



Classification of synoptic and local-scale wind patterns using k-means clustering in a Tyrrhenian coastal area (Italy)

Annalisa Di Bernardino¹ · Anna Maria Iannarelli² · Stefano Casadio² · Giovanna Pisacane³ · Gabriele Mevi² · Marco Cacciani¹

Received: 19 August 2021 / Accepted: 4 February 2022
© The Author(s) 2022

Abstract

In coastal regions, the complex interaction of synoptic-scale dynamics and breeze regimes influence the local atmospheric circulation, permitting to distinguish typical yet alternative patterns. In this paper, the k-means clustering algorithm is applied to the hourly time series of wind intensity and direction collected by in-situ weather stations at seven locations within 30 km from the western coastline of central Italy, in the proximity of Rome, over the period 2014–2020. The selection of both wind-integral quantities and ad hoc objective parameters allows for the identification of three characteristic clusters, two of which are closely related to the synoptic circulation and governed by persistent winds, blowing from either the northeast or the southeast direction throughout the day. In the latter case, synoptic and mesoscale contributions add up, giving rise to a complex circulation at the ground level. On the contrary, the third cluster is closely related to the sea breeze regime. The results allow the identification of some general information about the low-level circulation, showing that the synoptic circulation dominates in winter and, partly, in spring and autumn, when high ventilation and low recirculation conditions occur. Conversely, during summer the sea breeze regime is more frequent and stronger, generating intense air recirculation. Our analysis permits to discern rigorously and objectively the typical coastal meteorological patterns, only requiring anemological in-situ data.

1 Introduction

Coastal meteorology focuses on the meteorological phenomena that are directly caused—or significantly influenced—by the rapid changes in atmospheric conditions that occur at the land-sea transition zone (National Research Council 1994). As it is estimated that more than 600 million people (about 10% of the world population) today live in coastal areas less than 10 m.a.s.l. (UN 2017), a thorough understanding of weather conditions in coastal environments is crucial to

correctly evaluate the critical parameters that affect human health and wellness, e.g., air quality and thermal comfort.

The sea breeze regime (SBR) is a relevant topic in the context of coastal meteorology, as it characterizes the mesoscale circulation, accounting for many spatially and temporally nested phenomena (Miller et al. 2003), and with cascading impacts on the local environment (Masselink 1998). The onset of the SBR is triggered by the temperature difference between the colder sea surface and the warmer land, giving rise to a pressure gradient that is particularly intense during sunny, summer days. In addition to the thermal forcing, the SBR is also affected by other factors, such as the coastline curvature (Steele et al. 2013), the proximity of reliefs and orography (He et al. 2020), and, above all, the synoptic-scale weather conditions (Di Bernardino et al. 2021a). The separation of the breeze regime from the synoptic conditions is indeed one of the major problems encountered in SBR studies (Zhong and Takle 1993), whose investigation is made even harder in the urban environment of coastal cities, where the SBR also interacts with the urban heat island (UHI) effect (Bauer 2020). The UHI main characteristic are: (1) the local increase in temperature, especially at night

Responsible Editor: J.-F. Miao.

Marco Cacciani: Prematurely passed away on January 18, 2022.

✉ Annalisa Di Bernardino
annalisa.dibernardino@uniroma1.it

¹ Physics Department, Sapienza University, 00185 Rome, Italy

² SERCO Italia SpA, 00044 Frascati, Rome, Italy

³ Climate Modeling and Impacts Laboratory, ENEA-Casaccia Research Center, 00123 Rome, Italy

and during summer, due to the presence of heat-absorbing materials such as asphalt and concrete, which modify the natural albedo and heat capacity of the surfaces, and (2) the limited evapotranspiration and soil permeability, due to the reduction of green areas. The occurrence, intensity, and inland penetration of the SBR can strongly affect local air quality in cities, as wind conditions are among the main factors responsible for the accumulation of pollutants near the ground. For instance, the ingress of cold and humid air into the inland in the mornings can determine the development of a thermal internal boundary layer, in which the contaminants produced at the ground level can be trapped (Wei et al. 2018). Similarly, the absence of the breeze regime, in conjunction with unfavourable synoptic weather conditions (i.e., wind calm, persistent atmospheric stability), can cause thermal inversions, with the consequent increase in pollutants concentration near the ground (Fortelli et al. 2016). On the other hand, during the night hours, the air in contact with the mainland cools more rapidly than that over the ocean, due to the higher heat capacity and consequent thermal inertia of the sea. This leads to a situation of local high pressure over the land, and low pressure above the sea, with a consequent displacement of air from the land to the sea, typically defined as land breeze regime. In this case, the wind can transport atmospheric pollutants out of the cities, improving the air quality.

Although the SBR has been widely studied in recent years from both the experimental (Cenedese et al. 2000; Puygrenier et al. 2005) and the numerical (Fu et al. 2021; Di Bernardino et al. 2021b) point of view, its identification only relies on subjective and site-specific criteria, impairing the objective separation of the pure sea-breeze contribution to the local circulation from the concurrent synoptic influence, which can disturb, limit or exacerbate the inland penetration of the breeze (Azorin-Molina et al. 2011). Only recently, researchers have attempted to isolate the pure sea breeze component by analysing high-resolution models and observations and developing less site-specific criteria and no-threshold methods (Cafaro et al. 2019).

The objective classification of synoptic weather patterns, and the subsequent isolation of the SBR from synoptic-scale phenomena, can be carried out by means of clustering techniques. In recent years, clustering algorithms have been widely used for meteorological and climatological studies, although the choice of the input parameters represents a critical constraint for successful algorithm setup (see the review by Huth et al. 2008 for details about clustering techniques and meteorological applications). Among these, thanks to its simplicity of implementation and computation efficiency, the k-means algorithm is widely used for the classification of environmental data—e.g., observations from meteorological stations (Li et al. 2020), evapotranspiration measurements (Ferreira et al. 2019), and air quality parameters (Adame

et al. 2012)—as well as for wind energy resource assessments (Al-Shammari et al. 2016) and rainfall prediction (Geetha and Nasira 2014).

