Journal de la Société Française de Statistique
Vol. 156 No. 1 (2015)

Environmental conditions in semi-enclosed basins: A dynamic latent class approach for mixed-type multivariate variables

Titre: Conditions environnementales dans les bassins semi-fermés : Une approche dynamique de structure latente pour les variables multivariées de type mixte

Jan Bulla[1,2], Francesco Lagona[3], Antonello Maruotti[4] and Marco Picone[5]

Abstract: The identification of typical environmental conditions from multiple time series of linear and circular observations requires classification methods that account for the dependence across variables and in time. Motivated by a case study of sea conditions, we take a latent-class approach to classification, relying on a multivariate hidden Markov model. The model integrates multivariate von Mises and log-normal densities to describe the distribution that wind speed and wave height as well as wind and wave direction take under different latent regimes, with parameters that depend on the evolution of an unobserved Markov chain. The estimation of the model is facilitated by a hybrid algorithm that combines an EM algorithm with direct maximization of the log-likelihood.
Our analysis of marine data from two locations in the Mediterranean shows that a hidden Markov approach to classification can be successfully employed for identifying interpretable marine conditions in complex orographic settings.

Résumé : L'identification des conditions environnementales typiques basée sur plusieurs séries chronologiques d'observations linéaires et circulaires nécessite des méthodes de classification qui tiennent compte des dépendances aussi bien temporelles que dans l'ensemble des variables. Motivé par une étude de cas centré autour de caractéristiques marines, nous adoptons une analyse de structure latente comme classification, en s'appuyant sur un modèle de Markov caché multivarié. Le modèle intègre des densités multivariées von Mises et log-normales pour décrire la distribution que la vitesse du vent et la hauteur des vagues ainsi que la direction du vent et des vagues prennent sous différents régimes latents, avec des paramètres qui dépendent de l'évolution d'une chaîne de Markov non observée. L'estimation du modèle est facilitée par un algorithme hybride qui combine un algorithme EM avec maximisation directe de la log-vraisemblance.
Notre analyse des données marines de deux régions de la Méditerranée montre qu'une approche de Markov caché comme classification peut être utilisée avec succès pour identifier les conditions marines interprétables dans les milieux orographiques complexes.

Keywords: Circular data, EM algorithm, Hidden Markov model, Model-based clustering, Skewness, Unsupervised classification, Wave, Wind

AMS 2000 subject classifications: 62P12, 62M05, 91C20

[1] Department of Mathematics, University of Bergen, Bergen, Norway, E-mail: jan.bulla@math.uib.no
[2] LMNO, CNRS UMR 6139, Université de Caen, Caen, France
[3] DPS, Università Roma Tre, Rome, Italy
[4] Southampton Statistical Sciences Research Institute & School of Mathematics, University of Southampton, Southampton, Great Britain
[5] Marine Service, Institute for Environmental Protection and Research, Rome, Italy

 ISSN: 2102-6238

1. Introduction

In the past decades, hidden Markov models (HMMs), introduced in the second half of the sixties Baum and Petrie (1966) Baum and Eagon (1967), have found widespread application in the analysis of environmental phenomena. Details can be found, e.g., in Zucchini and Guttorp (1991), Hughes and Guttorp (1994), Hughes et al. (1999), Bellone et al. (2000), Robertson et al. (2004), Lagona (2005), Ailliot et al. (2009), Betró et al. (2008), Lagona et al. (2011). In an environmental context, HMMs assume the existence of weather states, which are identified as latent classes when fitting the model.
Most of the existing approaches, however, have been specified in a univariate setting, where environmental variables are modelled separately or, in a multivariate setting, by assuming conditional independence given the weather state (see, e.g., Holzmann et al., 2006), with few exceptions Lagona and Picone (2013), Lagona and Picone (2012), Bulla et al. (2012). The advantage of using a multivariate HMM approach for environmental data modelling consists in the ability to capture complex data correlation structures and to decompose them into interpretable relations. In the following, a multivariate HMM serves for studying air-sea interactions between mixed-type (linear-circular) environmental variables. We specify a multivariate HMM in the spirit of Bulla et al. (2012) for multivariate wind/wave data where wind and wave directions as well as wind speed and wave heights are recorded at multiple buoys over a period of one year. This work focuses on the detection of time-stable environmental conditions in areas of basin and semi-basin type, respectively, in the Mediterranean Sea. We were driven by the assumption that the wind-wave data analyzed possesses unobserved characteristics, most notably being generated by distinct weather conditions. We accordingly propose a dynamic latent variable model in an HMM framework for the analysis of wind-wave measurements. More specifically, we use a mixture of multivariate densities, given by the product of bivariate circular and log-normal distributions, and assume that the parameters of the mixture evolve in time according to an unobserved (hidden) Markov chain. This assumption allows clustering measurements within toroidal groups (when circular measurements are considered) and skewed ellipses (when linear measurements are modelled).
We cluster the circular and linear data part of the data separately to avoid the definition of hypercylindrical clusters (which would be obtained without the previous assumption), as such clusters may lead to highly non-interpretable results. Nevertheless, we would remark that toroidal and skew-elliptical clusters are not independent, just conditionally independent, as they are paired according to the weather states of the Markov chain. The inclusion of a dynamic latent process (described by the Markov chain) has several appealing features. The transition probability matrix of the Markov chain captures weather-switching in time, accounting for temporal autocorrelation. Consequently, the classification is not only based on similarities in the variables space, but also on similarities that occur in a temporal neighbourhood.

The paper is outlined as follows: We first discuss the environmental data in Section 2. Then, we present our modelling approach in Section 3, including the formal specification of bivariate circular and linear distributions along with the estimation procedure and a brief discussion on model selection. Subsequently, we present our results in Section 4 and close with a discussion in Section 5.

 ISSN: 2102-6238

2. Wind Wave Data

The data motivating this work are time series of hourly wave and wind directions as well as wind speeds and significant wave heights, recorded in 2011 by the Italian data buoy network (Rete Ondametrica Nazionale, RON), owned and managed by the Institute for Environmental Protection and Research (Istituto Superiore per la Protezione e la Ricerca Ambientale, ISPRA). This network was completely redesigned in 2009 to acquire oceanographic variables and to increase nautical capabilities, reducing accidental buoy adrift events.

In particular, we examine two different cases, which are characterized by specific features that need to be properly accounted for. The first time series of environmental measurement were recorded at Ancona, where the buoy is located in the Adriatic Sea about 30 km from the coast. The second data set was generated at Civitavecchia, where the buoy is in the Tyrrhenean Sea about 5 km from the coast (see Fig. 1)

The environmental situation at these two locations is considerably different. Ancona is located at the Adriatic Sea, which is a semi-enclosed basin. This area is subject to three relevant wind events: Bora, Mistral and Sirocco winds (see roses in Fig. 2). Sirocco arises from a warm, dry, tropical air mass that is pulled northward by low-pressure cells moving eastward across the Mediterranean Sea. It typically blows in the period March-October and generates effects along the major axis of the Adriatic basin (i.e., along the southeast-northwest direction). Bora episodes occur when a polar high-pressure area sits over the snow-covered mountains of the interior plateau behind the coastal mountain range and a calm low-pressure area lies further south over the warmer Adriatic. It transfers a great amount of energy to the northern portion of the Adriatic basin essentially in wintertime. Finally, the Mistral is a sea-breeze wind blowing northwesterly when the east Adriatic coast gets warmer than the sea.
While Bora and Sirocco episodes are usually associated with high-speed flows, Mistral is in general linked with good meteorological conditions. The orography of the Adriatic Sea plays a key role and most of the waves tend to travel from north, northwest and southeast along the major axis of the basin, where they can travel freely, without being obstructed by physical obstacles, such as coastlines. In wintertime (November to March) the dominant winds are Mistral and Bora. The first one generates low waves that travel along the Italian coast from northwest to southeast. The Bora wind generates high waves coming from north, reaching heights up to five meters. In summertime (April to October) the dominant winds are Mistral and Sirocco with waves never higher than three meters.

Civitavecchia is characterized by different geographical features. The Tyrrhenian Sea is bounded by the Sardinia and the Corsica islands (west), the Tuscan Archipelago (north), the coast of the Italian peninsula (east) and partially by Sicily (south). However, the Sardinian Channel, the Strait of Bonifacio, and the Strait of Messina allow water exchange with the Mediterranean Sea. Again, three relevant wind events characterize this area: Tramontane, Sirocco and Libeccio winds (see roses in Fig. 3). Tramontane is a northeasterly or northerly wind, which is prominent on the west coast of Italy, typically in winter. It is a fresh wind of the fine-weather mistral type, it does not often reach gale force, and it is associated with a depression over the Adriatic that acts simultaneously with an anticyclone further west. The Libeccio is the westerly or southwesterly wind, which predominates central Tyrrhenian sea all the year round; it may frequently give violent westerly squalls. It is persistent during summer, while it alternates with the Tramontane in

 ISSN: 2102-6238

The two locations where data were recorded are Ancona in the north-east and Civitavecchia in the southwest

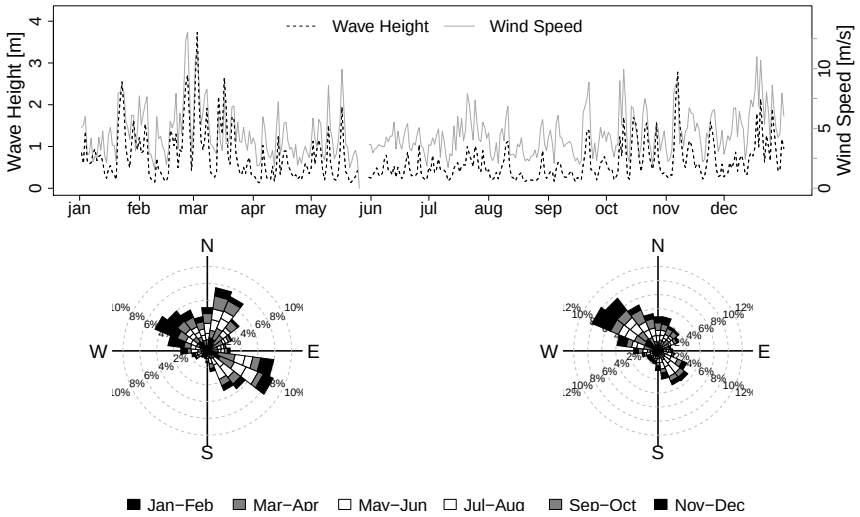
FIGURE 1. *Map of Italy*

wintertime.

Even in this case the orography plays a key role. The Tramontane, for example, has very low effects on waves because of the Tuscan archipelago north of Civitavecchia that narrows the formation of the waves. On the other hand, the Libeccio frequently raises high seas blowing through the Sardinia Channel and the Strait of Bonifacio.

