe, 2003. Fitting mixtures of von Mises distributions: a case study involving sudden infant death syndrome. *Computational Statistics and Data Analysis*, **41**: 505-513.

[6] Spurr, B.D. and M. A. Koutbeiy, 1991. A comparison of various methods for estimating the parameters in mixtures of von Mises distributions. *Communications in Statistics*. **20**: 725-741.