The principal aim of this study is the operative identification of the anemological patterns over the Tyrrhenian coast of central Italy, in the Lazio region, by applying the k-means clustering algorithm. The investigated domain also includes the city of Rome, which is the largest and most populous Italian city. Following the procedure proposed by Li et al. (2020), the clustering procedure only requires wind speed and direction data at the surface, measured at coastal stations. The minor modifications that have been introduced to better account for the region's peculiarities are described in Sect. 2.2. With respect to other clustering algorithms, this method offers the advantage of affordable data requirements, anemological observations being generally available with high temporal resolution and spatial coverage. In addition, the objective parameters used to guide the clustering procedure also provide valuable diagnostics to investigate the relative roles of ventilation and air recirculation in the region. The resulting classification, based on the analysis of the anemological data over the period 2014–2020, provides a representative picture of the synoptic and low-level circulation in the area of interest. Future applications envisage the characterization of the relationship between anemological patterns and air quality in the urban area of Rome and its surroundings and the assessment of the wind energy resource in the region. If sufficiently long time series are available (at least 30-year period, according to WMO 2017), this method will also allow to investigate the impact of climate change on the inter-annual variability of typical weather conditions (Ogaya and Peñuelas 2021; Pohl et al. 2021). Moreover, the procedure could be easily extended and applied to other coastal regions with similar conditions (e.g., shoreline orientation, breeze onset, and offset times).

The paper is organized as follows. Section 2 presents the study domain, the meteorological dataset, and the k-means clustering technique. The main results are described and discussed in Sect. 3, especially focusing on the seasonal and monthly occurrence of the identified clusters. Finally, Sect. 4 summarizes the conclusions of the study and possible future developments.

2 Methodology and dataset

2.1 Study area and meteorological data

The domain investigated in this study covers the coastal area of the Lazio region, central Italy, extending for about 100 km in the North-West/South-East direction. The area is generally flat near the coast and hilly in the inland, where four main orographic systems are located, i.e., the Tofia and

Sabatini Mountains in the North and the Alban Hills and the Lepini Mountains in the South (see Fig. 1). The domain includes the city of Rome, the capital of Italy, and its metropolitan area, which together host about 4.3 million inhabitants, besides several smaller towns.

In the present study, seven in-situ weather stations are selected, belonging to the Regional Agency for the Development and Innovation of Agriculture of Lazio (ARSIAL, <http://www.arsial.it/arsial/>) meteorological network. The distribution of the meteorological sites is depicted in Fig. 1, and their topographic characteristics are summarized in Table 1. All the stations are located within a maximum distance of 30 km from the coast, where the effects of the sea breeze have been proved to be still relevant, also reaching the innermost Rome area (Colacino 1982; Di Bernardino et al. 2021c). Stations 01, 02, and 07 are equipped with Campbell instruments (Campbell Scientific Europe, Loughborough, United Kingdom), while the sensors of stations 03, 04, 05, and 06 are produced by Siap+Micros (Siap+Micros

S.p.A., San Fior, Treviso, ITALY). All sensors comply with WMO (World Meteorological Organization) requirements, although detailed information about data uncertainty is currently unavailable.

Although the selected datasets had been previously validated by ARSIAL, additional pre-processing and quality control checks have been carried out. First, data were visually inspected and screened to eliminate gross errors, to standardize the formats and units of measurement, and to assess the presence of missing data. In all the selected stations, the percentage of missing hourly averaged wind speed and direction data was found to be less than 1.7%, with the exception of stations 01 and 07, for which the missing data amounted to 5.9% and 4.8% of the total, respectively. Missing data appeared to be randomly distributed over time and not to affect the results of the study. Next, the time hourly averaged series were inspected for errors attributable to non-climatological signals, such as instrumental artefacts, changes in the position of the weather stations,

Fig. 1 Map of the selected stations. Numbers refer to the Station ID reported in Table 1. The main geographical elements of the region are shown, while the red circle depicts the municipality of Rome



Table 1 List of the surface weather stations considered in the present study

Station ID	Station	Latitude (°)	Longitude (°)	Altitude (m.a.s.l.)	Distance from coastline (km)
01	Montalto	42.31	11.60	4	< 1
02	Canale Monterano	42.15	12.10	363	19
03	Cerveteri	41.99	12.04	7	1
04	Fiumicino	41.82	12.25	3	3
05	Roma	41.92	12.52	36	28
06	Marino	41.75	12.62	173	20
07	Velletri	41.69	12.78	322	25

and non-routine maintenance. As a result of such screening, only the hourly averaged wind speed and direction data covering the period 01/01/2014–31/12/2020 were retained for subsequent analysis.

2.2 K-means clustering

The k-means clustering algorithm is a method of vector quantization, developed by Wong and Hartigan (1979), which allows for the grouping of data within a given number of subsets, called clusters, which are defined following the seven objective parameters described below.

The technique requires two user-defined input parameters: (1) the number of clusters (k), into which the data will be divided, and (2) the initial (random) position of the centroid for each cluster. Once the centroids have been randomly initialized, the algorithm assigns each measurement to the closest cluster. Then, the positions of the cluster centroids are iteratively updated, according to the values of the objective parameters computed for cluster members, and optimized until convergence is achieved. Results are assumed to converge when the position of the centroids stabilizes, that is, when the SSE or inertia (i.e., the sum of the squared deviations of the computed variables from the correspondent cluster mean) is minimized. It should be emphasized that the k-means algorithm forcibly assigns each analysed element to one of the clusters, without the contemplating situations that do not fall within the predefined clusters.

The definition of the objective parameters that guide the clustering procedure moves from the work of Allwine and Whiteman (1994), who proposed a method for the investigation of air mass ventilation, recirculation, and stagnation. Using wind observations from a single station, they defined three characteristic quantities: (1) the “transport distance”,

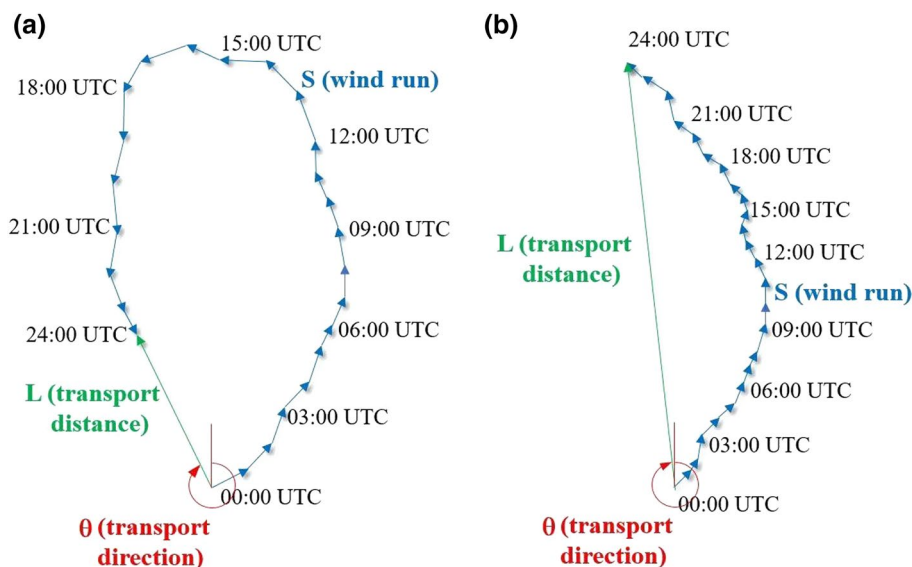
L , which is the net displacement of the wind trajectory, i.e., the actual distance between the start point and the endpoint of the path travelled by the air particle during a day, (2) the “wind run”, S , defined as the scalar sum of the transport distance at each time step (1 hour, in the present case), and (3) the “recirculation factor”, R , which is the ratio between L and S . Low (high) values of S describe situations of stagnation (ventilation). R tends to one when the wind has a persistent direction, while approaches zero in the case of substantial atmospheric recirculation (Fig. 2). In addition, Allwine and Whiteman (1994) defined the wind rotation angle, θ , measured from the North clockwise.