The identification of relevant regimes in a particular area is usually exploited by using numerical models, such as WAM (WAve Models), introduced in the late 1980s Hasselmann et al. (1988). Several works Christopoulos (1997), Ponce de Leòn and Guedes Soares (2008) highlight that the Mediterranean Basin has a unique character, resulting from geographic conditions. It is a semi-enclosed basin surrounded by a complex orography, which strongly affects and often controls the local climate and the wind/wave conditions in particular. This results in many interactions and feedback among marine-atmosphere-land processes, which make the Mediterranean area a unique

 ISSN: 2102-6238



The top panel shows wave height and wind speed. The lower panel displays wind direction (left) and wave direction (right). For the directional plots, the periods start with January/February on the inside.

FIGURE 2. *Wind and Wave date from the buoy of Ancona*

system.

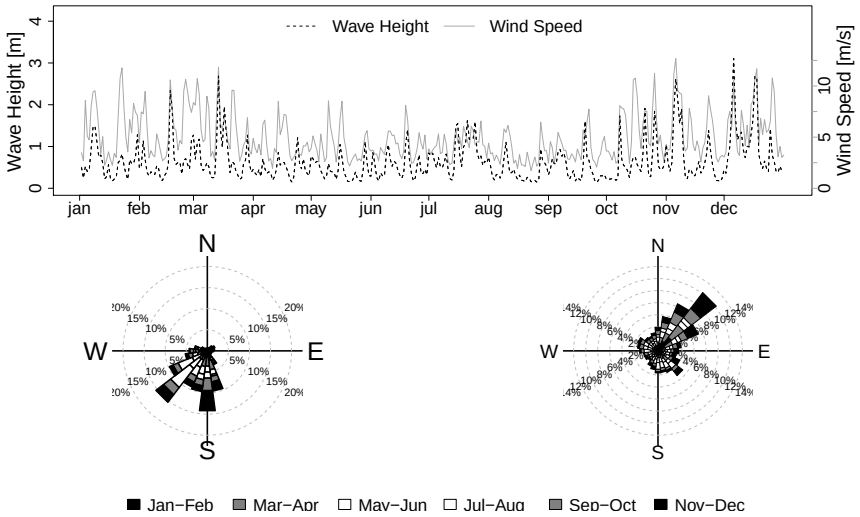
In particular, waves and winds are only partially correlated due to the orographic conditions. Therefore, the observed environmental regimes are a combination of atmospheric events, marine circulation, orography of coastal areas, and local conditions. As a result, numerical wind-wave models, traditionally used for ocean waves, are not flexible enough and provide biased results in studies of the Adriatic (Bertotti and Cavaleri, 2009). Other statistical techniques of wave data clustering are based on distance-based methods (Boukhanovsky et al., 2007) or hierarchical agglomerative clustering methods (Hamilton, 2010). The statistical properties of these methods are generally unknown, precluding the possibility of formal inference on the clustering results (Fraley and Raftery, 2002). An interesting alternative could be provided by state-space models. Ailliot et al. (2011) consider such a framework (strictly related to our approach) to capture wind effects in generating waves and to account for heterogeneity leading to different sea-states.

3. Methods

3.1. Model

For describing our model, we use the following notation. Let $\{\mathbf{z}_t\}_{t=1,2,\ldots,T} = \{\mathbf{x}_t, \mathbf{y}_t\}_{t=1,2,\ldots,T}$ be a multivariate time series with bivariate circular, $\mathbf{x}_t = (x_{1t}, x_{2t}) \in [0, 2\pi]^2$, and linear, $\mathbf{y}_t = (y_{1t}, y_{2t}) \in \mathbb{R}^2$, components. Thus, in our context $\mathbf{x}_t$ corresponds to wave and wind direction, while $\mathbf{y}_t$ includes wave height and wind speed. Similarly, let S_t be an unobserved latent variable representing the weather state, and taking values from 1 to K, for $t = 1, 2, \ldots, T$. Sequences of circular and linear variables and hidden weather state are denoted by $\mathbf{x}_{1:T}, \mathbf{y}_{1:T}, S_{1:T}$. Finally, let $f(\mathbf{z}_t \mid S_{1:t}, \mathbf{z}_{1:t-1})$ be the conditional distribution of $\mathbf{z}_t$, given the latent and the observed process.

 ISSN: 2102-6238



The top panel shows wave height and wind speed. The lower panel displays wave direction (left) and wind direction (right). For the directional plots, the periods start with January/February on the inside.

FIGURE 3. *Wind and Wave date from buoy of Civitavecchia*

Furthermore, let $f(S_t \mid S_{1:t-1})$ be the conditional distribution of S_t given the latent process. By imposing

$$f(\mathbf{z}_t \mid S_{1:T}, \mathbf{z}_{1:t-1}) = f(\mathbf{z}_t \mid S_t) \tag{1}$$
$$f(S_t \mid S_{1:t-1}) = f(S_t \mid S_{t-1}) \tag{2}$$

this time series becomes a multivariate HMM. Equation (1) states that for fixed t the observed process $\mathbf{z}_t$ is conditionally independent of the other elements of the time series given the current weather state. In other words, all the temporal dependence in environmental conditions is captured by the persistence in the weather state described by equation (2). Furthermore, equation (2) states that, given the history of the weather state up to time $t-1$, the weather state at time t depends only on the previous weather state. Of course, several extensions of (2) can be introduced (see e.g. Bartolucci and Farcomeni, 2010).

3.2. *Parameterization for the environmental measures*

The bivariate von Mises distribution (Singh et al., 2002) is adopted for the circular components $\mathbf{x}_t$, i.e. wind and wave directions, while a bivariate log-normal distribution is employed to model wind speeds and wave heights. Instead of the selected distributions, a wrapped normal (Lasinio et al., 2012) and a Gamma distribution (Lagona and Picone, 2011) can be adopted for the circular and linear components, respectively. Loosely speaking, conditional on the current weather state, the environmental measurements are modelled as bivariate von Mises and log-normal distributions with state-specific parameters. Furthermore, circular and linear components are assumed conditionally independent, given the weather state.

 ISSN: 2102-6238

Formally, the resulting parameterization for the circular components is

$$f(\mathbf{x}_t;\beta) = \frac{\exp(\beta_{11}\cos(x_{1t}-\beta_1)+\beta_{22}\cos(x_{2t}-\beta_2)+\beta_{12}\sin(x_{1t}-\beta_1)\sin(x_{2t}-\beta_2))}{C(\beta)} \tag{3}$$

where β indicates the parameter vector characterizing the location and the shape of a toroidal cluster and $C(\beta)$ is a normalizing constant. The density in (3) can be viewed as a bivariate generalization of the von Mises distribution, where β_{12} accounts for the statistical dependence between x_{1t} and x_{2t}.

For the linear part, we specify a bivariate log-normal density by

$$f(\log(\mathbf{y}_t);\gamma) = \frac{1}{2\pi|\Sigma|} exp\left(-\frac{1}{2}(\mathbf{y}_t-\mu)^t\Sigma^{-1}(\mathbf{y}_t-\mu)\right), \tag{4}$$

where

$$\mu = \begin{pmatrix}\gamma_1\\ \gamma_2\end{pmatrix} \text{ and } \Sigma = \begin{pmatrix}\gamma_{11} & \gamma_{12}\\ \gamma_{21} & \gamma_{22}\end{pmatrix}.$$

3.3. *Parameterization of the weather state process*

Let $q_{tjk} = f(S_t = k \mid S_{t-1} = j)$ for $j,k = 1,2,\ldots,K$ be the probability that state k is visited at time t given that at time $t-1$ the j-th weather state was visited. The simplest specification in this framework is the homogeneous Markov chain, which assumes constant (over time) transition probabilities, that is

$$q_{tjk} = q_{jk} \tag{5}$$

where $q_{jk} \in [0,1]$ for all $j,k = 1,2,\ldots,K$ and $\sum_{k=1}^{K} q_{jk} = 1$ for all $j = 1,2,\ldots,K$. One may note that this setting could be easily extended to a non-homogeneous model, e.g., by taking exogenous (atmospheric) variables into account. Then, the transition probabilities of the hidden Markov chain could be parameterized as functions of the exogenous covariates v_t by

$$q_{tjk} = \frac{\exp(c_{jk0}+v_t'c_{jk})}{1+\sum_{h:h\neq j}^{K}\exp(c_{jh0}+v_t'\mathbf{c}_{jh})}, \tag{6}$$

where $\mathbf{c}_{jk}$ is a vector of fixed regression parameters (of course, $\mathbf{c}_{jk} = \mathbf{0}$ and $c_{jk0} = 0$ if $j = k$). This model specification, also known as non-homogeneous hidden Markov model, is usually adopted when the interest is to gain understanding of the evolution of weather conditions which are affected by the exogenous covariates. Other specifications can be used (see e.g. Maruotti and Rocci, 2012). Unfortunately, we do not have any (reliable) covariate available. Thus, throughout the paper, we consider parameterization of type (5) to model the dynamics of the hidden process.

To complete the specification of the hidden process, we need to define the initial probabilities $\{\delta_k\}_{k=1,\ldots,K}$, where $\delta_k = f(S_1 = k)$, i.e., the probability of being in state k at time 1. These quantities may be parameterized via a logit-type specification close to the one in Equation (6) as well. However, in this study we do not estimate them separately from the transition probability matrix, but set them to be the stationary distribution of this matrix, following, e.g. Bellone et al. (2000).

 ISSN: 2102-6238

3.4. Parameter estimation

Let θ be the vector of the unknown model parameters, the likelihood of the observed data is given by

$$\begin{aligned} L(\theta) &= f(\mathbf{z}_{1:T};\theta) = \sum_{S_1,\ldots,S_T} f(S_{1:T};\theta) f(\mathbf{z}_{1:T} \mid S_{1:T};\theta) \\ &= \sum_{S_1,\ldots,S_T} f(\mathbf{z}_1 \mid S_1) f(S_1) \prod_{t=2}^{T} f(S_t \mid S_{t-1}) f(\mathbf{z}_t \mid S_t). \end{aligned} \tag{7}$$

In environmental data modelling, the number of observation time points T is usually large and computation of the (log-)likelihood is intensive. Using the statistical software `R` (R Development Core Team, 2013), we maximize the log-likelihood via a hybrid algorithm: First, we executed several iterations of the EM algorithm, and then we switched to direct numerical maximization via quasi-Newton methods. This approach has the advantage of combining stability of the EM algorithm with high convergence speed of the quasi-Newton approach in the neighbourhood of a maximum (Bulla and Berzel, 2008). More precisely, we started the EM algorithm from 20 random initial parameter values and stopped after 50 iterations. Subsequently, the resulting parameters served for initializing a direct maximization of the likelihood via the function `nlm()`. This maximization procedure was stopped when the relative increase of the log-likelihood in two successive steps fell below 10^{-6}. This stopping criterion was determined in preliminary experiments to ensure that the estimated parameters are sufficiently stable to conclude that a maximum has been reached. The technique worked well: We observed that direct maximization of the log-likelihood is numerically stable and rapid when initial parameters are in the close neighbourhood of a maximum. Moreover, in most cases the same maximum was reached, independent of the initial values.
From a more technical perspective, one needs to take care of some computational problems in order to maximize the likelihood with respect to the parameters. For example, numerical underflow has to be avoided and the model needs to be re-parameterized in terms of unconstrained parameters when using unconstrained maximization algorithms. For details on how to deal with these problems see, e.g., Chapter 3 of Zucchini and MacDonald (2009). Details on computational aspects for model fitting of multivariate HMM with linear-circular components, in particular detailed steps of the EM algorithm, can be found in Bulla et al. (2012).

3.5. Model order

Fitting HMMs to environmental data involves the choice of a model order, i.e., the number of weather states. However, while the estimation of the parameters of an HMM has been studied extensively, the consistent estimation of the number of hidden states is still an unsolved problem. The AIC and BIC methods are used most commonly, but their application in this context has not been justified theoretically. Therefore, we use these criteria mainly as indicator to avoid over-fitting. More importantly, we consider the interpretability of the resulting weather states as main criterion. This concerns in particular the situation where our model is fitted to the full data

 ISSN: 2102-6238

set, where AIC and BIC would allow justifying a significantly higher number of states, which, however, provide only little additional insight. Thus, the algorithm is defined for a fixed number of states K and reaches a K-based solution which can be successively used to estimate model parameters as the number of components is increased to $K+1$. Various authors have discussed alternative algorithms for joint estimation of K and model parameters, such as VEM or VDM (Böhning, 2000). However, a possible solution is to update estimates for a fixed K and improving step by step (as in EMGFU; Böhning, 2003). We follow the latter approach by evaluating the obtained results for varying K-based model.

4. Results

For each buoy, the environmental dataset has been split into six bimesters, and for each bimester we fit a different multivariate HMM following the strategy described in the previous section. We decided to split the data into several shorter samples to conveniently capture seasonal differences within the dominant weather states. A preliminary analysis indicated that fitting the model to the full sample is a less attractive alternative, because this requires a much higher number of states which basically result from slightly different 'sub-states' of the main weather states and occur only during specific periods of the year. The choice of the period length of two months results from a compromise: on the one hand, the data available has to be of sufficient length for stable estimation results of our relatively complex model. On the other hand, the periods have to be sufficiently short to account for intra-annual heterogeneity patterns.
Four- and three-state models have been selected for fitting the data from Ancona and Civitavecchia, respectively. This choice is driven by physical interpretability, which is one of the major requirements of an empirical work. The different numbers of states for the two locations result from different environmental characteristics of the data at the two buoys. Two different environmental settings can be inferred from the fitted models, and they seem to be closely related to the orography of the Adriatic and the Tyrrenean Sea, respectively. In the Adriatic, waves origin from two main directions (northwest and southeast), whilst at Civitavecchia the prevailing direction of origin of the waves is only one (south to southwest). In the models finally selected, weather states correspond to well-recognized environmental conditions in terms of sea-states and wind-fields.

4.1. Weather states at Ancona

At Ancona, the following weather conditions can be recognized see Figure 4 for the marginal distributions and Table 3 for parameter estimates): State 1 is characterized by high wave heights in the same direction as the wind. This environmental condition is mainly due to Sirocco, which – blowing south/southeast – generates waves on the same axis of the basin. State 2 identifies Bora winds (from north to northeast) which correspond to waves from the north. Waves are very high due to the high energy of the wind, which typically blows in the form of powerful bursts. State 3 identifies to Mistral winds (north to northwest), which generate modest waves from the north. At last, State 4 corresponds to calm sea conditions. All the combinations of winds, not capable of transferring large amounts of energy to the sea, are clustered in such a State. These conditions are common, e.g., when wind blows from land.

 ISSN: 2102-6238

Looking at estimated transitions, states are extremely persistent, i.e., essentially diagonal transition probability matrices are estimated (Table 1). We would like to point out that Sirocco (State 1) and Mistral (State 2), linked to waves from southeast and northwest, are distributed throughout the year. On the contrary, Bora, generating waves from the north (State 3), occurs mainly in autumn and winter and is substituted by calm wind-sea conditions (State 4) during spring and summer. A dynamic evolution of weather states can be also discussed. Once the State 4 is visited, State 1 and State 3 are most likely to follow. This condition is well reflected in the observed and well-known alternations between cyclones and anticyclones over central and southern Europe, which affect weather conditions on the Adriatic Sea. Cyclonic events generally proceed in the direction west-east (State 4) and may lead to winds from the southern quadrants (Sirocco, State 1). Persistence of such weather conditions yields higher wave heights. When cold temperatures come from northern Europe into the semi-basin, strong Bora wind conditions (State 2) arise as anticyclones up to stabilize the weather with mistral events. This phenomenon is less evident in summertime when there is an alternation of winds (usually weak) from the south (Sirocco) and north (Mistral). The following Figures 5 and 6 report the classification of data in the second and sixth bimester for the buoy of Ancona. Comparing all bimesters, these two selected periods show the two most distinct situations in terms of heterogeneity.

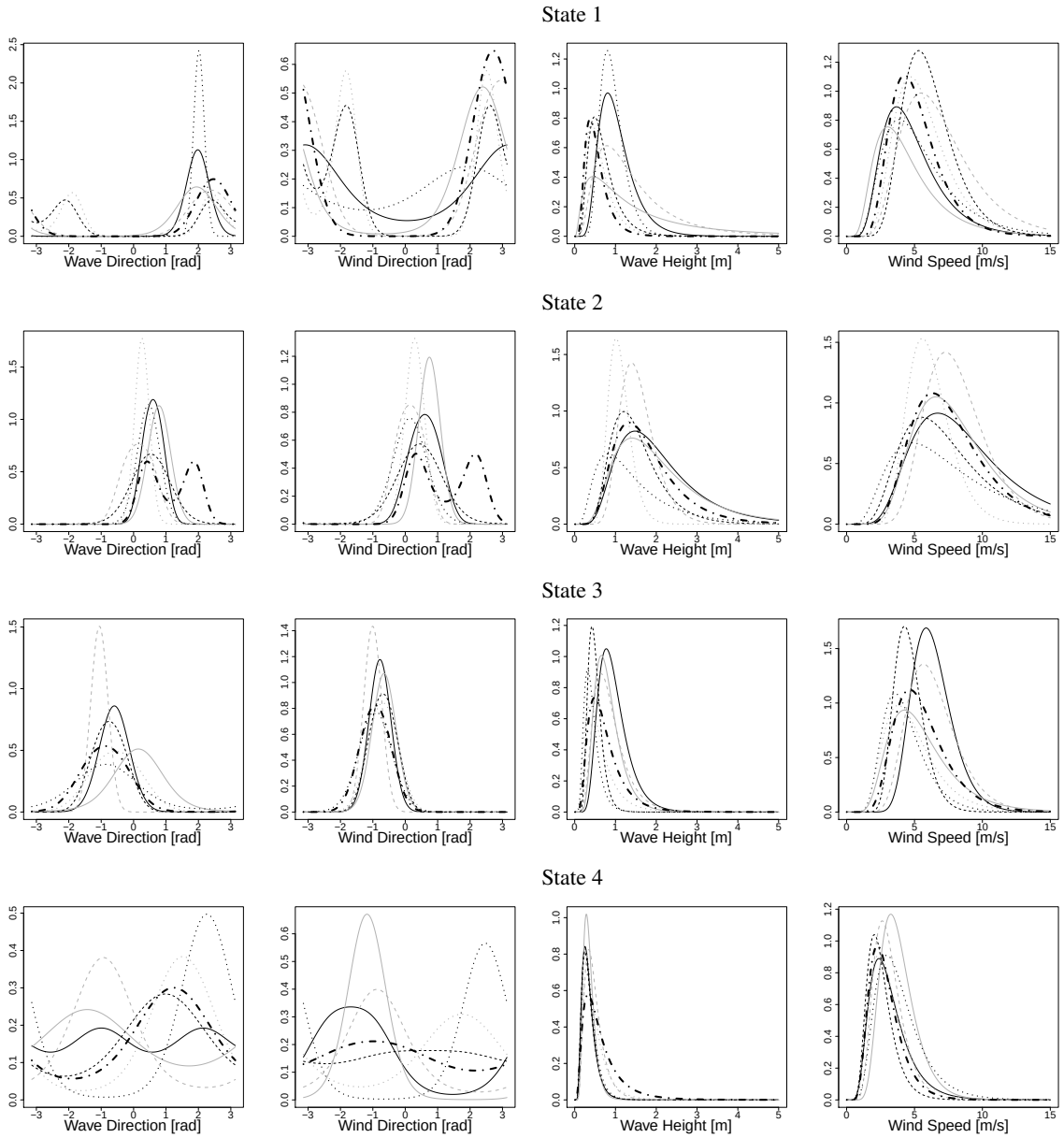
As to comparing the annual distributions in Figure 4 to those from the different bimonthly periods, one can observe the following: The weather phenomena identified on annual and bimonthly basis, respectively, seem not too different in general. An exception is the second State 2, which captures Bora conditions for the bimonthly data, but also captures some of the stronger southeastern Sirocco events on annual level. Apart from this, one may note that the seasonal distributions have higher concentrations and lower variances, respectively, than the annual distribution. This is a direct consequence of the slightly varying inter-period characteristics present in the four states, and teir period-varying proportions. Naturally, bimonthly distributions are able to better capture the variability of the directional and linear data in time. For example, Sirocco winds blow mostly from southeast in wintertime, but predominantly from south during summertime. As a result, the waves generated by Sirocco winds come from southeast in wintertime and from a wide sector from southeast to southwest during summertime. As expected, winds blow stronger in wintertime than in summertime, resulting also in higher waves during winter. Therefore, the marginal distributions of both linear variables are approximately centred at zero for summertime periods (see Table 4). Last but not least, all states show a high persistence with all elements of the main diagonal being greater than 0.9 (see Table 2).

4.2. Weather states at Civitavecchia

A similar analysis can be provided on the data from Civitavecchia (see Figure 7 for the marginal distributions and Table 7). For each bimester, State 1 corresponds to north winds (north and northwest). As these winds blow very strongly from the mainland, they do not transfer a large amount of energy to the sea, and, as a result, we observe waves of low height. State 2 identifies Libeccio, which blows from southwest along the main axis of the basin. It generates high waves resulting from winds occurring in bursts. Finally, State 3 identifies calm sea and south winds (with a modal south-southeast direction).