However, the above integral quantities are insufficient to fully describe the dynamics of coastal circulation and to discriminate between different weather regimes. Additional diagnostics are thus needed to better characterize the SBR and isolate it from the concomitant synoptic weather conditions (Li et al. 2020). We therefore adopted the following list of relevant objective parameters:

- (i) u_{morn} , averaged zonal wind component in the period 0800–1200 UTC;
- (ii) v_{morn} , averaged meridional wind component in the period 0800–1200 UTC;
- (iii) u_{after} , averaged zonal wind component in the period 1500–1800 UTC;
- (iv) v_{after} , averaged meridional wind component in the period 1500–1800 UTC;
- (v) R , recirculation factor;
- (vi) $\cos(\theta)$;
- (vii) $\sin(\theta)$.

The first four parameters were selected on consideration of the typical hour range that comprises the onset and

Fig. 2 Schematic illustration of the wind run S (blue arrows), transport distance L (green arrow), and transport direction θ (red arrow). **a, b** Show the case of high and low recirculation, respectively



decline of the SBR in the study area (Di Bernardino et al. 2021c), regardless of the seasonal trend in insolation. They enable the clustering algorithm to capture events of persistent wind throughout the day and to separate them from the breeze-dominated days. Rather than the wind rotation angle, θ , $\cos(\theta)$ and $\sin(\theta)$ were used, so as to avoid discontinuities in the North quadrants and the consequent clustering errors. Before applying the k-means algorithm, all the parameters were normalized, so as to have zero mean and unit variance.

Accordingly, the time series from the seven meteorological stations, now spanning a total of 2482 days, were further processed to derive the regional time series of the objective parameters. Spatial averaging ensures that local effects due to the orographic differences across the selected sites are reduced and that the resulting values coherently represent the mesoscale patterns.

3 Results

3.1 Identification of clustering patterns

In this section, we describe the methods adopted for the evaluation of the most suitable number of clusters and the main characteristics of the identified patterns.

The greatest difficulty for the application of k-means clustering lies in the prescription of the optimal number of clusters, k_O , for data classification. Here, k_O is obtained by applying the “elbow” and the “silhouette” methods proposed by Kodinariya and Makwana (2013), and comparing the results. In both methods, the clustering algorithm is run several times with the same input data and varying k , so as to construct alternative k -dependent metrics. The “elbow” method derives inertia (or SSE) as a function of k , and identifies k_O as the value after which inertia starts decreasing following a quasi-linear trend. On the other hand,

the “silhouette” method measures the cohesion separation distance between the k clusters, by estimating how close each element of one cluster is to the elements of the neighbouring clusters. In this case, k_O is the value that maximizes the average silhouette coefficient. The resulting alternative metrics are shown in Fig. 3. For the area of interest, $k_O = 3$ appears to be a reasonable choice.

Afterwards, the algorithm was run with the following setup: (1) random choice for the initial centroid positions, (2) a maximum of 300 iterations for the k-means algorithm, (3) maximum relative tolerance of the difference in the cluster centres of 2 consecutive iterations with respect to Frobenius norm of 10^{-4} . The results shown here were achieved after 18 iterations and exhibit reasonable internal coherence, as indicated by a within-cluster SSE of about $10e4$.

Figure 4 presents the daily trend of wind speed (represented by the length of the vectors and expressed in m/s) and direction (indicated by the orientation of the vectors and expressed in degrees) for the three clusters detected by the k-means algorithm. The plots show the diurnal variation of the wind vectors averaged over the population of each cluster.

Figure 4a shows the wind evolution along the day for the cluster hereinafter indicated as “Northeasterly Cluster”. The prevailing wind direction is stably from the northeast throughout the day. Wind intensity ranges between 1.75 (1500 UTC) and 3.42 m/s (1000 UTC), reaching a maximum in the central hours of the day, with a daily-average of 2.37 m/s. The local circulation appears to be dominated by synoptic forcing, and the SBR fails to develop. The second typical pattern detected, shown in Fig. 4b, is hereinafter called “Breeze Cluster”. During the night, the wind blows from the Northeast quadrant, i.e., from inland, with very low intensity (a minimum of 1.03 m/s is reached at 0500 UTC). At 0800 UTC, the stations record a sharp change in wind direction, associated with generally calm wind conditions.

Fig. 3 Results of the **a** “elbow” and **b** “silhouette” methods used for the identification of the correct number of clusters