The estimated transition probability matrices (Table 5) indicate that persistence in single

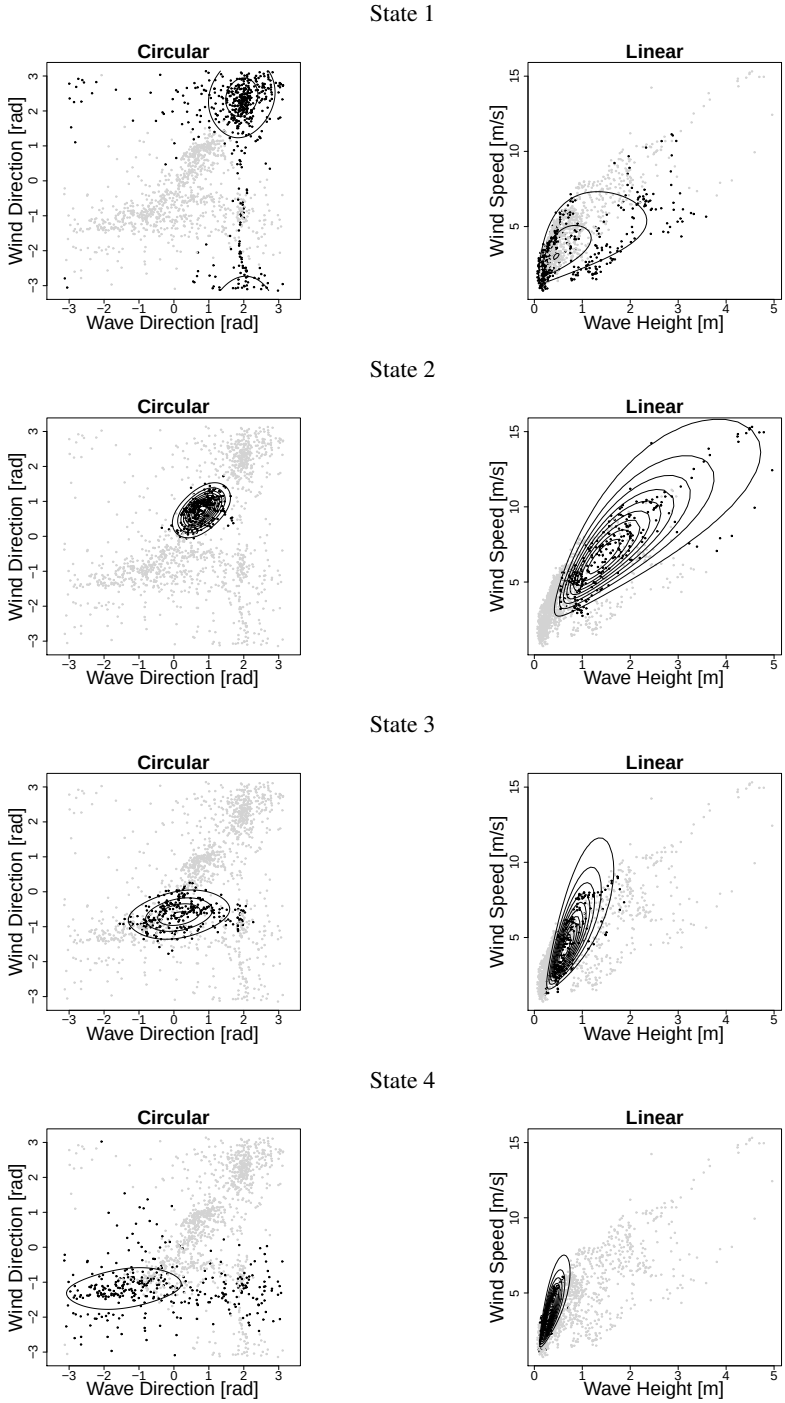
 ISSN: 2102-6238



The figure displays the marginal distributions of the models fitted to annual data (dot-dashed solid black lines) and bimonthly data (black solid Jan/Feb, grey solid Mar/Apr, black dotted May/Jun, grey dotted Jul/Aug, black dashed Sep/Oct, grey dashed Nov/Dec) at Ancona.

FIGURE 4. *Univariate (marginal) distribution at Ancona*

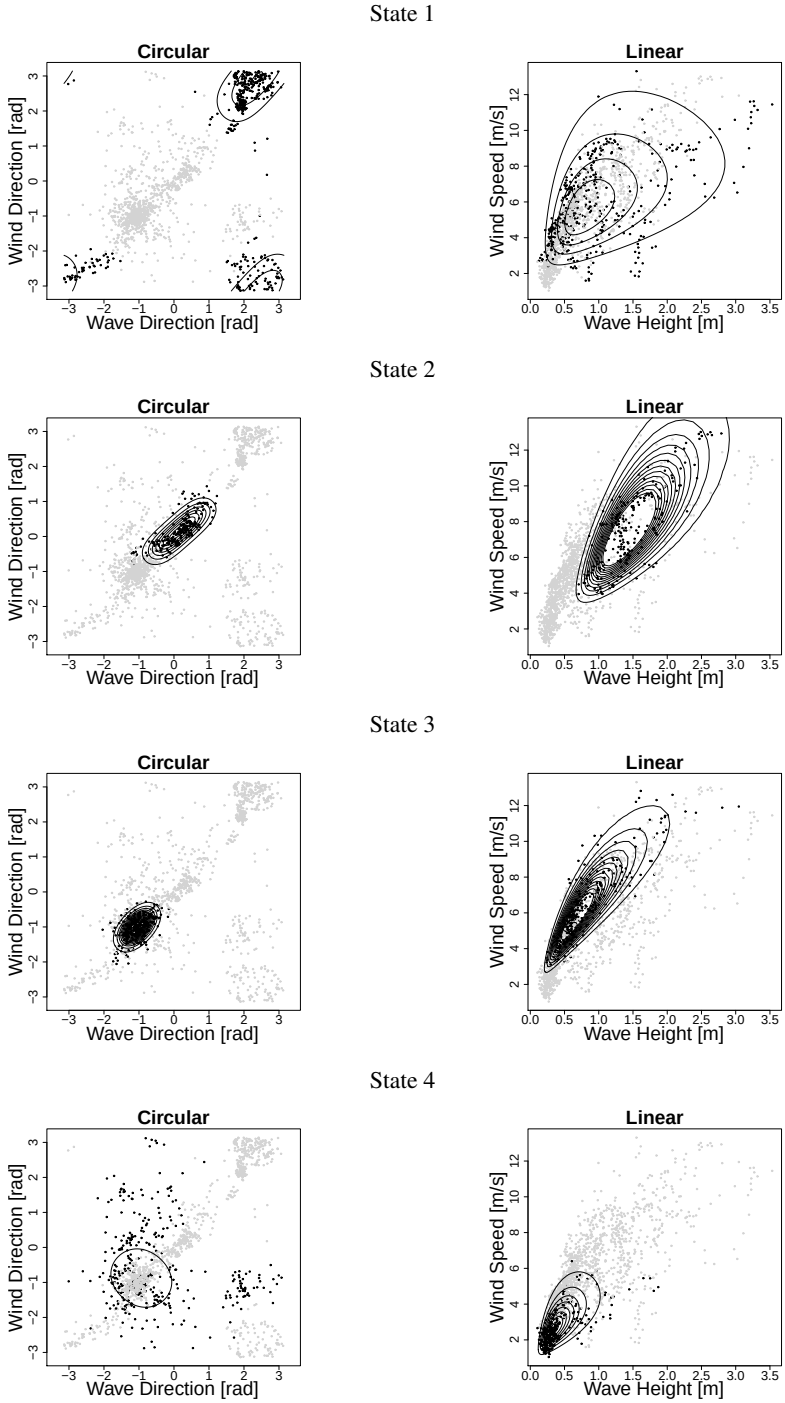
 ISSN: 2102-6238



The figure displays the conditional distributions of the fitted model at Ancona in the bimester March/April. The circular components are on the left, and the linear components on the right, starting with the first state from the top. Observations classified to the respective state by local decoding are coloured black, while the remaining observations are gray.

FIGURE 5. *Conditional distributions at Ancona in March/April*

 ISSN: 2102-6238



The figure displays the conditional distributions of the fitted model at Ancona in the bimester November/December. The circular components are on the left, and the linear components on the right, starting with the first state from the top. Observations classified to the respective state by local decoding are coloured black, while the remaining observations are gray.

FIGURE 6. *Conditional distributions at Ancona in November/December*

 ISSN: 2102-6238

TABLE 1. *Estimated transition probability matrices for Ancona, bimester data*

The table summarizes the estimated transition probability matrices for Ancona. Standard errors, reported in parentheses below each estimated, result from parametric bootstrap with 200 repetitions. Rows and columns, respectively, indicate the origin and destination states.

	Jan-Feb				Mar-Apr				May-Jun			
State	1	2	3	4	1	2	3	4	1	2	3	4
1	0.973	0.013	0.007	0.007	0.968	0.007	0.008	0.018	0.936	0.026	0.02	0.018
	(0.012)	(0.01)	(0.005)	(0.005)	(0.008)	(0.003)	(0.004)	(0.006)	(0.029)	(0.019)	(0.012)	(0.015)
2	0.014	0.969	0.014	0.003	0.003	0.972	0.025	0	0.009	0.956	0.022	0.014
	(0.006)	(0.008)	(0.005)	(0.003)	(0.003)	(0.013)	(0.012)	(0)	(0.005)	(0.015)	(0.012)	(0.008)
3	0	0.019	0.957	0.024	0.003	0.012	0.957	0.028	0.002	0.015	0.963	0.02
	(0)	(0.01)	(0.015)	(0.01)	(0.004)	(0.008)	(0.012)	(0.01)	(0.002)	(0.005)	(0.008)	(0.006)
4	0.007	0.002	0.02	0.971	0.04	0.006	0	0.954	0.012	0.004	0.02	0.964
	(0.005)	(0.003)	(0.01)	(0.012)	(0.014)	(0.006)	(0)	(0.017)	(0.005)	(0.004)	(0.006)	(0.009)
	Jul-Aug				Sep-Oct				Nov-Dec			
State	1	2	3	4	1	2	3	4	1	2	3	4
1	0.927	0	0.031	0.043	0.949	0.012	0.023	0.016	0.971	0.002	0.014	0.013
	(0.01)	(0)	(0.009)	(0.008)	(0.011)	(0.006)	(0.009)	(0.008)	(0.009)	(0.003)	(0.008)	(0.006)
2	0	0.941	0.044	0.015	0.003	0.969	0.025	0.003	0.004	0.975	0.016	0.005
	(0)	(0.042)	(0.039)	(0.021)	(0.003)	(0.012)	(0.011)	(0.004)	(0.006)	(0.016)	(0.008)	(0.008)
3	0.004	0.005	0.937	0.054	0	0.017	0.938	0.045	0	0.008	0.951	0.041
	(0.003)	(0.003)	(0.011)	(0.011)	(0)	(0.008)	(0.011)	(0.011)	(0.003)	(0.004)	(0.011)	(0.01)
4	0.07	0.004	0.022	0.904	0.038	0	0.023	0.939	0.033	0.003	0.037	0.928
	(0.012)	(0.003)	(0.006)	(0.012)	(0.013)	(0)	(0.007)	(0.014)	(0.012)	(0.004)	(0.012)	(0.017)

TABLE 2. *Estimated transition probability matrices for Ancona, annual data*

The table summarizes the estimated transition probability matrices for Ancona. Standard errors, reported in parentheses below each estimated, result from parametric bootstrap with 200 repetitions. Rows and columns, respectively, indicate origin and destination states.

	Jan-Dec			
State	1	2	3	4
1	0.942	0.009	0.022	0.027
	(0.007)	(0.003)	(0.004)	(0.004)
2	0.001	0.960	0.017	0.022
	(0.001)	(0.005)	(0.003)	(0.004)
3	0.002	0.010	0.951	0.037
	(0.001)	(0.002)	(0.004)	(0.003)
4	0.034	0.012	0.032	0.922
	(0.004)	(0.002)	(0.004)	(0.006)

 ISSN: 2102-6238

TABLE 3. *Estimated parameters for Ancona, bimester data*

The table summarizes the estimated parameters for Ancona. The parameters β_1 - β_{12} belong to the circular component, and γ_1 - γ_{12} are the parameters of the linear component. In addition, the last column reports the initial/stationary distribution hidden Markov chain. Standard errors, reported in parentheses below each estimated, result from parametric bootstrap with 200 repetitions.

		β_1	β_2	β_{11}	β_{22}	β_{12}	γ_1	γ_2	γ_{11}	γ_{22}	γ_{12}	δ
	1	1.981	-3.088	8.258	0.881	0.028	-0.212	1.301	0.169	0.2	0.08	0.206
		(0.023)	(0.116)	(0.71)	(0.102)	(0.326)	(0.022)	(0.024)	(0.015)	(0.017)	(0.014)	(0.081)
	2	0.599	0.604	21.818	9.629	12.259	0.39	1.903	0.235	0.19	0.182	0.266
Jan		(0.021)	(0.026)	(2.02)	(0.745)	(1.285)	(0.023)	(0.02)	(0.017)	(0.013)	(0.014)	(0.095)
Feb	3	-0.595	-0.781	6.135	11.246	4.137	-0.256	1.767	0.144	0.056	0.076	0.246
		(0.021)	(0.021)	(0.481)	(0.852)	(0.592)	(0.022)	(0.012)	(0.01)	(0.004)	(0.006)	(0.068)
	4	0.575	-1.681	0	1.414	-1.48	-1.378	0.864	0.224	0.201	0.123	0.282
		(0.086)	(0.063)	(0)	(0.09)	(0.124)	(0.025)	(0.024)	(0.018)	(0.013)	(0.013)	(0.089)
	1	1.942	2.393	2.942	2.055	0.348	-0.794	1.106	0.97	0.28	0.341	0.353
		(0.033)	(0.034)	(0.166)	(0.14)	(0.166)	(0.043)	(0.023)	(0.067)	(0.02)	(0.031)	(0.075)
	2	0.787	0.745	11.066	12.311	6.316	0.339	1.883	0.275	0.144	0.166	0.211
Mar		(0.017)	(0.02)	(0.834)	(1.047)	(0.94)	(0.03)	(0.018)	(0.029)	(0.014)	(0.019)	(0.082)
Apr	3	0.146	-0.654	2.076	7.933	1.253	-0.427	1.452	0.157	0.182	0.135	0.187
		(0.048)	(0.021)	(0.209)	(0.704)	(0.373)	(0.024)	(0.025)	(0.017)	(0.018)	(0.016)	(0.06)
	4	-1.432	-1.181	0.489	3.194	0.431	-1.254	1.179	0.153	0.117	0.108	0.248
		(0.163)	(0.035)	(0.085)	(0.219)	(0.177)	(0.021)	(0.02)	(0.01)	(0.01)	(0.01)	(0.077)
	1	2.017	1.909	37.37	0.487	0.729	-0.222	1.351	0.101	0.245	0.104	0.102
		(0.017)	(0.323)	(5.976)	(0.121)	(0.832)	(0.026)	(0.049)	(0.012)	(0.027)	(0.015)	(0.039)
	2	0.463	0.145	9.001	4.129	1.678	-0.206	1.573	0.412	0.354	0.344	0.218
May		(0.021)	(0.037)	(0.662)	(0.35)	(0.446)	(0.044)	(0.044)	(0.036)	(0.028)	(0.031)	(0.073)
Jun	3	-0.858	-0.929	1.358	4.548	1.159	-1.147	1.271	0.189	0.134	0.133	0.358
		(0.048)	(0.021)	(0.069)	(0.304)	(0.184)	(0.018)	(0.015)	(0.011)	(0.008)	(0.009)	(0.065)
	4	2.263	2.468	2.107	2.622	1.16	-1.333	1.086	0.232	0.192	0.16	0.322
		(0.04)	(0.03)	(0.141)	(0.155)	(0.164)	(0.019)	(0.021)	(0.017)	(0.01)	(0.012)	(0.08)
	1	-2.89	-2.795	5.568	5.949	10.799	-0.879	1.548	0.274	0.132	0.178	0.327
		(0.015)	(0.013)	(0.366)	(0.334)	(0.537)	(0.026)	(0.018)	(0.017)	(0.008)	(0.011)	(0.052)
	2	0.259	0.292	21.145	11.982	3.893	0.015	1.718	0.059	0.068	0.043	0.047
Jul		(0.029)	(0.053)	(4.784)	(2.641)	(1.545)	(0.04)	(0.033)	(0.01)	(0.017)	(0.011)	(0.03)
Aug	3	-0.551	-0.693	2.841	7.684	3.697	-0.974	1.337	0.337	0.166	0.214	0.303
		(0.031)	(0.021)	(0.167)	(0.489)	(0.26)	(0.026)	(0.02)	(0.017)	(0.009)	(0.011)	(0.05)
	4	1.524	1.749	1.369	0.941	0.843	-1.211	0.869	0.337	0.19	0.137	0.323
		(0.054)	(0.067)	(0.11)	(0.079)	(0.108)	(0.025)	(0.017)	(0.021)	(0.013)	(0.014)	(0.041)
	1	-2.986	-2.747	4.768	4.187	7.57	-0.722	1.67	0.243	0.097	0.137	0.215
		(0.031)	(0.024)	(0.34)	(0.293)	(0.47)	(0.024)	(0.017)	(0.018)	(0.009)	(0.012)	(0.048)
	2	0.53	0.442	3.737	2.839	1.754	0.186	1.717	0.161	0.205	0.128	0.233
Sep		(0.032)	(0.042)	(0.269)	(0.26)	(0.232)	(0.023)	(0.029)	(0.014)	(0.014)	(0.01)	(0.082)
Oct	3	-0.756	-0.69	6.759	10.265	6.538	-0.859	1.449	0.112	0.055	0.064	0.276
		(0.024)	(0.021)	(0.539)	(0.761)	(0.633)	(0.016)	(0.011)	(0.008)	(0.004)	(0.005)	(0.058)
	4	0.98	0.687	0.749	0.155	0.556	-1.397	0.712	0.239	0.147	0.073	0.276
		(0.084)	(0.158)	(0.084)	(0.083)	(0.1)	(0.026)	(0.022)	(0.018)	(0.01)	(0.01)	(0.062)
	1	2.374	2.944	3.843	3.201	2.7	-0.238	1.704	0.424	0.163	0.13	0.29
		(0.031)	(0.037)	(0.232)	(0.163)	(0.264)	(0.035)	(0.022)	(0.027)	(0.013)	(0.014)	(0.082)
	2	0.149	0.149	10.844	13.412	11.005	0.332	1.985	0.078	0.079	0.06	0.157
Nov		(0.037)	(0.03)	(0.895)	(1.335)	(1.1)	(0.018)	(0.021)	(0.008)	(0.009)	(0.007)	(0.085)
Dec	3	-1.054	-1.007	19.659	17.858	10.069	-0.437	1.733	0.201	0.086	0.117	0.312
		(0.026)	(0.013)	(2.999)	(2.636)	(1.886)	(0.088)	(0.106)	(0.018)	(0.009)	(0.011)	(0.062)
	4	-0.933	-0.876	1.218	1.318	-0.204	-1.067	0.962	0.234	0.125	0.115	0.24
		(0.072)	(0.053)	(2.301)	(2.518)	(1.357)	(0.096)	(0.113)	(0.02)	(0.011)	(0.011)	(0.059)

 ISSN: 2102-6238

lt]