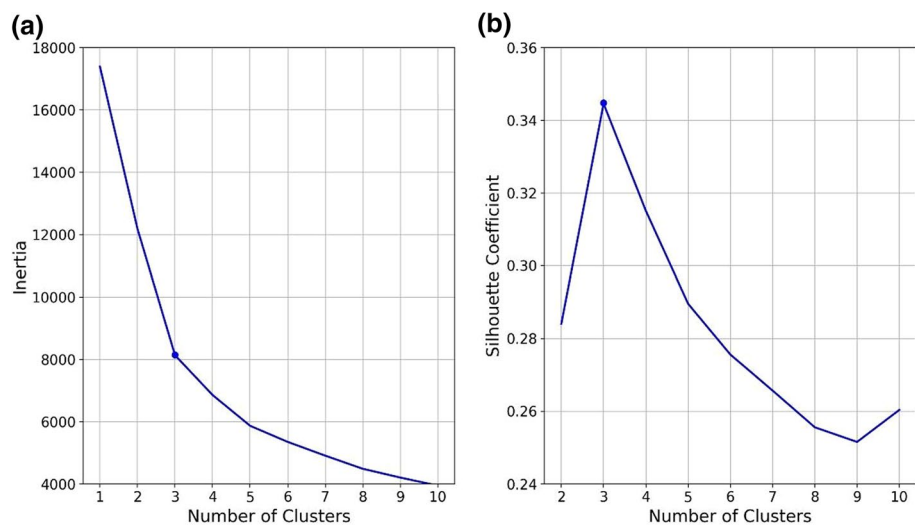
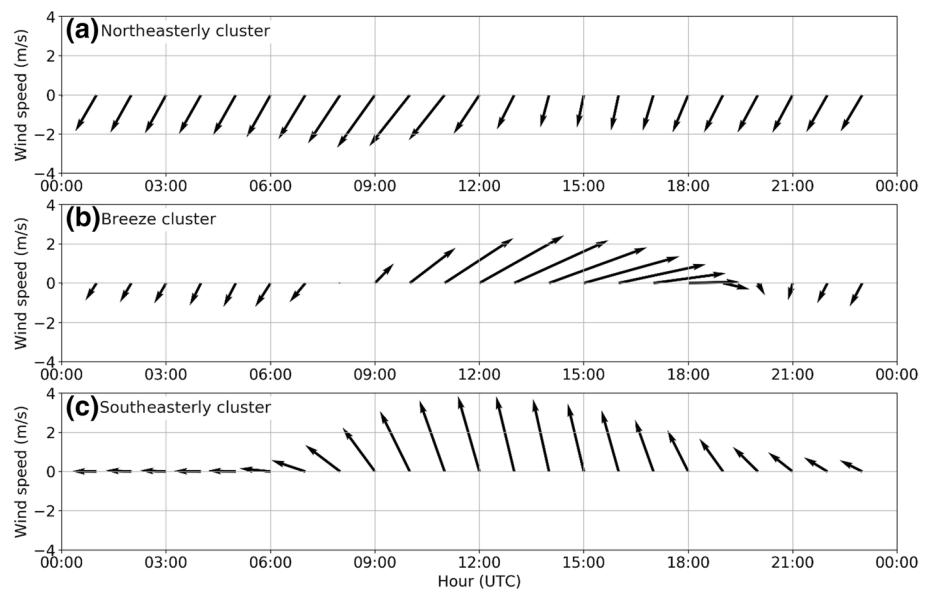


Fig. 4 Temporal variation of hourly wind speed (m/s, vector length) and direction (degrees, vector orientation) for the three clusters identified by the k-means algorithm



From 0900 UTC onwards, the wind progressively veers, until it blows perpendicular to the coastline. Such rotation corresponds to a rapid increase in wind speed, as expected at the onset of the sea breeze, which reaches a maximum of 3.72 m/s at 1400 UTC. On average, the breeze is present until 1800 UTC, when the wind rotates clockwise and the speed decreases back to the night values typical of the land breeze. It is worth pointing out that the graphed trend was obtained by pooling all data together, without accounting for seasonal variability, and therefore smoothing out any seasonal difference in the onset and offset time of the breeze. This second cluster is representative of pure-breeze events, in which the mesoscale and the local circulation prevail over the synoptic-scale dynamics. Figure 4c depicts the third pattern detected, hereinafter named “Southeasterly Cluster”. Here, the synoptic and mesoscale circulations interact, giving rise to a more complex anemological pattern. During the night, the wind blows from the East, with speeds below 1.5 m/s. From 0600 UTC onwards, the wind rotates clockwise, settling in the Southeast quadrant. The wind velocity gradually increases until 1300 UTC, and then it slowly decreases. The wind direction remains fairly constant until 1700 UTC. In the evening, the wind veers counterclockwise and a decrease in intensity is observed. The anemological trend of this cluster is attributable to the varying relative weights of the synoptic-scale and sea-breeze regimes in determining the overall wind pattern, with the low-intensity synoptic-scale wind from the East prevailing at night, and the SBR causing the acceleration and the diurnal rotation of the flow.

The results relative to the breeze cluster are in agreement with previous studies conducted in the same area. Colacino (1982) and Petenko et al. (2011) showed that this region is subject to two prevailing wind regimes: (1) the

drainage effect from the bottom of the Tiber valley, which generates a flow from the North and is mainly observed in the eastern side of the city, where the proximity of the reliefs influences the circulation, and (2) the SBR, which affects the region during daytime throughout the year, but more frequently in summer. Furthermore, these results are in accordance with the findings by Di Bernardino et al. (2021c), who showed that the wind blows from the Southwest when the urban center of Rome is reached by the sea breeze, and remains almost constant in direction for several hours.

Table 2 summarizes the relative occurrence of each cluster over the period of interest, together with the time means and standard deviations of the associated objective parameters. The three distinct patterns occur with comparable frequencies, the Breeze Cluster being present 37% of the time, followed by the Southeasterly Cluster (32.6%) and by the Northeasterly Cluster (30.4%). As expected, the average values of L , S , and R are comparable across the latter two clusters, with R equal to 0.79 and 0.65 for the Northeasterly and the Southeasterly clusters, both indicating high ventilation and low recirculation conditions in consequence of the persistent wind direction (Fig. 2, panel b). This result confirms the internal coherence and robustness of the classification obtained via the k-means algorithm. On the other hand, the Breeze Cluster is characterized by a markedly lower R (0.49), consistently with the recirculation conditions induced by wind rotation (Fig. 2, panel a), which also account for the significantly lower L (30.2 km). A decrease in S (58.5 km) is also observed. These values are in agreement with the results in Li et al. (2020): although the regions of interest have different morphological and climatological characteristics, the wind-integral quantities have similar values for comparable patterns.

Table 2 Occurrence and mean (standard deviation) of input parameters for the three clusters

Cluster	<i>n</i> , days	Days ratio (%)	u_{norm} (m/s)	u_{norm} (m/s)	u_{after} (m/s)	v_{after} (m/s)	<i>L</i> (km)	<i>S</i> (km)	<i>R</i> (-)	θ (°)
Northeast. cluster	755	30.4	- 1.8 (1.6)	- 2.4 (2.3)	- 0.5 (1.9)	- 1.8 (2.1)	60.0 (38.3)	71.8 (34.3)	0.79 (0.19)	209.7 (23.2)
Breeze cluster	917	37.0	1.9 (1.5)	1.3 (1.6)	2.7 (1.4)	0.5 (1.4)	30.2 (23.3)	58.5 (23.5)	0.49 (0.21)	96.4 (59.2)
Southeast. cluster	810	32.6	- 1.4 (1.8)	3.2 (1.8)	- 0.9 (1.9)	2.8 (1.9)	55.5 (39.2)	77.7 (35.1)	0.65 (0.21)	289.4 (85.9)