TABLE 4. *Estimated parameters for Ancona, annual data*

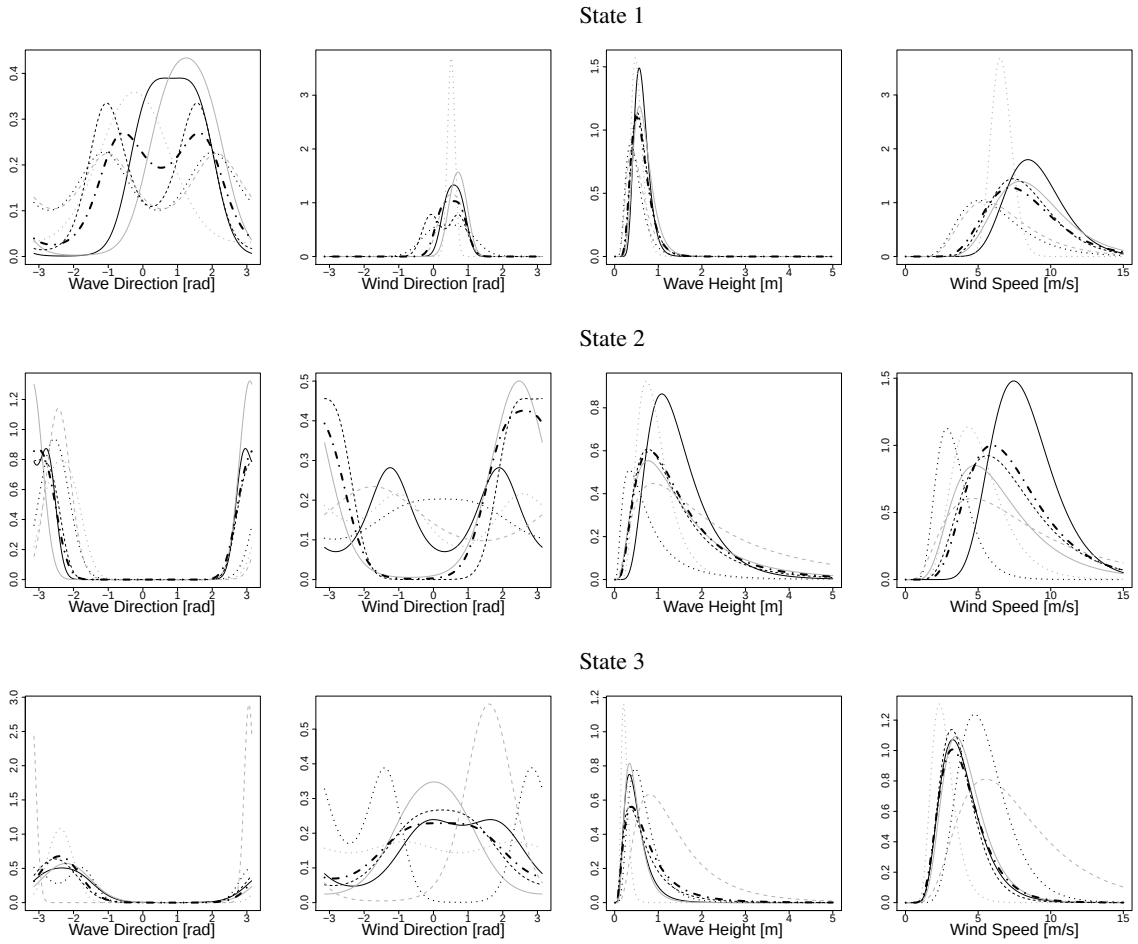
The table summarizes the estimated parameters for Ancona. The parameters β_1 - β_{12} belong to the circular component, and γ_1 - γ_{12} are the parameters of the linear component. In addition, the last column reports the initial/stationary distribution hidden Markov chain. Standard errors, reported in parentheses below each estimated, result from parametric bootstrap with 200 repetitions.

		β_1	β_2	β_{11}	β_{22}	β_{12}	γ_1	γ_2	γ_{11}	γ_{22}	γ_{12}	δ
	1	2.465	2.710	5.307	4.091	3.079	-0.978	1.451	0.254	0.129	0.162	0.179
		(0.013)	(0.016)	(0.200)	(0.148)	(0.157)	(0.014)	(0.010)	(0.009)	(0.005)	(0.006)	(0.022)
	2	1.139	1.247	9.770	5.701	11.363	0.287	1.845	0.197	0.136	0.121	0.209
Jan		(0.008)	(0.010)	(0.260)	(0.134)	(0.239)	(0.010)	(0.008)	(0.007)	(0.005)	(0.005)	(0.027)
Dec	3	-0.882	-0.922	3.199	7.144	3.620	-0.717	1.536	0.294	0.126	0.173	0.332
		(0.013)	(0.009)	(0.077)	(0.187)	(0.116)	(0.010)	(0.007)	(0.007)	(0.003)	(0.005)	(0.028)
	4	1.239	-0.982	0.832	0.346	-0.525	-1.122	0.819	0.477	0.169	0.159	0.280
		(0.033)	(0.063)	(0.034)	(0.026)	(0.042)	(0.014)	(0.009)	(0.014)	(0.005)	(0.006)	(0.019)

weather conditions is the norm. State 1, termed Tramontane, is very common in winter and is often followed by of State 2, the Liberico, especially during summer. As for Ancona, a physical interpretation of the states is possible as well. The cyclonic circulation directly affects the Tyrrhenian Sea basin producing winds from southwest (State 2) and a strong swell from the south to southeast. After these rather extreme conditions, Sirocco (State 3) with calm sea conditions follows. At the opposite, the anticyclonic circulation generates events from the north (State 1) with relatively low waves. The following Figures 8 and 9 show the data classification in the first and fifth bimesters for the buoy of Civitavecchia. Similar to Ancona, these two bimesters represent two very different situations, and are most representative for showing the heterogeneity of the six bimesters analyzed.

When comparing the results from fitting a model with three states to annual and bimonthly data at Civitavecchia, the comments provided in the previous section for Ancona are in principle the same. Tables 8 and 6 summarize the estimated parameters. However, for Civitavecchia the three states do not mix as it happens for State 2 in Ancona.

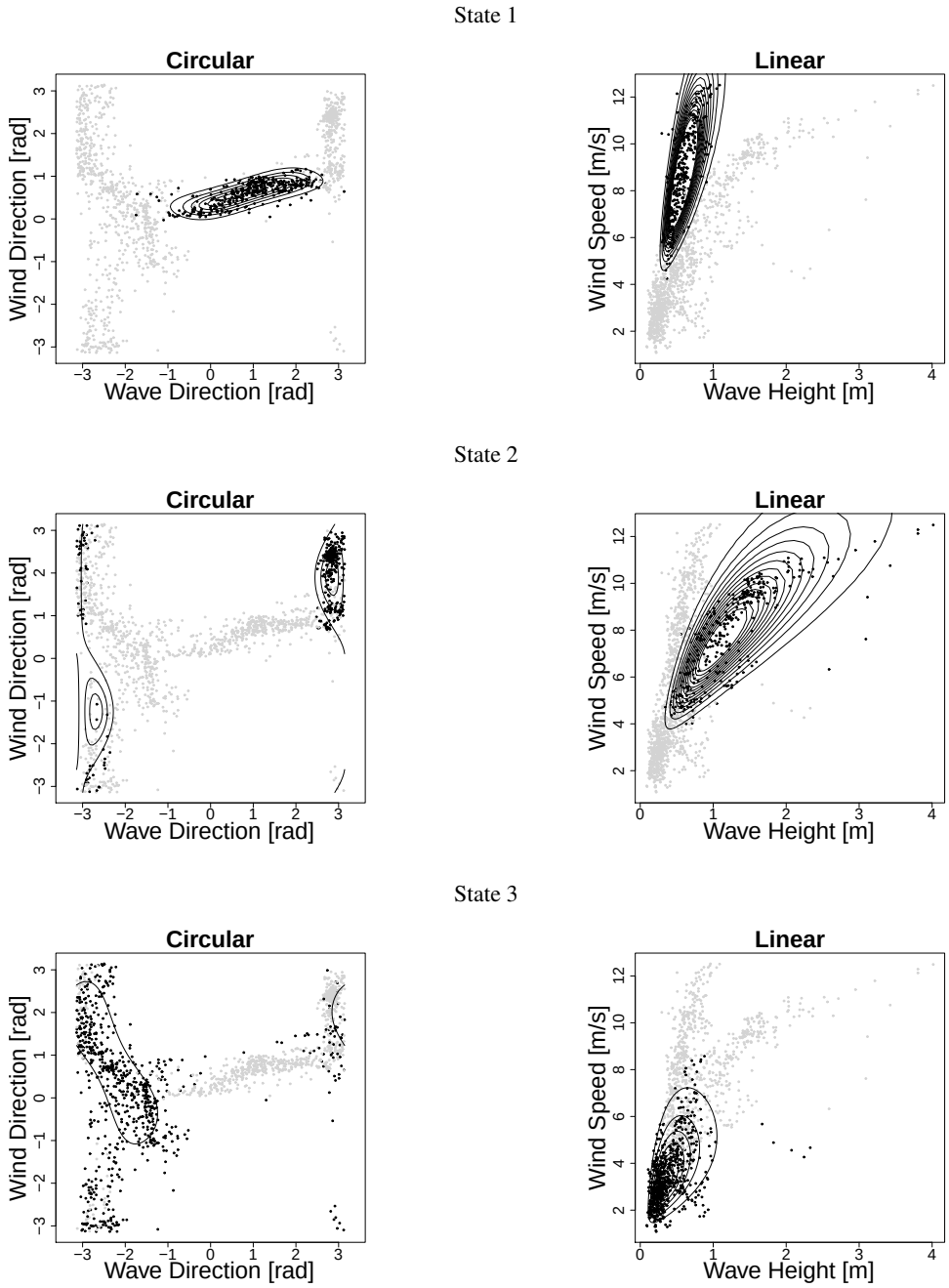
 ISSN: 2102-6238



The figure displays the marginal distributions of the models fitted to annual data (dot-dashed solid black lines) and bimonthly data (black solid Jan/Feb, grey solid Mar/Apr, black dotted May/Jun, grey dotted Jul/Aug, black dashed Sep/Oct, grey dashed Nov/Dec) at Civitavecchia.

FIGURE 7. *Univariate (marginal) distribution at Civitavecchia*

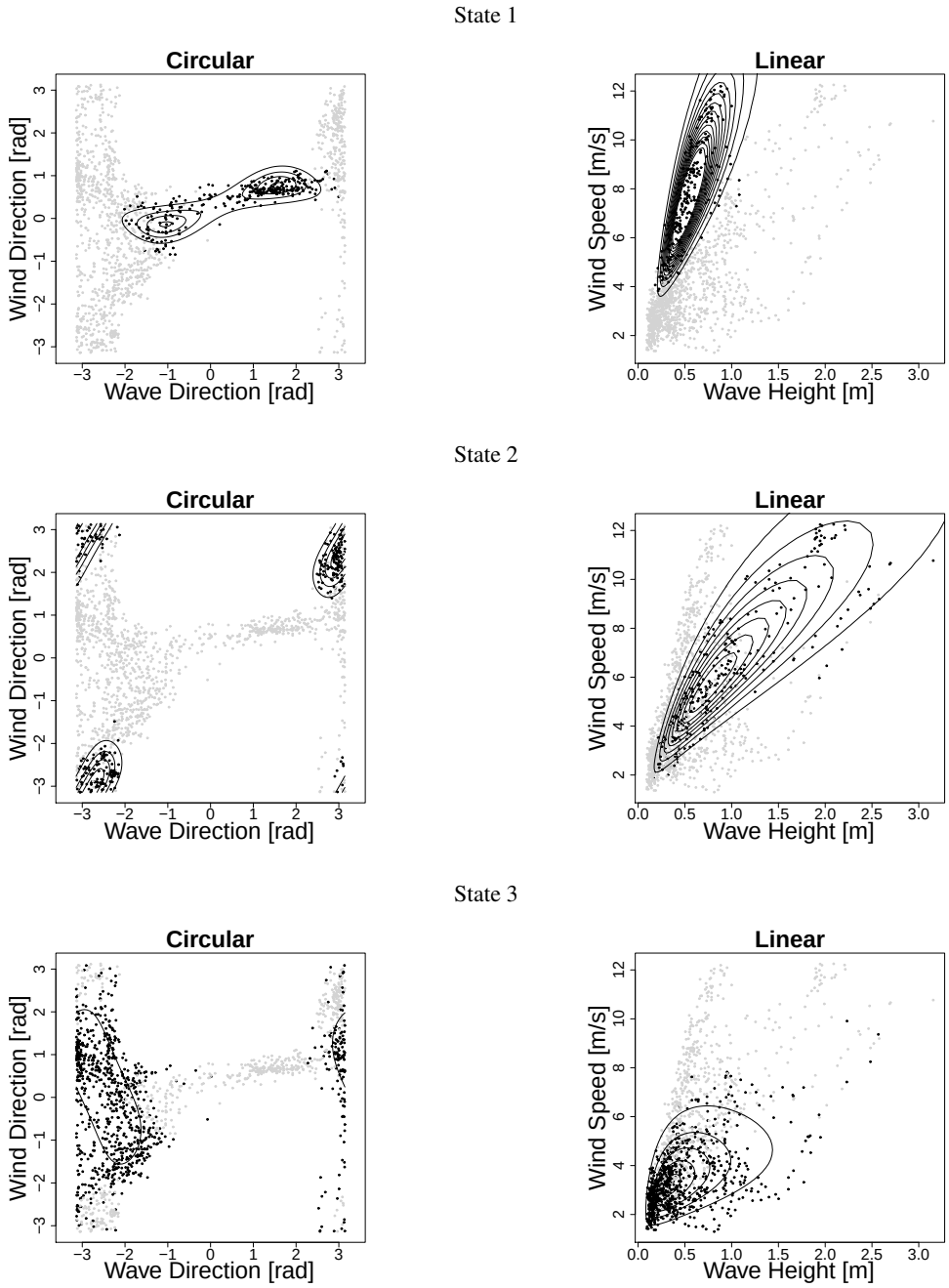
 ISSN: 2102-6238



The figure displays the conditional distributions of the fitted model at Civitavecchia in the bimester January/February. The circular components are on the left, and the linear components on the right, starting with the first state from the top. Observations classified to the respective state by local decoding are coloured black, while the remaining observations are gray.

FIGURE 8. *Conditional distributions at Civitavecchia in January/February*

 ISSN: 2102-6238



The figure displays the conditional distributions of the fitted model at Civitavecchia in the bimester September/October. The circular components are on the left, and the linear components on the right, starting with the first state from the top. Observations classified to the respective state by local decoding are coloured black, while the remaining observations are gray.

FIGURE 9. *Conditional distributions at Civitavecchia in September/October*

 ISSN: 2102-6238

TABLE 5. *Estimated transition probability matrices for Civitavecchia, bimester data*

The table summarizes the estimated transition probability matrices for Civitavecchia. Standard errors, reported in parentheses below each estimated, result from parametric bootstrap with 200 repetitions. Rows and columns respectively indicate origin and destination states.

	Jan-Feb			Mar-Apr			May-Jun		
State	1	2	3	1	2	3	1	2	3
1	0.975	0.009	0.016	0.948	0.011	0.041	0.933	0.05	0.018
	(0.009)	(0.007)	(0.007)	(0.015)	(0.008)	(0.013)	(0.015)	(0.017)	(0.015)
2	0.006	0.969	0.025	0.008	0.966	0.026	0.009	0.924	0.067
	(0.004)	(0.011)	(0.01)	(0.005)	(0.01)	(0.009)	(0.007)	(0.011)	(0.013)
3	0.01	0.012	0.978	0.016	0.021	0.963	0.019	0.077	0.904
	(0.005)	(0.005)	(0.007)	(0.004)	(0.006)	(0.007)	(0.01)	(0.018)	(0.014)
	Jul-Aug			Sep-Oct			Nov-Dec		
State	1	2	3	1	2	3	1	2	3
1	0.96	0	0.04	0.958	0.005	0.037	0.958	0.022	0.02
	(0.087)	(0.004)	(0.087)	(0.013)	(0.005)	(0.012)	(0.008)	(0.006)	(0.008)
2	0.001	0.987	0.012	0	0.952	0.048	0.026	0.953	0.021
	(0.001)	(0.003)	(0.003)	(0)	(0.013)	(0.013)	(0.008)	(0.01)	(0.007)
3	0	0.023	0.977	0.014	0.015	0.971	0.012	0.024	0.964
	(0.001)	(0.007)	(0.007)	(0.004)	(0.004)	(0.006)	(0.004)	(0.008)	(0.01)

TABLE 6. *Estimated transition probability matrices for Civitavecchia, annual data*

The table summarizes the estimated transition probability matrices for Civitavecchia. Standard errors, reported in parentheses below each estimated, result from parametric bootstrap with 200 repetitions. Rows and columns respectively indicate the origin and destination states.

	Jan-Dec		
State	1	2	3
1	0.958	0.007	0.035
	(0.005)	(0.002)	(0.005)
2	0.005	0.955	0.040
	(0.002)	(0.005)	(0.004)
3	0.010	0.018	0.972
	(0.001)	(0.002)	(0.002)

 ISSN: 2102-6238

TABLE 7. *Estimated parameters for Civitavecchia, bimester data*

The table summarizes the estimated parameters for Civitavecchia. The parameters β_1 - β_{12} belong to the circular component, and γ_1 - γ_{12} are the parameters of the linear component. In addition, the last column reports the initial/stationary distribution hidden Markov chain. Standard errors, reported in parentheses below each estimated, result from parametric bootstrap with 200 repetitions.

		β_1	β_2	β_{11}	β_{22}	β_{12}	γ_1	γ_2	γ_{11}	γ_{22}	γ_{12}	δ
	1	0.847	0.589	2.939	30.968	9.736	-0.575	2.132	0.072	0.049	0.046	0.25
		(0.046)	(0.014)	(0.195)	(2.206)	(0.645)	(0.015)	(0.011)	(0.006)	(0.003)	(0.004)	(0.095)
Jan	2	-3.06	0.323	18.978	0	-7.492	0.078	2.009	0.213	0.073	0.105	0.263
Feb		(0.009)	(0.04)	(1.302)	(0)	(0.516)	(0.027)	(0.015)	(0.014)	(0.005)	(0.007)	(0.092)
	3	-2.34	0.822	3.188	0.773	-2.189	-1.087	1.18	0.282	0.139	0.114	0.487
		(0.024)	(0.051)	(0.159)	(0.064)	(0.123)	(0.025)	(0.016)	(0.015)	(0.008)	(0.01)	(0.097)
	1	1.251	0.707	2.485	28.617	7.136	-0.568	2.068	0.113	0.082	0.07	0.197
		(0.044)	(0.014)	(0.262)	(2.429)	(0.737)	(0.018)	(0.015)	(0.009)	(0.007)	(0.007)	(0.061)
Mar	2	3.084	2.471	13.76	2.253	2.939	-0.267	1.565	0.519	0.22	0.229	0.345
Apr		(0.014)	(0.036)	(0.815)	(0.141)	(0.312)	(0.034)	(0.023)	(0.038)	(0.013)	(0.019)	(0.088)
	3	-2.219	0.018	3.037	1.322	-1.6	-1.079	1.243	0.24	0.133	0.101	0.458
		(0.028)	(0.045)	(0.188)	(0.068)	(0.137)	(0.019)	(0.017)	(0.012)	(0.006)	(0.008)	(0.066)
	1	0.438	0.436	0	5.644	3.296	-1	1.628	0.197	0.146	0.142	0.17
		(0.465)	(0.029)	(0.888)	(0.944)	(0.737)	(0.034)	(0.087)	(0.061)	(0.013)	(0.016)	(0.059)
May	2	-2.557	0.273	6.667	0.345	-1.454	-1.086	1.047	0.61	0.126	0.194	0.472
Jun		(0.491)	(0.077)	(1.001)	(0.718)	(0.64)	(0.034)	(0.084)	(0.064)	(0.007)	(0.015)	(0.074)
	3	-2.456	-2.426	6.558	2.648	6.88	-0.758	1.559	0.259	0.104	0.147	0.359
		(0.018)	(0.027)	(0.324)	(0.186)	(0.4)	(0.019)	(0.014)	(0.017)	(0.006)	(0.01)	(0.048)
	1	-0.238	0.514	1.232	98.993	5.667	-0.746	1.875	0.064	0.012	0.002	0.016
		(0.345)	(0.029)	(0.787)	(54.216)	(4.495)	(0.06)	(0.022)	(0.021)	(0.004)	(0.007)	(0.041)
Jul	2	-2.356	-2.367	8.179	0.286	3.395	-0.315	1.463	0.186	0.123	0.044	0.634
Aug		(0.014)	(0.05)	(0.352)	(0.08)	(0.292)	(0.251)	(0.123)	(0.015)	(0.008)	(0.006)	(0.081)
	3	-2.383	-2.42	9.307	0	2.002	-1.592	0.85	0.118	0.093	0.046	0.35
		(0.014)	(0.104)	(0.751)	(0.065)	(0.359)	(0.249)	(0.122)	(0.018)	(0.009)	(0.005)	(0.076)
	1	0.256	0.322	1.003	16.644	8.514	-0.663	2.004	0.123	0.076	0.081	0.194
		(0.048)	(0.016)	(0.135)	(1.339)	(0.71)	(0.021)	(0.019)	(0.01)	(0.006)	(0.007)	(0.063)
Sep	2	-2.981	2.883	13.808	4.683	8.359	-0.274	1.73	0.431	0.187	0.256	0.211
Oct		(0.022)	(0.033)	(1.218)	(0.445)	(0.847)	(0.037)	(0.026)	(0.041)	(0.019)	(0.027)	(0.062)
	3	-2.53	0.246	3.781	0.845	-1.723	-1.068	1.156	0.505	0.123	0.134	0.595
		(0.018)	(0.044)	(0.157)	(0.062)	(0.118)	(0.025)	(0.013)	(0.027)	(0.006)	(0.01)	(0.058)
	1	0.555	0.531	0	19.81	5.648	-0.914	1.701	0.223	0.152	0.156	0.31
		(0.052)	(0.01)	(0)	(1.228)	(0.391)	(0.022)	(0.018)	(0.015)	(0.01)	(0.012)	(0.061)
Nov	2	-2.438	-1.737	8.481	0.436	0.307	-0.109	1.533	0.795	0.436	0.452	0.329
Dec		(0.017)	(0.134)	(0.606)	(0.065)	(0.19)	(0.045)	(0.031)	(0.052)	(0.029)	(0.036)	(0.074)
	3	3.06	1.602	53.176	2.381	0.435	-0.221	1.708	0.396	0.242	0.281	0.36
		(0.005)	(0.039)	(2.921)	(0.107)	(0.558)	(0.031)	(0.024)	(0.024)	(0.014)	(0.018)	(0.08)

 ISSN: 2102-6238

TABLE 8. *Estimated parameters for Civitavecchia, annual data*

The table summarizes the estimated parameters for Civitavecchia. The parameters β_1 - β_{12} belong to the circular component, and γ_1 - γ_{12} are the parameters of the linear component. In addition, the last column reports the initial/stationary distribution hidden Markov chain. Standard errors, reported in parentheses below each estimated, result from parametric bootstrap with 200 repetitions.