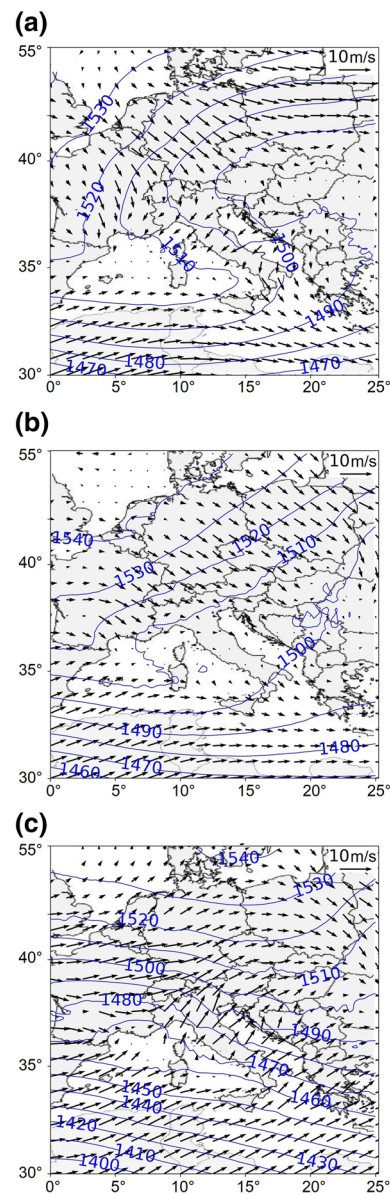


Fig. 5 Average synoptic conditions for **a** northeasterly cluster, **b** breeze cluster, and **c** southeasterly cluster at 1200 UTC from ERA5 reanalysis. Contours are isobars (in geopotential metres), while vectors depict wind intensity and direction at the geopotential height at 850 hPa

To appreciate the relative influence of synoptic and local atmospheric circulation on the different clusters we identified, Fig. 5 shows the corresponding average synoptic maps, as derived from the ECMWF atmospheric reanalysis ERA5 data (<https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5>), by only considering the days belonging to each single cluster. Geopotential data, used to compute geopotential heights, and wind components at the 850 hPa level with a spatial resolution of ~ 30 km

(Albergel et al. 2018) are considered. Maps refer to 1200 UTC, chosen as representative of the whole day.

For the Northeasterly Cluster, Fig. 5a shows a high-pressure system located over the North Sea and a low-pressure system over the Balkans, as well as a trough over the western Mediterranean. In this baric configuration, Italy is at the convergence zone between a western anticyclonic system and an eastern cyclonic system. These conditions typically give rise to northeasterly winds. Figure 5b depicts the distinctive synoptic circulation occurring in the Breeze Cluster. The map depicts the typical summer conditions of the Italian peninsula: no evident pressure gradients are observed over Italy, and, consequently, there are no relevant synoptic winds, ensuring intense heat and good weather. These conditions, coupled with the intense atmospheric heating, responsible for the difference in temperature (and, therefore, pressure) between ocean waters and the dry land, favour the predominance of mesoscale or local-scale phenomena, such as the SBR. Finally, Fig. 5c presents the characteristic conditions occurring in the case of the Southeasterly Cluster. A high pressure region is located over northern Europe and a high baric gradient is observable between northern and southern Italy, giving rise to winds from the southern quadrant. Again, we underline that the map shows the average conditions at 1200 UTC, i.e., when the SBR interacts with the synoptic circulation, as shown by the daily wind trend in Fig. 4c. Southern winds determine persistent high-pressure conditions and the advection of air masses from North Africa, also responsible for the transport of Saharan dust up to the middle latitudes (Gobbi et al. 2019).

The aforementioned results suggest that the clustering algorithm properly worked, allowing for the identification

of anemological low-level patterns affected by both the synoptic and the mesoscale circulation, by only requiring the measurements of wind speed and direction from coastal stations. In what follows, the cluster analysis is deepened by considering the monthly variability of the wind integral quantities and the seasonal and monthly distribution of pattern occurrences.

3.2 Seasonal and monthly analysis

For each cluster, Fig. 6a shows the distribution of the number of occurrences across seasons. As already remarked, during the analysed period, the three clusters have comparable frequencies, with a slight predominance of the Breeze Cluster (37%) with respect to the Southeasterly and Northeasterly Clusters (32.6% and 30.4%, respectively). From the figure, it is evident that the Northeasterly Cluster mainly occurs during winter (43.8% of its total occurrences) and that the autumn and spring samples still representing a significant fraction of the total (29.9% and 18.4%, respectively). The relative number of occurrences sharply decreases in summer (7.9%). On the other hand, the Southeasterly Cluster has a more homogeneous seasonal distribution, the relative number of occurrences exhibiting limited variations across seasons (spring: 29.5%, autumn: 25.1%, winter: 23.9%, summer: 21.5%). Finally, the overall slightly prevalent Breeze Cluster is mostly observed in summer (44% of its total occurrences) and spring (29.1%), with a less populated autumn sample (18.3% of the total) and minimum manifestation in winter (6.6%). The results are in agreement with those presented by Mastrantonio et al. (2006), who studied the anemological conditions of the same area, showing

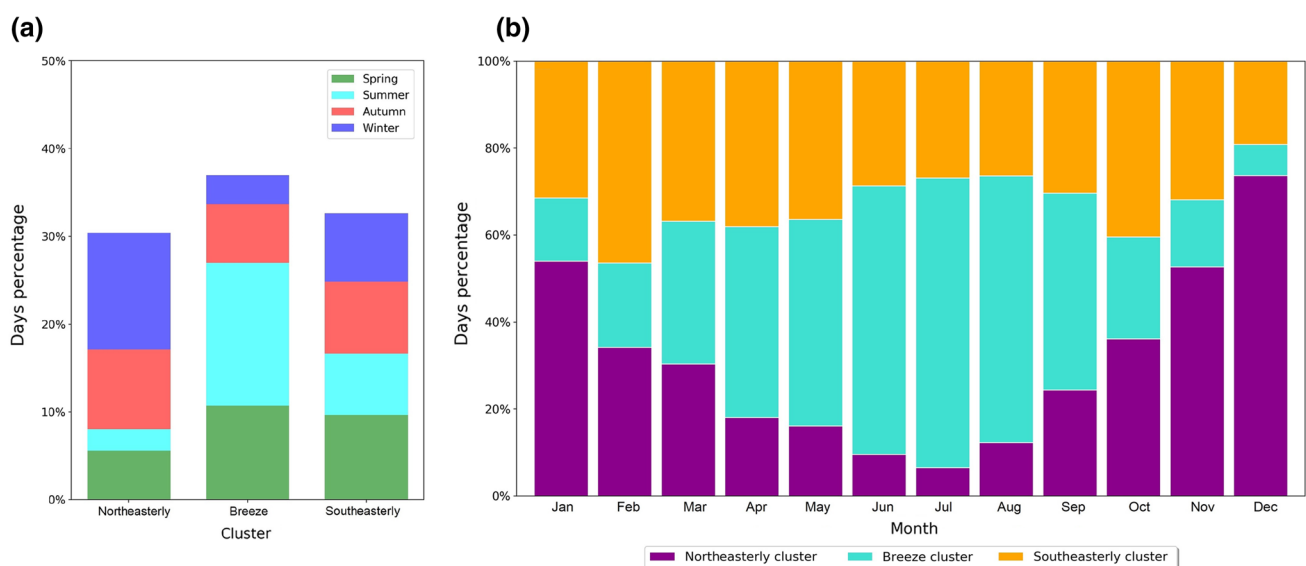


Fig. 6 Histogram of the **a** seasonal and **b** monthly occurrence of different clusters

the persistence of sea-breeze circulation for a large part of the year, with a pronounced diurnal cycle. Moreover, in the same region Petenko et al. (2011) detected the SBR throughout the year, with higher frequency in summer and spring, by exploiting both ground-based observations and numerical simulations.