		β_1	β_2	β_{11}	β_{22}	β_{12}	γ_1	γ_2	γ_{11}	γ_{22}	γ_{12}	δ
	1	0.528	0.512	1.008	21.044	6.991	-0.685	1.980	0.131	0.097	0.093	0.171
		(0.024)	(0.006)	(0.048)	(0.747)	(0.246)	(0.009)	(0.007)	(0.004)	(0.004)	(0.004)	(0.023)
Jan	2	-3.038	2.622	10.563	2.775	5.116	-0.223	1.786	0.443	0.159	0.245	0.256
Dec		(0.008)	(0.015)	(0.283)	(0.082)	(0.170)	(0.015)	(0.009)	(0.015)	(0.005)	(0.009)	(0.026)
	3	-2.411	0.260	4.145	0.596	-1.695	-0.951	1.175	0.506	0.157	0.162	0.573
		(0.008)	(0.020)	(0.087)	(0.022)	(0.053)	(0.010)	(0.006)	(0.010)	(0.003)	(0.005)	(0.025)

5. Discussions

We propose a dynamic-latent class approach to identify weather states under various wind and wave conditions, by estimating state-specific parameters, observed at different locations. Using mixtures of product densities to model multivariate data allows for a relatively simple specification of the dependence between mixed-type variables. We exploited bivariate von Mises and log-normal distributions, but a variety of alternative parametric families could be considered.

Our analysis shows that the HMM framework can be successfully employed for identifying different environmental conditions in rather complex settings. In our case study, we provide evidence that our approach is able to distinguish different environmental conditions at two geographical locations. This helps in reconstructing environmental conditions, as well as the dependence among environmental phenomena.

The estimated parameters and inferred states of the model allow for establishing a direct link to various, commonly accepted weather conditions. At Ancona, the four states correspond to Sirocco, Boral, Mistral, and a calm state. The three states at Civitavecchia are associated with Tramontane, Sirocco, and Libeccio conditions. These different states can be well-identified and observed in varying frequencies over the entire year. Nevertheless, the parameters of the model, in particular those of conditional distributions of the observations, evolve in time to adapt to the different periods. These patterns seem natural given the impact of different seasons throughout a year.

The comparably simple structure implied by our approach allows for a straightforward implementation for different environmental problems for which a hidden time-dependent structure could be assumed. A limitation of our approach concerns the so-called sojourn time distribution. In standard HMMs, sojourn time is geometrically distributed as a consequence of the Markov property of the underlying Markov chain. However, in some real-world problems this is an unrealistic assumption because the probability of a state change depends on time spent in the current state. Moreover, the median of the geometric distribution is always at one, an assumption to be discussed. An interesting alternative may be represented by hidden semi-Markov models, HSMMs (see, e.g., Barbu and Limnios, 2008, Bulla et al., 2010, Langrock and Zucchini, 2011) which represent a generalization of the HMMs, allowing for more general sojourn-time distributions. Moreover, the implementation of seasonal dependence in the parameters of the observation com-

 ISSN: 2102-6238

ponent or relationships to covariates may allow for better accounting for seasonal heterogeneity of the model.

Acknowledgements

Thanks are expressed to the Fédération Normandie-Mathématiques (FR CNRS 3335) for the continuous support of Antonello Maruotti and Marco Picone, as well as to the Italian Institute for Environmental Protection and Research for providing the data and financial support. Moreover, we would like to render thanks to two anonymous reviewers and Ingo Bulla for helpful questions and comments.

References

Ailliot, P., Baxevani, A., Cuzol, A., Monbet, V., and Raillard, N. (2011). Space-time models for moving fields. application to significant wave height. *Environmetrics*, 22(3):354–369.

Ailliot, P., Thompson, C., and Thomson, P. (2009). Space-time modelling of precipitation by using a hidden markov model and censored gaussian distributions. *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, 58(3):405–426.

Barbu, V. and Limnios, N. (2008). *Semi-Markov Chains and Hidden Semi-Markov Models Toward Applications: Their Use in Reliability and DNA Analysis*. Preliminary Entry 191 (Lecture Notes in Statistics). Springer-Verlag, New York - Heidelberg - Berlin.

Bartolucci, F. and Farcomeni, A. (2010). A note on the mixture transition distribution and hidden markov models. *Journal of Time Series Analysis*, 31(2):132–138.

Baum, L. and Eagon, J. (1967). An inequality with applications to statistical estimation for probabilistic functions of markov processes and to a model for ecology. *Bulletin of the AMS*, 73:360–363.

Baum, L. and Petrie, T. (1966). Statistical inference for probabilistic functions of finite state markov chains. *The Annals of Mathematical Statistics*, 37(6):1554–1563.

Bellone, E., Hughes, J., and Guttorp, P. (2000). A hidden markov model for downscaling synoptic atmospheric patterns to precipitation amounts. *Climatology Research*, 15:1–12.

Bertotti, L. and Cavaleri, L. (2009). Wind and wave predictions in the adriatic sea. *Journal of Marine Systems*, 78:227–234.

Betró, B., Bodini, A., and Cossu, Q. (2008). Using hidden markov model to analyse extreme rainfall events in central-east sardinia. *Envirionmetrics*, 19:702–713.

Böhning, D. (2000). *Computer-assisted Analysis of Mixture and Applications: Meta-analysis, Disease Mapping and Others*. Chapman & Hall.

Böhning, D. (2003). The em algorithm with the gradient function update for discrete mixtures with known (fixed) number of components. *Statistics and Computing*, 13:257–263.

Boukhanovsky, A., Lopatouhkin, L., and Guedes Soares, C. (2007). Spectral wave climate of the north sea. *Applied Ocean Research*, 29:146–154.

Bulla, J. and Berzel, A. (2008). Computational issues in parameter estimation for stationary hidden Markov models. *Computational Statistics*, 23(1):1–18.

Bulla, J., Bulla, I., and Nenadić, O. (2010). `hsmm` - an **R** package for analyzing hidden semi-Markov models. *Computational Statistics and Data Analysis*, 54(3):611–619.

Bulla, J., Lagona, F., Maruotti, A., and Picone, M. (2012). A multivariate hidden markov model for the identification of sea regimes from incomplete skewed and circular time series. *Journal of Agricultural, Biological, and Environmental Statistics*, 17(4):544–567.

Christopoulos, S. (1997). Wind-wave modelling aspects within complicate topography. *Annales Geophysicae*, 15:1340–1353.

Fraley, C. and Raftery, A. (2002). Model-based clustering, discriminant analysis and density estimation. *Journal of American Statistical Association*, 97:611–631.

 ISSN: 2102-6238

Hamilton, L. (2010). Characterising spectral sea wave conditions with statistical clustering of actual spectra. *Applied Ocean Research*, 32:332–342.

Hasselmann, S., Hasselmann, K., Janssen, P., Bauer, E., Komen, G., Bertotti, L., Lionello, P., Guillaume, A., Cardone, V., and Greenwood, J. (1988). The wam model - a third generation ocean wave prediction model. *Journal of Physical Oceanography*, 18:1775–1810.

Holzmann, H., Munk, A., Suster, M., and Zucchini, W. (2006). Hidden markov models for circular and linear-circular time series. *Environmental and Ecological Statistics*, 13(3):325–347.

Hughes, J. and Guttorp, P. (1994). A class of stochastic models for relating synoptic atmospheric patterns to regional hydrologic phenomena. *Water Resources Research*, 30(5):1535–1546.

Hughes, J., Guttorp, P., and Charles, S. (1999). A non-homogeneous hidden markov model for precipitation occurrence. *Journal of the Royal tatistical Society - Series C*, 48:15–30.

Lagona, F. (2005). Air quality indices via non homogeneous hidden markov models. In *Proceeding of the Italian Statistical Society Conference on Statistics and Environment, Contributed Papers, CLEUP*, pages 91–94, Padova.

Lagona, F., Maruotti, A., and Picone, M. (2011). *A Non-Homogeneous Hidden Markov Model for the Analysis of Multi-Pollutant Exceedances Data*. IntechOpen.

Lagona, F. and Picone, M. (2011). A latent-class model for clustering incomplete linear and circular data in marine studies. *Journal of Data Science*, 9:585–605.

Lagona, F. and Picone, M. (2012). Model-based clustering of multivariate skew data with circular components and missing values. *Journal of Applied Statistics*, 39:927–945.

Lagona, F. and Picone, M. (2013). Maximum likelihood estimation of bivariate circular hidden markov models from incomplete data. *Journal of Statistical Computation and Simulation*, 83(7):1223–1237.

Langrock, R. and Zucchini, W. (2011). Hidden markov models with arbitrary state dwell-time distributions. *Computational Statistics and Data Analysis*, 55(1):715–724.

Lasinio, G., Gelfand, A., and Lasinio, M. (2012). Spatial analysis of wave direction data using wrapped gaussian processes. *The Annals of Applied Statistics*, 6(4):1474–1498.

Maruotti, A. and Rocci, R. (2012). A mixed non-homogeneous hidden markov model for categorical data, with application to alcohol consumption. *Statistics in Medicine*, 31(9):871–886.

Ponce de Leòn, S. and Guedes Soares, C. (2008). Sensitivity of wave model predictions to wind fields in the western mediterranean sea. *Coastal Engineering*, 35:920–929.

R Development Core Team (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Robertson, A., Kirshner, S., and Smyth, P. (2004). Downscaling of daily rainfall occurrence over northeast brazil using a hidden markov model. *Journal of Climate*, 17(22):4407–4424.

Singh, H., Hnizdo, V., and Demchuk, E. (2002). Probabilistic model for two dependent circular variables. *Biometrika*, 89(3):719–723.

Zucchini, W. and Guttorp, P. (1991). A hidden markov model for space-time precipitation. *Water Resources Research*, 27(8):1917–1923.

Zucchini, W. and MacDonald, I. (2009). *Hidden Markov for Time Series: An Introduction Using R*. CRC Monographs on Statistics and Applied Probability. Chapman & Hall, London.

 ISSN: 2102-6238