From the point of view of seasonal time frequencies, it is observed that the Breeze and the Southeasterly Clusters have similar frequencies in spring (41.3% and 37.1%, respectively), as compared to the more sporadic Northeasterly Cluster (21.6%). In summer, the SBR prevails (63.3% of the time), followed by the southeasterly pattern (27.3%), which is associated to the interaction of the atmospheric flow with the North African promontories, resulting in the typical summer high pressure and fair weather conditions (Fig. 5). In autumn, the three patterns have a homogeneous distribution, with a slight prevalence of the Northeasterly and Southeasterly Clusters (37.9% and 34.1%, respectively). During winter these two patterns definitely prevail, with the former appearing 54.5% of the time and latter exhibiting a relative frequency of 32%.

The variability of the relative frequencies associated to the three clusters is more evident when considering their monthly variations along the year, as shown in Fig. 6b. The monthly frequency of the Northeasterly Cluster decreases along the transition from winter to spring, reaching a minimum in summer (6% in July), and increasing again in autumn and winter. In December, this cluster occurs about 75% of the time. The plot highlights that the monthly frequency of the Southeasterly Cluster is fairly homogeneous throughout the year, ranging from 20 to 40% of the total monthly cases, with minima in August (27%) and December (20%). The evolution in the monthly frequency of the Breeze Cluster, as expected, is in counter phase with respect to that of the Northeasterly Cluster, reaching its minimum in

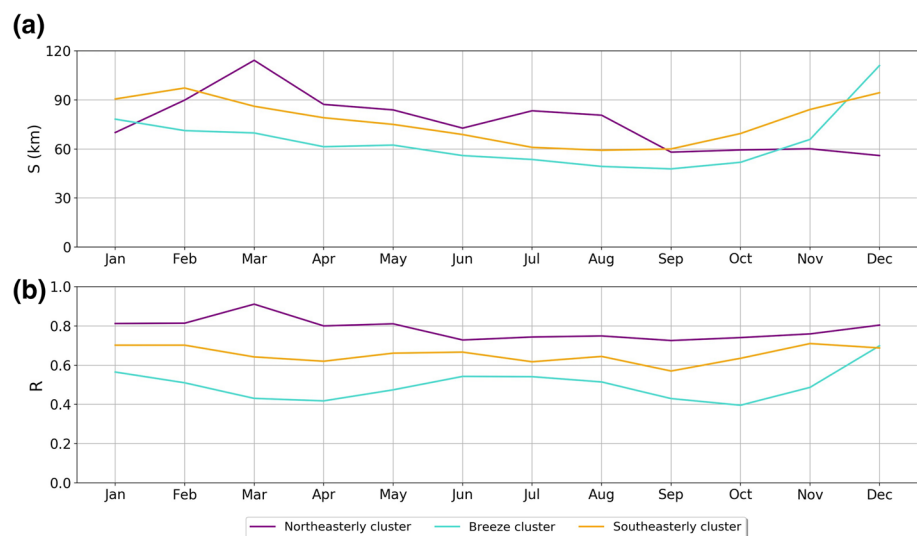
winter and its maximum in summer (9% in December and 68% in July).

Finally, to evaluate the temporal variability of the wind-integral quantities defined in Sect. 2.2, Fig. 7 shows the monthly evolution of the average wind run in km, S (panel a), and recirculation factor, R (panel b), obtained for each cluster.

For the Northeasterly Cluster, S shows a strong monthly variability, reaching a maximum of about 112 km in spring (March), then progressively decreasing in summer, autumn, and winter, when it reaches a minimum of approximately 58 km (December). The monthly evolution of R is in line with that of S , although with more limited variability (maximum in March, 0.90, minima in June and September, 0.72). During spring and summer, when the occurrence of the Northeasterly Cluster is anyway rarer, R assumes the largest values in the year, despite the comparatively higher values of S with respect to the other clusters. This implies that the spring and summer months are characterized by more intense ventilation, mainly due to persistent synoptic wind conditions, as highlighted in the previous section.

The Breeze Cluster exhibits S values that are comprised between about 50 km (September) and 80 km (January), with a peak of approximately 112 km in December. S decreases in spring, reaches a minimum in summer and then rises again during autumn and winter. R is relatively lower in spring and autumn and relatively higher in summer, when the breeze is more frequent and intense, and air particle displacement is anyway enhanced. In general, R ranges between 0.40 and 0.55, with a somewhat unexpected peak of about 0.70 in December, that is still sustained in January. However, it is important to note that the relatively higher winter values of R are still lower than those found for the other two clusters, and may be associated with the passage of cold fronts, which align with the breeze front and are not distinguished

Fig. 7 Monthly trend of **a** wind run, S , and **b** recirculation factor, R , for different clusters



by the algorithm, due to their having similar anemological characteristics. This aspect will be further investigated in future research.

For the Southeasterly Cluster, S reaches the highest values (about 95 km) in autumn and winter (February and December) and the lowest (about 60 km) in summer (July and August). The values of R are constantly confined between the curves derived for the two other clusters, and range between 0.58 (September) and 0.70 (November), with limited monthly variability.

The analysis of seasonal and monthly variability therefore allows to conclude that, in the examined region, the synoptic and the mesoscale meteorological conditions alternatively prevail and govern the local ground-level circulation, depending on the time of year. In winter, the persistence of synoptic wind from either the northeast or the southeast dominates, while in the summer months, i.e., when the temperature and pressure gradients across the coastline increase, the mesoscale sea breeze circulation takes the lead. In spring and autumn, the atmospheric conditions vary rapidly, and the occurrences of either cluster are equally distributed.

4 Conclusions

In this work, the observations of wind intensity and direction provided by seven ground-based meteorological stations, located in a coastal area of central Italy, are used to describe the anemological peculiarities of the region. The data, covering a 7-year interval, are analysed using the k-means clustering algorithm, which is widely used in environmental and meteorological studies.

Clustering is carried out by considering seven objective parameters, based on the wind-integral quantities proposed by Allwine and Whiteman (1994) and on the diagnostics suggested by Li et al. (2020). The former allow the evaluation of the daily ventilation/stagnation conditions, while the latter provide useful parameters to identify the onset and the cessation of the SBR. This method allows discriminating between days when the atmospheric low-level circulation is dominated by the synoptic conditions and days when the SBR prevails.

For the region of interest, the algorithm identified three characteristic clusters:

- the Northeasterly Cluster, dominated by the synoptic conditions and characterized by prevailing winds from Northeast throughout the day;
- the Breeze Cluster, when the mesoscale circulation takes over and the onset of the sea breeze can be recognized from the increase in wind velocity and the wind rotation towards the southwest quadrant, i.e., perpendicularly to the coastline;

- the Southeasterly Cluster, dominated by the synoptic conditions and occurring when the wind blows from the southeast quadrant for the entire duration of the day.

From this information, general conclusions can be addressed. When the Northeasterly Cluster occurs, Italy appears to be located at the convergence of a high-pressure system located over the North Sea and a low-pressure system over the Balkans. Conversely, when the Southeasterly Cluster arises, the Italian peninsula is crossed by a large baric gradient, which generates winds from the southern quadrant. For both the Northeasterly and the Southeasterly Clusters, the synoptic circulation dominates over the local and the mesoscale dynamics, and the SBR fails to develop. Nonetheless, in the latter case the atmospheric flows at the different scales interact, generating a more complex anemological pattern, in which the SBR causes wind intensity to increase and wind direction to veer clockwise during the central hours of the day, after which the process is, to some extent, reversed in the afternoon. Moreover, the Southeasterly Cluster can occur in conjunction with Saharan dust outbreaks. Conversely, the Breeze Cluster is characterized by the prevalence of the local circulation over the synoptic one. The synoptic circulation mainly governs the coastal weather conditions from October to March, when the Northeasterly and the Southeasterly Clusters occur in more than half of the selected cases. On the contrary, the SBR prevails in summer, even if the sea breeze regime can be detected during the whole year.

It should be emphasized that all available data are forcibly classified according to the prescribed parameter k_O , which is set a priori as the number of clusters which better accounts for most possible weather configurations, based on the analysis of the specific time series in use. This implies that ambiguous events (e.g., rotations of the synoptic-scale wind along the day for reasons other than the interaction with sea breeze) can be erroneously attributed to one of the clusters, when they in fact would call for their own category. Another possible confounding factor arises from the need of limiting the analysis to stations that lie within 30 km of the coastline, which only allows including one urban location. As a consequence, the effects of the Rome UHI on the low-level local circulation could not be evaluated in the present investigation. Future experimental research is planned (1) to assess whether accounting for the UHI requires increasing k_O , and (2) to generalize the clustering algorithm by linking the onset and cessation of the sea breeze to the hours of light, which also enables to consider the seasonal trend in insolation. The results of this work can help investigate the relationship between meteorological clusters and air quality in coastal areas, and characterize the local wind energy resource. If sufficiently long datasets are available, the proposed methods

can be applied to the analysis of the regional climatology of coastal areas.

Acknowledgements The authors gratefully acknowledge ARSIAL for the great effort in providing anemological data and the BAQUNIN (<https://www.baqunin.eu/>) team for the technical and scientific support.

Funding This research was supported by ESA BAQUNIN contract 4000126749/19/I-NS.

Data availability All the data used in the present paper are available on request from the corresponding author (annalisa.dibernardino@uniroma1.it). In-situ meteorological data are provided by ARSIAL agency and can be downloaded from <http://www.arsial.it/>.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Adame JA, Notario A, Villanueva F, Albaladejo J (2012) Application of cluster analysis to surface ozone, NO₂ and SO₂ daily patterns in an industrial area in CentralSouthern Spain measured with a DOAS system. *Sci Total Environ* 429:281291. <https://doi.org/10.1016/j.scitotenv.2012.04.032>
- Albergel C, Dutra E, Munier S, Calvet JC, Munoz-Sabater J, Rosnay PD, Balsamo G (2018) ERA-5 and ERA-Interim driven ISBA land surface model simulations: which one performs better? *Hydrol Earth Syst Sci* 22(6):3515–3532. <https://doi.org/10.5194/hess-22-3515-2018>
- Allwine KJ, Whiteman CD (1994) Single-station integral measures of atmospheric stagnation, recirculation and ventilation. *Atmos Environ* 28(4):713–721. [https://doi.org/10.1016/1352-2310\(94\)90048-5](https://doi.org/10.1016/1352-2310(94)90048-5)
- Al-Shammari ET, Shamshirband S, Petković D, Zalnezhad E, Yee L, Taher RS, Čojbašić Ž (2016) Comparative study of clustering methods for wake effect analysis in wind farm. *Energy* 95:573–579. <https://doi.org/10.1016/j.energy.2015.11.064>
- Azorin-Molina C, Tijm S, Chen D (2011) Development of selection algorithms and databases for sea breeze studies. *Theor Appl Climatol* 106(3):531–546. <https://doi.org/10.1007/s00704-011-0454-4>
- Bauer TJ (2020) Interaction of urban heat island effects and land–sea breezes during a New York City heat event. *J Appl Meteorol Climatol* 59(3):477–495. <https://doi.org/10.1175/JAMC-D-19-0061.1>
- Di Bernardino A, Iannarelli AM, Casadio S, Perrino C, Barnaba F, Tofful L, Campanelli M, Di Liberto L, Mevi G, Siani AM, Cacciani M (2021a) Impact of synoptic meteorological conditions on air quality in three different case studies in Rome. *Italy Atmos Pollut Res* 12(4):76–88. <https://doi.org/10.1016/j.apr.2021.02.019>
- Di Bernardino A, Mazzella V, Pecci M, Casasanta G, Cacciani M, Ferretti R (2021b) Analysis of local circulation and related UHI effect in the urban area of Rome: WRF simulations and ground-based observations, submitted to *Bound Layer Meteorol*
- Di Bernardino A, Iannarelli AM, Casadio S, Mevi G, Campanelli M, Casasanta G, Cede A, Tiefengraber M, Siani AM, Cacciani M (2021c) On the effect of sea breeze regime on aerosols and gases properties in the urban area of Rome, Italy. *Urban Clim* 37:100842. <https://doi.org/10.1016/j.uclim.2021.100842>
- Cafaro C, Frame TH, Methven J, Roberts N, Bröcker J (2019) The added value of convection-permitting ensemble forecasts of sea breeze compared to a Bayesian forecast driven by the global ensemble. *Q J R Meteorol Soc* 145(721):1780–1798. <https://doi.org/10.1002/qj.3531>
- Cenedese A, Miozzi M, Monti P (2000) A laboratory investigation of land and sea breeze regimes. *Exp Fluids* 29(1):S291–S299. <https://doi.org/10.1007/s003480070031>
- Colacino M (1982) Observations of a sea breeze event in the Rome area. *Arch Meteorol Geophys Bioclimatol Ser b* 30(1–2):127–139. <https://doi.org/10.1007/BF02323399>
- Ferreira LB, da Cunha FF, de Oliveira RA, Fernandes Filho EI (2019) Estimation of reference evapotranspiration in Brazil with limited meteorological data using ANN and SVM—a new approach. *J Hydrol* 572:556–570. <https://doi.org/10.1016/j.jhydrol.2019.03.028>
- Fortelli A, Scafetta N, Mazzarella A (2016) Influence of synoptic and local atmospheric patterns on PM10 air pollution levels: a model application to Naples (Italy). *Atmos Environ* 143:218–228. <https://doi.org/10.1016/j.atmosenv.2016.08.050>
- Fu S, Rotunno R, Chen J, Deng X, Xue H (2021) A large-eddy simulation study of deep-convection initiation through the collision of two sea-breeze fronts. *Atmos Chem Phys* 21(12):9289–9308. <https://doi.org/10.5194/acp-21-9289-2021>
- Geetha A, Nasira GM (2014) Data mining for meteorological applications: decision trees for modeling rainfall prediction. In: 2014 IEEE international conference on computational intelligence and computing research, pp 1–4. <https://doi.org/10.1109/ICCCIC.2014.7238481>
- Gobbi GP, Barnaba F, Di Liberto L, Bolignano A, Lucarelli F, Nava S, Perrino C, Pietrodangelo A, Basart S, Costabile F, Dionisi D, Rizza U, Canepari S, Sozzi R, Morelli M, Manigrasso M, Drewnick F, Struckmeier C, Poenitz K, Wille H (2019) An inclusive view of Saharan dusta dvections to Italy and the Central Mediterranean. *Atmos Environ* 201:242–256. <https://doi.org/10.1016/j.atmosenv.2019.01.002>
- He BJ, Ding L, Prasad D (2020) Relationships among local-scale urban morphology, urban ventilation, urban heat island and outdoor thermal comfort under sea breeze influence. *Sustain Cities Soc* 60:102289. <https://doi.org/10.1016/j.scs.2020.102289>
- Huth R, Beck C, Philipp A, Demuzere M, Ustrnul Z, Cahynová M, Kysely J, Tveito OE (2008) Classifications of atmospheric circulation patterns: recent advances and applications. *Ann N Y Acad Sci* 1146(1):105–152. <https://doi.org/10.1196/annals.1446.019>
- Kodinariya TM, Makwana PR (2013) Review on determining number of Cluster in KMeans Clustering. *Int J Adv Res Comput Sci Manag Stud* 1(6):90–95
- Li W, Wang Y, Bernier C, Estes M (2020) Identification of sea breeze recirculation and its effects on ozone in Houston, TX, During DISCOVER-AQ 2013. *J Geophys Res Atmos* 125(22):e2020JD033165. <https://doi.org/10.1029/2020JD033165>

- Masselink G (1998) The effect of sea breeze on beach morphology, surf zone hydrodynamics and sediment resuspension. *Mar Geol* 146(1–4):115–135. [https://doi.org/10.1016/S0025-3227\(97\)00121-7](https://doi.org/10.1016/S0025-3227(97)00121-7)
- Mastrantonio G, Viola A, Petenko I, Argentini S, Conidi A, Coniglio L (2006) Caratterizzazione della circolazione locale mediante analisi di dati di vento. Il sistema ambientale della tenuta presidenziale di Castelporziano, II Serie, Accademia Nazionale delle Scienze detta dei Quaranta. “Scritti e Documenti” XXXVII, Roma, pp 13–49
- Miller STK, Keim BD, Talbot RW, Mao H (2003) Sea breeze: Structure, forecasting, and impacts. *Rev Geophys* 41:1011. <https://doi.org/10.1029/2003RG000124>
- National Research Council (1992) Coastal meteorology: a review of the state of the science. The National Academy Press, Washington, D.C. <https://doi.org/10.17226/1991>
- National Research Council (1994) Environmental science in the coastal zone: issues for further research. The National Academy Press, Washington, D.C. <https://doi.org/10.17226/2249>
- Ogaya R, Peñuelas J (2021) Climate change effects in a Mediterranean forest following 21 consecutive years of experimental drought. *Forests* 12(3):306. <https://doi.org/10.3390/f12030306>
- Petenko I, Mastrantonio G, Viola A, Argentini S, Coniglio L, Monti P, Leuzzi G (2011) Local circulation diurnal patterns and their relationship with large-scale flows in a coastal area of the Tyrrhenian Sea. *Bound-Layer Meteorol* 139(2):353–366. <https://doi.org/10.1007/s10546-010-9577-x>
- Pohl B, Saucède T, Favier V, Pergaud J, Verfaillie D, Féral JP, Krasniqi Y, Richard Y (2021) Recent climate variability around the Kerguelen Islands (Southern Ocean) seen through weather regimes. *J Appl Meteorol Climatol* 60(5):711–731. <https://doi.org/10.1175/JAMC-D-20-0255.1>
- Puygrenier V, Lohou F, Campistron B, Saïd F, Pigeon G, Bénech B, Serça D (2005) Investigation on the fine structure of sea-breeze during ESCOMPTE experiment. *Atmos Res* 74(1–4):329–353. <https://doi.org/10.1016/j.atmosres.2004.06.011>
- Steele CJ, Dorling SR, Glasow RV, Bacon J (2013) Idealized WRF model sensitivity simulations of sea breeze types and their effects on offshore windfields. *Atmos Chem Phys* 13(1):443–461. <https://doi.org/10.5194/acp-13-443-2013>
- UN (2017) The Ocean Conference. “Factsheet: People and Oceans.” New York. June 5–9, 2017 <https://www.un.org/sustainabledevelopment/wp-content/uploads/2017/05/Ocean-factsheet-package.pdf>. Accessed 24 July 2021
- Wei J, Tang G, Zhu X, Wang L, Liu Z, Cheng M, Munkel C, Li X, Wang Y (2018) Thermal internal boundary layer and its effects on air pollutants during summer in a coastal city in North China. *Res J Environ Sci* 70:37–44. <https://doi.org/10.1016/j.jes.2017.11.006>
- WMO (2017) WMO Guidelines on the calculation of climate normal. WMO-No. 1203
- Wong MA, Hartigan JA (1979) Algorithm as 136: A k-means clustering algorithm. *J R Stat Soc Ser C Appl Stat* 28(1):100108
- Zhong S, Takle ES (1993) The effects of large-scale winds on the sea-land-breeze circulations in an area of complex coastal heating. *J Appl Meteor* 32:1181–1195. <https://doi.org/10.1007/s10546-007-9185-6>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.